



The Center for International Cooperation in E-Business
China University of Geosciences, Wuhan 430074, P.R.China,
Tel: 0086-27- 67883357, Fax: 0086-27- 87801763, www.whiceb.com

The Center for International Cooperation in E-Business

The Center for International Cooperation in E-Business is a leading research and development unit of China University of Geosciences, Wuhan, with strong ties to industries, universities and research institutes. Its mission is to advocate an integrated and comprehensive approach to E-Business and E-Commerce development in China, and its strategic, managerial and organizational impacts of relative IT technology. Emphasis is on collaboration with international partners in public or private institutions and other research centers. Our aim is to provide thoughtful and practical guidance to organizations in the format of research projects, education/training program, consultancy and international/domestic trade support services. Our network comprises researchers and specialists from other leading business schools and center in Europe, North America and also Asia. The center is charged with the advancement of knowledge in E-Business and building links with government, industry and commerce.



University of Calgary Press

The Seventeenth Wuhan International Conference on E-Business

Wuhan, P. R. China
May 25- May 27, 2018



The Seventeenth Wuhan International Conference on E-Business

Wuhan, P. R. China

May 25- May 27, 2018

Conference hosts

The Center for International Cooperation in E-Business, China University of Geosciences, Wuhan, China
School of Economics and Management, China University of Geosciences, Wuhan, China
Baden-Wuerttemberg Cooperative State University Heidenheim, Heidenheim, Germany
College of Business, Alfred University, Alfred, New York, USA

Organizers

The Center for International Cooperation in E-Business, China University of Geosciences, Wuhan, China
School of Economics and Management, China University of Geosciences, Wuhan, China
School of Management, Wuhan Textile University
Research center of Enterprise Decision Support, Key Research Institute of Humanities and Social Sciences in Universities of HuBei Province

Conference Co-Chairs

Jing Zhao, Professor, CICEB, School of Economics and Management, China University of Geosciences
(zhao5563@outlook.com)
Juergen Seitz, Professor, Baden-Wuerttemberg Cooperative State University Heidenheim, Germany
(seitz@dhbw-heidenheim.de)
Doug Vogel, Professor, Harbin Institute of Technology, China, AIS President 2012-2013
(isdoug@hit.edu.cn)

Conference Honorary Chair

Wilfred V. Huang, Professor, Alfred University (fhuang@alfred.edu)

Publicity chair & Proceedings Editor

Yiliu (Paul) Tu, Professor, University of Calgary, Canada paultu@ucalgary.ca

Secretary-General

Zhen Zhu, Associate Professor, China University of Geosciences, Wuhan China (zhuzhen@cug.edu.cn)

International Advisory Board

Chair: Joey George, Professor of Iowa State University, USA, AIS President 2010-2011,
Robert Kauffman, Professor, Singapore Management University, Singapore

-Pacific Asian

Patrick Chau, Professor, The University of Hong Kong, Hong Kong
Guoqing Chen, Professor, Tsinghua University, China
Robert Kauffman, Professor, Singapore Management University, Singapore
Wei Kwok Kee, Professor, City University of Hong Kong, Hong Kong
Ting-Peng Liang, Professor, National Sun Yat-Sen University, Taiwan
Feicheng Ma, Professor, Wuhan University, China
Jiye Mao, Professor, Renmin University, China
Michael D. Myers, Professor, AIS President 2006-2007, University of Auckland, New Zealand
Bernard Tan, Professor, AIS President 2009-2010, National University of Singapore, Singapore
Kanliang Wang, Professor, Renmin University, China
Nilmini Wickramasinghe, Professor, Deakin University, Australia
Kang Xie, Professor, Sun Yat-Sen University, China
Qiang Ye, Professor, Harbin Institute of Technology, China
J. Leon Zhao, Professor, City University of Hong Kong, Hong Kong

-North American

Bob Carasik, Wells Fargo Bank, USA
Weiguo Fan, Professor, Virginia Polytechnic Institute and State University, USA
Joey George, AIS President 2010-2011, Iowa State University, USA
Zhangxi Lin, Professor, Texas Tech University, USA

Ning Nan, Assistant Professor, University of British Columbia, Canada
Paul A. Pavlou, Professor, Temple University, USA
Arun Rai, Professor, Editor in Chief of MISQ, Georgia State University, USA
Richard Watson, Professor, University of Georgia, USA
Han Zhang, Professor, Georgia Institute of Technology, USA

-European

David Avison, Professor, ESSEC, France
Niels Bjorn-Andersen, Professor, Copenhagen Business School, Denmark
Marco De Marco, Professor, Università Cattolica, Italy
Reima Suomi, Professor, Turku School of Economics, Finland
Yao-Hua Tan, Professor, Vrije University Amsterdam, Netherlands
Hans-Dieter Zimmermann, Professor, FHS St. Gallen, University of Applied Sciences, Switzerland

Editorial Board of the Proceedings

Yiliu (Paul) Tu, University of Calgary, Canada, **Editor**

-Cross-border e-Commerce Initiatives under China's Belt and Road Initiative Track

Christopher Westland, University of Illinois – Chicago, USA
Rong Du, Xidian University, China
Jian Mou Xidian University, China

-Virtual Teams in Information Technology Project Management Track

Xiaobo (Bob) Xu, The American University of Sharjah, Sharjah, UAE
Weiyong Zhang, The American University of Sharjah, Sharjah, UAE

-Emerging Issues in e-Business Track

Zhongyun (Phil) Zhou, School of Economics and Management, Tongji University, China
Xiao-Liang Shen, Wuhan University, China
Yongqiang Sun, Wuhan University, China
Xiao-Ling Jin, Shanghai University, China

-Big Data and Business Analytics Track

John Qi Dong, University of Groningen, Netherlands
Chia-Han Yang, National Cheng Kung University, Taiwan

-e-Service & e-Customer Experience in 21st Century China Track

Edward Kasabov, University of Huddersfield, UK
Alexander Warlow, Institute of Direct and Digital Marketing, UK

-Crowdsourcing, Co-creation & Social Innovation Track

Yuxiang (Chris) Zhao, Nanjing University of Science & Technology, China
Jian Tang, Central University of Finance and Economics, China

-Changing Consumers in the Digital World Track

Guoxin Li, Harbin Institute of Technology, China
Jiaoju Ge, Harbin Institute of Technology, China

-Social Network and Commerce Track

Yaobin Lu, Huazhong University of Science & Technology, China
Jiang Wu, Wuhan University, China
Ling Zhao, Huazhong University of Science & Technology, China

-Social Media in the Workplace Track

Hefu Liu, University of Science and Technology of China, China
Qian Huang, University of Science and Technology of China, China

-Enterprise Social Media Track

Yi Wang, Southwestern University of Finance and Economics, China
Yang Chen, Southwestern University of Finance and Economics, China
Si Shi, Southwestern University of Finance and Economics, China

-e-Business Strategy with Data Driven Track

Liping Qian, Chongqing University, China
Xingyao Ren, Nankai University, China
Xiaoling Li, Chongqing University, China

-User Behavior and Personalized Service Track

Weijun Wang, Central China Normal University, China
Chunmei Gan, Sun Yat-sen University, China

-General Topics Track

Zhen Zhu, China University of Geosciences, China

Program Committee

Carl Adams, University of Portsmouth, UK
Dimitris Assimakopoulos, Grenoble Ecole de Management, France
Ezendu Ariwa, London Metropolitan University, UK
David Avison, ESSEC, France
Jalal Baghdadchi, Alfred University, USA
Niels Bjorn-Andersen, Copenhagen Business School, Denmark
Erran Carmel, American University, Washington DC, USA
Sally Chan, Principal, SMC Solutions, Audit and Operations Consulting, Canada
Hsi-Ya Chang, National Center for High-Performance Computing, Taiwan
Patrick Chau, The University of Hong Kong, Hong Kong
Wojciech Cellary, The Poznan University of Economics, Poland
Narciso Cerpa, University of Talca, Chile
Guoqing Chen, Tsinghua University, Beijing, China
Shiping Chen, CSIRO ICT Centre, Australia
Jinjin Chen, Swinburne University of Technology, Australia
Edwin Cheng, The Hong Kong Polytechnic University, Hong Kong
Roger Chiang, University of Cincinnati, Germany
Dickson K.W. Chiu, Dickson Computer Systems, Hong Kong
David Chou, Eastern Michigan University, USA
Huey-Wen Chou, National Central University, Taiwan
Dickson K.W. Chiu, Dickson Computer Systems, Hong Kong
Alessandro D'Atri, LUISS Guido Carli University, Italy
Rober Davison, City University of Hong Kong, Hong Kong
Haluk Demirkan, Arizona State University, USA
Weiguo Fan, Virginia Tech, USA
Ali Farhoomand, The University of Hong Kong,
Rimantas Gatautis, Kaunas University of Technology, Lithuania
Joanne Gellatly, George Brown University, Canada
Xianjun Geng, The University of Texas at Dallas, USA
Joey George, Iowa State University, USA
Philip Gordon, University of Berkeley, USA
Varun Grover, Clemson University, USA
Tor Guimaraes, Tennessee Technological University, USA

Guoqing Guo, People's University of China, China
 Alan R. Hevner, University of South Florida, USA
 Robert C. H. Hsu, Chung Hua University, Taiwan
 Kai-Lung Hui, Hong Kong University of Science and Technology, Hong Kong
 Roger D. Huang, University of Notre Dame, USA
 Wayne Wei Huang, Ohio University, USA
 Fengjie Jing, Wuhan University, China
 Jahng Jung Joo, Seoul National University, South Korea
 Diana Kao, University of Windsor, Canada
 Okyay Kaynak, Bogazici University, Turkey
 Wei Kwok Kee, City University of Hong Kong, Hong Kong
 Subodha Kumar, University of Washington, USA
 Ricky Y K Kwok, The Hong Kong University, Hong Kong
 Winfried Lamersdorf, University of Hamburg, USA
 Francis Lau, Hong Kong University, Hong Kong
 Kyoung Jun Lee, Kyung Hee University, South Korea
 Biqiang Li, Wuhan University of Technology, China
 Keqin Li, State University of New York at New Paltz, USA
 Qi Li, Xi'an Jiaotong University, China
 Zhangxi Lin, Texas Tech University, USA
 Ting-Peng Liang, National Sun Yat-Sen University, Taiwan
 Chunlu Liu, Deakin University, Australia
 Lu Liu, Beihang University, China
 Claudia Loebbecke, University of Cologne, Germany
 Mary Loyland, University of North Dakota, North Dakota, USA
 Feicheng Ma, Wuhan University, China
 Marco De Marco, Università Cattolica, Italy
 Rogers Matama, lucky family business, Uganda
 Rossi Matti, Aalto University, Finland
 Donald McCubbrey, University of Denver, USA
 Shaozu Mei, University of Science & Technology of Beijing, China
 Teo Yong Meng, Sun Microsystems Inc., National University of Singapore
 Maryska Milos, University of Economics, Prague
 Kazuo Misue, University of Tsukuba, Japan
 Vojislav B. Mistic, University of Manitoba, Canada
 Ali R. Montazemi, McMaster University, Canada
 San Murugesan, Southern Cross University, Australia
 Michael D. Myers, University of Auckland, New Zealand
 Charles Omagor, Makerere University, Uganda
 Khaled Negm, Internet Security Systems, Egypt
 Guihua Nie, Wuhan University of Technology, China
 Marko Nieminen, Aalto University, Finland
 Richard Nolan, University of Washington, USA
 Tim Norman, University of Aberdeen, UK
 Gal Oestreicher-Singer, New York University, USA
 Harri Oinas-Kukkonen, University of Oulu, Finland
 Dolun Oksoy, Ankara University, Turkey

Jim Oliver, INSEAD, France
 Charles Omagor, Makerere University Business School, Uganda
 Mario Paolucci, LABSS-ISTC-CNR, Italy
 Hugh Pattinson, University of Western Sydney, Australia
 Paul A. Pavlou, Temple University, USA
 Dunlu Peng, Fudan University, China
 Rajesh K Pillania, Management Development Institute, India
 David Pistruì, Acumen Dynamics, USA
 Thomas E. Potok, Oak Ridge National Laboratory, USA
 Michael Z Qiao, George Brown College, Canada
 Zheng Qin, Shanghai University of Finance and Economics, China
 G. Ramakrishna, Vidya Jyothi Institute of Technology, India
 Kai Rannenberg, Goethe University, Germany
 Sourav Ray, McMaster University, Canada
 Gatautis Rimantas, Kaunas University of Technology, Lithuania
 Xiaomin Ruan, University of New Mexico, USA
 Perry Sadorsk, York University, Canada
 Richard Scudder, Denver University, Denver, USA
 Juergen Seitz, University of Cooperative Education, Heidenheim, Germany
 Choon Ling Sia, City University of Hong Kong, Hong Kong
 Keng Siau, University of Nebraska-Lincoln, USA
 Edgar H Sibley, George Mason University, USA
 Yesha Sivan, The Academic College of Tel-Aviv Yaffo, Israel
 Detmar W. Straub, Georgia State University, USA
 Reima Suomi, Turku School of Economics, Finland
 Cheickna Sylla, New Jersey Institute of Technology, USA
 Liwen Tan, Wuhan University, China
 Yao-Hua Tan, Vrije University Amsterdam, Netherlands
 Domenico Talia, University of Calabria, Italy
 Tsutomu Terada, Kobe University, Japan
 Jason Thatcher, Clemson University, USA
 Hock Hai Toe, National University of Singapore, Singapore
 Paul Tu, University of Calgary, Canada
 Nigel Upton, London Business College, United Kingdom
 Kunal Verma, University of Georgia, USA
 Brian Vinter, Copenhagen University, Denmark
 Bernhard Volz, University of Bayreuth, Germany
 Eric Wang, National Central University, Taiwan
 Shan Wang, Renmin University of China, China
 Xuedong Wang, Central China Normal University, China
 Richard Watson, University of Georgia, USA
 Chih-Ping Wei, National Taiwan University, Taiwan
 Andrew Whinston, University of Texas, USA
 Brian Whitworth, Massey University (Albany), New Zealand
 Mu Xia, University of Illinois Urbana-Champaign, USA
 Tianliang Xu, Huazhong University of Technology and Sciences, China
 Ruiliang Yan, Indiana University Northwest, USA

Chunyang Ye, University of Toronto, Canada
Yufei Yuan, McMaster University, Canada
Daniel Zeng, University of Arizona, USA
Dongsong Zhang, University of Maryland, USA
Han Zhang, Georgia Institute of Technology, USA
Jinglong Zhang, Huazhong University of Science and Technology, China
Lixin Zhang, IBM Austin Center for Advanced Studies, USA
Xinguo Zhang, Zhongnan University of Economics and Law, China
Fang Zhao, American University of Sharjah, UAE
J. Leon Zhao, Eller College of Management, University of Arizona, USA
Lina Zhou, University of Maryland, USA
Daniel Zhu, Iowa State University, USA
Haibin Zhu, Nipissing University, Canada
Xingrong Zhu, Zhongnan University of Economics and Law, China
Hans-Dieter Zimmermann, FHS St. Gallen, University of Applied Sciences, Switzerland
Ralf Zimmermann, ZBS, Germany
YanJun Zuo, University of North Dakota, USA
Vladimir Zwass, Fairleigh Dickinson University, USA

Sponsors

The Association for Information Systems (AIS)
China Association for Information Systems (CNAIS)
The China Information Economics Society
Baden-Wuerttemberg Cooperative State University Heidenheim, Heidenheim, Germany
School of Information Systems, Singapore Management University, Singapore
University of Calgary, Canada
Deakin University, Australia
University of North Dakota, USA
New Jersey Institute of Technology, New Jersey, USA
University of Turku, Finland
Huazhong University of Sciences & Technology, China
Wuhan University, China
Wuhan University of Technology, China
Zhongnan University of Economics and Law, China
Huazhong Normal University, China

Sponsoring journals (alphabetical order)

Electronic Commerce Research (SSCI index)
Electronic Commerce Research and Applications (SSCI index)
Electronic Markets-The International Journal on Networked business (SSCI index)
International Journal of Information Systems and Change Management (EI index)
International Journal of Information Technology and Management (EI index)
International Journal of Networking and Virtual Organizations (EI index)
International Journal of Services Technology and Management (EI index)
The Journal of Virtual Worlds Research

Preface

The annual Wuhan International Conference on E-Business (WHICEB) is an AIS affiliated conference. WHICEB 2018 is held in Wuhan, P. R. China from May 25 to 27, 2018. WHICEB promotes intellectual research and facilitates academic and corporate networking in e-business and related fields. The intent is to encourage academic research and business development through exchanging ideas about the e-business, global and corporate financial issues, and necessity for continuous innovation. The conference aims at presenting innovative research findings, solutions and approaches to make the Internet a productive and efficient vehicle for global commerce. Whether running an e-business or transforming a business into e-business, we constantly encounter challenges ranging from technological to behavioral issues, from marketing to data analysis issues, and from effectiveness to security issues. The past years all over the world initiatives have been started for the next step of development. Some people have already talked about the fourth industrial revolution. After consumer oriented mass production we focus nowadays on personalized products and services, which demands cyber physical systems, cloud computing and big data. There are integration issues for management of technology, management of supply chains, management of human resources and management of knowledge and intelligence that are being resolved in an e-business environment. Organizations, regardless of its locations and sizes, should consider having a strategic decentralized planning effort that includes e-business as a pillar for sustainable competitive advantage.

Proceedings of the Seventeenth WHICEB document the breadth and depth of research from different aspects of business and from different disciplines that have major implications for e-business. The tracks in the proceedings include: Cross-border e-Commerce Initiatives under China's Belt and Road Initiative; Virtual Teams in Information Technology Project Management; Emerging Issues in E-Business; Big Data and Business Analytics; e-Service & e-Customer experience in 21st century China; Crowdsourcing, Co-creation & Social Innovation; Changing Consumers in the Digital World; Social Network and Commerce; Social Media in the Workplace; Enterprise Social Media; e-Business Strategy with Data Driven; User Behavior and Personalized Service; and General Topics. The proceedings will be listed in the appropriate indexes. The selected best papers from the proceedings will be recommended to international academic journals/special issues including but not limited to the following: Electronic Commerce Research and Applications (SSCI), International Journal of Networking and Virtual Organizations (EI), and International Journal of Services Technology and Management (EI).

The research papers in the proceedings went through a double blind peer review process. Papers are accepted based upon a clear research methodology and contributions to the knowledge of e-business including but not limited to case study, experiment, simulation or survey. The efforts made by our track chairs in reviewing submissions are really appreciated, which ensures the quality of the proceedings. I personally thank them for their professional diligence. They are: Christopher Westland, Rong Du, and Jian Mou, Cross-border e-Commerce Initiatives under China's Belt and Road Initiative; Xiaobo (Bob) Xu, Weiyong Zhang, Virtual Teams in Information Technology Project Management; Zhongyun (Phil) Zhou, Xiao-Liang Shen, Yongqiang Sun, Xiao-Ling Jin, Emerging Issues in e-Business; John Qi Dong, Chia-Han Yang, Big Data and Business Analytics; Edward Kasabov, Alexander Warlow, e-Service & e-Customer Experience in 21st Century China; Yuxiang (Chris) Zhao, Jian Tang, Crowdsourcing, Co-creation & Social Innovation; Guoxin Li, Jiaoju Ge, Changing Consumers in the Digital World; Yaobin Lu, Jiang Wu, Ling Zhao, Social Network and Commerce; Hefu Liu, Qian Huang, Social Media in the Workplace; Yi Wang, Yang Chen, Si Shi, Enterprise Social Media; Liping Qian, Xingyao Ren, Xiaoling Li, e-Business Strategy with Data Driven; Weijun Wang, Chunmei Gan, User Behavior and Personalized Service; Zhen Zhu, General Topics.



Yiliu (Paul) Tu

Editor, Proceedings of Sixteenth Wuhan International Conference on E-Business
University of Calgary, Canada,

Welcome Message from the President of China University of Geosciences

The world today is in the wave of economic globalization and informatization. The Internet and Information Technology (IT) have provided new opportunities to the world economy with tremendous achievements. In this globalizing information era, IT and the Internet are silently changing the way people work, live and learn. In recent years, many scholars have conducted research on e-business from many different disciplines such as computer science, telecommunications, economics, management, human resources, law and sociology. As a result of these efforts, e-business has created a bright perspective for further development in the world and is expected to continue to contribute to global growth and stability.

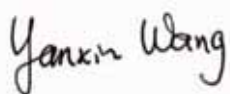
The Wuhan International Conference on E-business is jointly organized by the International Cooperation Center for E-business and the College of Economy and Management, China University of Geosciences (Wuhan), and by the College of Business, Alfred University in the United States and Baden-Wuerttemberg Cooperative State University Heidenheim, Heidenheim, Germany, and by the School of Management, Research center of Enterprise Decision Support(Key Research Institute of Humanities and Social Sciences in Universities of HuBei Province) ,Wuhan Textile University.

This conference has become an AIS Affiliated Conference In 2011, and the Proceedings of the Seventeenth Wuhan International Conference on E-Business (WHICEB 2018) will be included into AIS library. The focus of this conference is related to realizing the full potential of technology in the globalization era. The conference will demonstrate the latest achievement in the fields of Internet economy and organizations. Conference tracks include Cross-border e-Commerce Initiatives under China's Belt and Road Initiative; Virtual Teams in Information Technology Project Management; Emerging Issues in E-Business; Big Data and Business Analytics; e-Service & e-Customer experience in 21st century China; Crowdsourcing, Co-creation & Social Innovation; Changing Consumers in the Digital World; Social Network and Commerce; Social media in the workplace; Enterprise social media; E-business strategy with data driven; User behavior and personalized service; and general topics. It will providing scholars from home and abroad an academic exchange platform for the promotion of technological innovation and international cooperation, and hence the healthy development of e-business in the world.

CUG, a state key university in China with geosciences as its world-famous academic programs, always attaches great importance to the studies of the interaction among population, economy, society, resources and the environment, in an attempt to promote the harmonious development of both human and nature, and pays great attention to the development of international academic exchange platforms for promoting the international collaboration, and facilitating the process of internationalization of the university.

In recent years, great progress has been made on e-business and the development of management science, which not only promotes the development of management science in CUG, but also provides an important support for the construction of a complete disciplinary system based on geosciences for CUG.

The prodigious economic growth in China will provide essentials and supports for the studies of e-business and related fields. E-business will have a major impact on our future prosperity in all facets of life, business and government. Let us work together to strive for a more dynamic e-environment and a more enjoyable life for mankind in this magic e-era!



Yanxin Wang

PhD, Professor

President

China University of Geosciences, Wuhan, China

Welcome Message from the AIS President

Greetings to all WHICEB 2018 Delegates!


I am delighted to welcome each of you to the 17th annual Wuhan International Conference on E-Business! I have been championing local communities and their conferences and it is great to see strong conferences, such as WHICEB, emerging and growing steadily. They afford opportunities for faculty and students to come together, form friendships, and enjoy exchanging ideas. Through fostering such relationships and ideas, WHICEB cultivates high quality scholarship in China, in Asia, and around the globe.

On behalf of the Association for Information Systems (AIS), I would like to thank the conference organizers, Jing Zhao, from the China University of Geosciences and Doug Vogel from Harbin Institute of Technology as well as Juergen Seitz from Baden-Wuerttemberg Cooperative State University in Heidenheim, Germany for their service to the Information Systems community this conference. I would like to extend a special thanks to Wilfred V. Huang, honorary conference chair, for his many years of service. The conference organizers have assembled a world-class conference committee, attracted a strong set of globally recognized journal outlets, and attracted significant global sponsorship to ensure that the conference, and its delegates is an ongoing success.

The AIS is honored to count WHICEB as an affiliated conference. It provides an exemplary example of how a group of exemplary scholars can come together and have a lasting impact on the quality of research and networking that is necessary to advance the Information Systems discipline. The AIS invites all WHICEB delegates to participate in the global academic community at conferences such as the International Conference on Information Systems (ICIS) or the Pacific Asia Conference on Information Systems (PACIS) and more focused venues such as the Workshop on Information Systems Economics (WISE) or the Workshop on Information Technology and Systems (WITS).

As AIS President, I personally invite each of you to familiarize yourself with the Association. We offer a growing set of digital services, such as the eLibrary, where the WHICEB conference proceedings will be available to download, webinars on research methods and career management, and access to workshops that can help develop your skills for research and teaching. For members and non-members alike, I urge you to visit www.aisnet.org to learn more about the benefits of membership.

I cordially invite you to join the WHICEB 2018, and look forward to reading your work in the future.



Matti Rossi
President, Association for Information Systems (2017-2018)
Professor of Information Systems
Aalto University School of Business

Keynote Speakers



Marshall Van Alstyne, coauthor of the international bestseller *Platform Revolution*. He is one of the world's experts on network business models and is Everett Lord Distinguished Scholar at Boston University's Questrom School of Business. He is a frequent speaker, board level advisor, and consultant to startups and global firms. His research has received half a dozen academic awards and appeared in journals such as *Science*, *Nature* and *Harvard Business Review*. Interviews appear regularly across Bloomberg, The Economist, The New York Times, The Wall Street Journal and National Public Radio. He studied computer science at Yale and information technology at MIT. He holds multiple patents; was among the first to measure the dollar value of social networks, and his theories of network businesses are taught worldwide. He is a husband and dad, who loves dogs, exercise, travel, and questions of governance.



Bin Gu, Earl and Gladys Davis Distinguished Professor and associate dean of China Programs at the W P Carey School of Business at Arizona State University. Professor Gu's research interests are in data analytics, online platforms, online social media and social network, mobile commerce and IT-enabled business models. His work has appeared in *Management Science*, *MIS Quarterly*, *Information Systems Research*, *Journal of Management Information Systems*, *Production and Operations Management*, *Journal of Retailing*, *Decision Support Systems*, *Journal of Financial Service Research* and other academic journals.

In addition to his research activities at the W. P. Carey School of Business, Professor Gu maintains an active international research agenda. He has received multiple research grants from the Robert Wood Johnson Foundation, the NET Institute, Center for the Studies of Economic Liberty, Center for Service Leadership and National Science Foundation of China. Professor Gu teaches Emerging Technology, IT-enabled Business Model and Business Analytics Strategy across multiple online and onsite platforms at the graduate level both in the US and in China. Prior to joining ASU, Professor Gu served on the faculty at McCombs School of Business at The University of Texas at Austin.

Professor Gu currently serves on the editorial boards of *MIS Quarterly* as a senior editor, and *Information Systems Research* as an associate editor. He also serves as an associate editor or program committee member of International Conference on Information Systems (ICIS), INFORMS Conference on Information Systems Technology (CIST) and ACM Conference on Electronic Commerce (ACM EC).

Professor Gu's research was awarded the 2017 Hawaii International Conference on System Sciences Best Paper Award, the 2016 W P Carey Faculty Research Award, the 2012 and 2014 Emerald Citations of Excellence Award, the 2014 Americas Conference on Information Systems Best Research-in-Progress Award, the Journal of Retailing Top Cited Article 2007-2011, the 2008 Information Systems Research Best Published Paper Award, and the 2007 International Conference on Information Systems (ICIS) Best Paper-in-Track Award. Before joining academia, Professor Gu had worked for Arthur Andersen as a consultant.

CONTENTS

Cross-border e-Commerce Initiatives under China's Belt and Road Initiative

The Influencing Factors Model of Cross-Border E-commerce Development: A Theoretical Analysis

Li Junbo, Zhang Jue, Qu Fei, Zhao Yuehan.....1

New Weather Indices for China: Tool of Risk Control of International Supply Chain

Qing Zhu, Jiarui Li, Jian Chai.....9

An Exploration of Cross-border E-commerce Consumer Feedbacks: An LDA Approach

Jian Mou, Gang Ren, Chunxiu Qin, Kerry Kurcz22

Understanding IS Success Model and Valence Framework in Sellers' Acceptance of Cross-border E-commerce

Yi Cui, Jian Mou, Jason Cohen, Yanping Liu.....31

Policy Recommendations for Promoting the Development of Cross-Border E-Commerce between China and Central Asian Countries

Tingting Xiao, Shizhong Ai, Weili Zhang39

Research on The Optimization Strategy of Cross-border B2B Supply Chain with Service Cost Information Sharing

Dan Fang, Wensheng Yang, Jun Ma, Yiliu Tu48

Empirical Research on the Relationship between Foreign Trade Growth and Cross-Border Electronic Commerce in Fujian Province

Feifei Huang, Lu Chen.....56

Virtual Teams in Information Technology Project Management

Research on Influence Factors of Information Technology Enterprise's Operating Performance

Jiangping Wan, Jiawen Huang, Zhong Wang63

An Analysis with Evolutionary Game of the Resource Sharing in Supply Chain Under Cloud Platform

Ling-wu Zeng, Yi Hu, Xing-jian Zhou71

The Analysis on Multimodal Transport Mode of Cross-border E-commerce with 'the Belt and Road' Strategy Based on Personalized Recommendation

Mingli Zhang, Yanling Fan, Man Chen79

Scheduling Limited Resources in Engineering Projects

Samer Ben Issaa, Yiliu Tu88

Emerging Issues in E-Business

An Analysis on the Development Model of China's County-level E-commerce

Kegui Liao, Yuzhen Zhang, Xuewen Gui.....99

Understanding WeChat User's Intention to Use Various Functions: from Social Cognitive Perspective <i>Shenhao Zhou, Xiaoli Yu, Chuan Luo</i>	109
Online-to-offline (O2O) Platforms, Proprietary Platforms and Firm Performance <i>Xing Wan, Xue Feng</i>	117
Research on the Influencing Factors of the Willingness to Pay for Knowledge Consumers in the Knowledge Payment Platform <i>Xinyan Liu, Jun Feng</i>	124
Research on the Evaluation of Green Logistics Based on Cloud Model <i>Xiulian Jiang, Yining Song, Ling Ding, Zihan Jiang</i>	132
Determinants of online sellers' advertising: An empirical study on Chinese C2C market <i>Weijia You</i>	140
Research on Poverty Alleviation of County E-Commerce-A Case Study of Luotian County <i>Li Li, Bin Xu, Huiming Xiao, Yang Li</i>	145
 Big Data and Business Analytics	
A Hybrid Model to Analyze Air Pollution Spread Scales in Xi' an and Surrounding Cities <i>QingZhu, TingWang, Yiqiong Wu, JianChai</i>	155
Fine-grained Aspect Extraction for Online Reviews of E-commerce Products Based on Semi- supervised Learning <i>Huosong Xia, Yitai Yang</i>	165
The Relationship Between Investor Sentiment and Stock Market Volatility: Based on the VAR Model <i>Ge Zhang, Jishun Wang, Hao Guo, Xin Zhang</i>	173
E-commerce Service and Supply Chain Finance in Rural China: A Value Captured Perspective <i>Yi Chen</i>	181
Research on Zipf's Law of Hot Events in Search Engines <i>XU Ying-fan, SHI Ming-liang</i>	189
Quality Prediction of Answers in Community of Questions and Answers of Zhihu <i>Ming Li, Yi Zhang, Kaijuan Xing, Xiaoyu Qi</i>	196
A Study on Online-Shopping Opinion Leaders Based on Massive Data <i>Chong Zhang</i>	204
The Evaluation of E-commerce Efficiency in China using DEA-Tobit model: evidence from Taobao data <i>Zongwei Li, Dongqing Luan, Yanhui Zhang, Peng Guo</i>	211
 e-Service & e-Customer experience in 21st century China	
The Differences of Online Review Textual Content: A Cross-Cultural Empirical Study <i>Liyi Zhang, Li Li</i>	219

e-Control of Online Customers:When Companies Make Customers Compliant ... Lessons for 21st Century China <i>Edward Kasabov</i>	227
A Research about Consumer's Usage Intention to Green Finance Products —— Taking the Alipay's Ant forest as the Example <i>Hong Peng, Jun Mi</i>	235
Can the Adjustment of Consumer Finance Credit Line Cause Impulse Purchase? ——An Analysis from the Perspective of E-commerce Promotion <i>Hong Peng, Yushan Yusufujiang</i>	243
Crowdsourcing, Co-creation & Social Innovation	
Research on Innovation Ecosystem of the Crowd Innovation Space Based on System Dynamics <i>Bin Chen, Xin Jin, Jing Wu, Hu Yang</i>	252
Inter-organizational Information Technology and Joint Competitive Advantages: An Integrative Model of Co-creating IT value <i>Lu Sun, Chengyi Yue</i>	260
Exploring Gender Effects on Peer Rating in Open Innovation and Crowdsourcing: A Case of Website Evaluation <i>Liang Chen</i>	268
Changing Consumers in the Digital World	
A Text Mining Based Approach for Mining Customer Attribute Data on Undefined Quality Problem <i>Qing Zhu, Yiqiong Wu, Yuze Li, Renxian Zuo</i>	276
The Influence of Media News Frame on Consumer's Brand Attitudes in the Context of Product-harm Crisis <i>Silan Li, Shiyu Liu, Mengzhi Xian</i>	290
A Study on Consumers' Learning Effect in the Price Reduction Auction: a case study of Gongtianxia <i>Yanan Xu, Bo Yang, Rong Zhang, Pengfei Xie</i>	298
Consumer Coupon Redemption Behavior Prediction on B2C E-commerce <i>Wenmin Qi, Li Li, Ruibo Yao</i>	310
Measurement on Short-term Effect and Purchase Conversion Mechanism of Online Advertising <i>Weiwei Liao, Li Li, Ruibo Yao</i>	312
How do Patients with Chronic Diseases Make Usage Decisions Regarding Mobile Health Monitoring Services? <i>Fanbo Meng, Xitong Guo, Kee-hung Lai, Xinli Zhao</i>	321
The Motivation of Capital-giving in Crowdfunding Market: A Self-determination Theory Perspective <i>Xiang Yuan, Hongwei Wang, Yuan Meng</i>	330

Social Network and Commerce

What Factors Will Determine Users' Knowledge Payment Decision? An Theoretical and Empirical Research	
<i>Yang Zhao, Liqiang Wu, Yu Zhao, Xini Yuan</i>	333
Network Structure Mining and Evolution Analysis - Based on BA Scale-Free Network Model	
<i>Rui chen , Ting He</i>	335
Exploring Users' Intention to use QQ's Various Functions based on Social Cognitive Theory	
<i>Xun Wang, Ying Wu, Chuan Luo, Yalan Li</i>	341
Study on Enterprises' Internet Public Opinion Area Hotspots Based on Social Network Analysis	
<i>Yuanheng Li, Yan Tu, Xuefeng Li</i>	350
A Bibliometric Review of Research Trends in Social CRM	
<i>Jiangping Wan, Leqi Xie</i>	358
The Impacts of Environmental Cues and Browsing Experience on Impulse Buying on Social Shopping Website	
<i>Yanhong Chen, Jingwen Liu, Si Fan</i>	366
Role of Danmu Function in User Experience and Engagement: A Double-edged Sword	
<i>Si Fan, Jingwen Liu, Ting Zhu</i>	375
Self-disclosure in Social Network Sites: An Integration of Stimulus-Organism-Response Paradigm and Privacy Calculus Model	
<i>Ning Zhang, Jiaqi Pang, Jinlin Wan</i>	382
Assessing Customer-Based Brand Equity of E-Commerce Platform Based on Promotion Benefits	
<i>Ting Zhu, Jingwen Liu, Yanhong Chen</i>	391

Social media in the workplace

Network Structure and Employee Creativity in a Heterogeneous Team	
<i>Jianping Peng, Taiye Luo, Jing Quan, Peiwen Guo</i>	398
A Net Loan Monitoring Platform for University Students Based on Visual Micro-blog	
<i>Ruijun Zhang, Caiyan Lin, Xinxin Zheng, Lixin Gao, Jiaxin Tong, Shangyang Yu</i>	406
Inhibitors of Continuance Intention to Use Mobile Social Networking Sites: The Effects of Stress and Regret	
<i>Chunmei Gan, Tingting Li</i>	415
Product Information Diffusion in a Social Network and Marketing Implications: A Case Study of Huawei Mobile Phone	
<i>Zhang Ling, Luo Manman, Zhu Lijun</i>	423

Enterprise social media

The Analyzing on the Connotation, Properties and Research Trend of Enterprise Social Media	
<i>Yu Wang, Daqing Zheng, Lin Huang</i>	432

Convert Traffic to Purchase: the Impact of Social Network Information on Trust and Purchase Intention in Social Commerce <i>Yanli Pei, Min Zhang, Yuqi Zhang, Shan Wang</i>	440
Exploring Users' Interactive Behaviors in Online Group: A Case Study of QQ Group "TuanRenTang" <i>Lei Zhu, Chunxiang Xue, Xiuzhi Wu</i>	449
E-business strategy with data driven	
The Evaluation on Home Improvement Website's Efficiency Based on SBM-DEA <i>Jiangping Wan, Jiawen Huang</i>	457
Seller Channel Choice and Optimal Pricing on Heterogeneous Platforms under Online Price Comparison System <i>Qingfang Sang</i>	465
Quantifying the Effect of Mobile Channel Visits on Firm Revenue <i>Yingnan Zhao, Quansheng Wang, Peijian Song, Fan Feng</i>	473
Research on Spillover Effect of Paid Search Advertising Channels <i>Chaofan Yang, Li Li, Ruibo Yao</i>	476
A Research on the Influential Factors of Listing Sales Based on Online Information in Short Rental Markets <i>Hong Peng, Shuang Tao</i>	484
Dual-Channel Supply Chain Network Equilibrium Model with Consumer-Driven <i>Jun Ma, Ding Zhang, Ding Feng, Yiliu Tu</i>	490
The Impacts of Open and Proprietary IT on Vertical Firm Boundaries <i>Fengfeng Lou, Wen Guang Qu</i>	499
User behavior and personalized service	
Impacts of Psychological Distance-based Sales Promotion on Online Purchasing Behaviors under Different Involvement <i>Liyi Zhang, Shan Huang, Qihua Liu</i>	507
Privacy as a Commodity Is Not the Case: Privacy Calculus Model for Connected Cars <i>Minho Ju, Jian Mou</i>	515
Considering correlation retarded growth for personalized recommendation in social tagging <i>Xuwei Pan, Ling Ding, Leimin Ye</i>	523
Will the Recommendation Information be Sure to Improve the Brand Evaluation: The Impact of Distrust <i>Xinyan Liu, Cheng Hu</i>	531
A Personalization-Privacy Paradox in Usage of Mobile Health Services: A Game Theoretic Perspective <i>Fanbo Meng, Xitong Guo, Kee-hung Lai, Xinli Zhao</i>	539

Understanding continuance intention to use travel review websites <i>Ping Wang, Hongxiu Li, Yong Liu</i>	546
Gender Differences in Selective Attention and Shopping Intention in the case of Taobao Live-show: An Eye-Tracking Study <i>Qing Xu, Mengqi Fei, Huizhong Tan</i>	555
Signaling Peer Trust in Accommodation-sharing Services: Effects of Similarity and Reviews on Listing Sales <i>Jiang Wu, Mengmeng Jin</i>	564
General	
The Impact of Board Capital on Performance of China's Listed Tourism Companies Based on the Moderating Effect of Leadership Structure <i>Li Muchun, Xu Tianqi, Liu Tingting</i>	572
The Business Talents Cultivation of Cross-border E-commerce under “the Belt and Road Initiative” <i>Ai Xu, Shi Tang, Yijia Gao, Zongqing Zhou</i>	583
Study on Social Network for College Students' Job Hunting <i>Li Li, Zhenyan Hu, Huiming Xiao, Meilin Chen, Yitang Zeng</i>	589

The Influencing Factors Model of Cross-Border E-commerce Development:

A Theoretical Analysis^{*}

Li Junbo¹, Zhang Jue¹, Qu Fei, Zhao¹ Yuehan¹

¹School of Business, Guilin University of Technology, Guilin, 541004, China

Abstract: Cross-border e-commerce (CBEC) is the future trend of cross-border trade. Although China is at the forefront of CBEC development, its transaction volume is still not satisfactory. The purpose of this paper is to study factors influencing the development of CBEC industry from the macro-environment perspective. First, we commented and summarized relevant literature at home and abroad about business ecosystem and factors determining the development of CBEC, then proposed a model of factors influencing CBEC development by combining business ecosystem theory with PEST framework, followed by interpretation and discussion. The model consists of core species, key species, supporting species, parasitic species in the CBEC ecosystem, and they are affected by external environmental factors from political, economic, social and technological perspectives.

Keywords: cross-border e-commerce, influencing factors, e-commerce ecosystem, PEST

1. INTRODUCTION

Cross-border e-commerce (CBEC) is regarded as a new type of trade combined with electronic commerce (e-commerce) and cross-border trade. With the development of the Internet, information technology and social economy, cross-border e-commerce is developing rapidly. Amazon, eBay, Made-in-China, and Alibaba are the early explorers of CBEC. However, Chen & Yang^[1] argued that there was no leading company in CBEC in the current application situation.

In recent years, China has strongly advocated the “Internet Plus” plan of action and positively pursued the Belt and Road Initiative to encourage “Encourage people to do business creatively and drive innovation”. Accordingly, investment in CBEC is flourishing and booming, and all kinds of traditional e-commerce businesses, foreign trade enterprises, express delivery companies are taking active participation. As a new way of trade, CBEC can promote the development of small and medium-sized enterprises (SMEs), improve the level of opening up, encourage the optimization and upgrading of the industrial structure, give full play of female talent, and benefit the economic development in poor areas practically^[2]. Cross-border B2B e-commerce is still the mainstream of business mode, but cross-border B2C, C2C and O2O e-commerce have gradually grown.

Cross-border e-commerce can cover a wide range of global markets, as Lewis & Cockrill^[3] suggested that where there is Internet, there is within the online commercial boundaries. According to the PayPal Cross-Border Consumer Research 2016, China is the most popular cross-border online shopping site for online shoppers worldwide. According to China E-Commerce Report 2016 by People's Republic of China Ministry of Commerce, China's total import and export trade volume in 2016 was 24.33 trillion yuan with a year-on-year decrease of 0.9 percent, and in the meanwhile, the domestic e-commerce transaction volume in China was 26.1 trillion yuan with a year-on-year growth of 19.8 percent. And as reported by China E-Commerce Research Center's 2016 E-Commerce Market Data Monitoring Report, China's cross-border e-commerce volume in 2016 was 6.7 trillion yuan, up 24 percent. In contrast to these three data, there is a big gap between cross-border e-commerce transactions and international trade volume, national e-commerce transactions, so that we assumed that the development space of cross-border e-commerce in China is really huge.

However, the reality is that although there are constant practice attempts and theoretical studies in the field of cross-border e-commerce in China, the development of cross-border e-commerce still lags behind that of traditional e-commerce. As a result, the following three questions arise: What are the participants in CBEC? What factors affect CBEC development? And which is drivers or impediments of CBEC? Despite recent progress in CBEC research, what is main factors influencing CBEC remains to be deeply explored. We aim to comb main literatures to identify the factors model influencing CBEC and pave the way for its further development for future research. A brief summary of some of the relevant literature in business ecosystem and influencing factors of CBEC is presented in Section 2. Section 3 gives a concrete factors model and explains its details on factors of CBEC development. Finally, Section 4 concludes the paper and suggests directions for future research.

2. LITERATURE REVIEW

2.1 E-COMMERCE ECOSYSTEM

Ecosystem theory was first introduced in natural studies, by Tansley (1935). Then Moore ^[4] applied it to business, and proposed business ecosystem. That is, there was a kind of dynamic competition symbiotic relationship among ecological members, whose sustainable competition advantage coming from innovation. Ding & Wu ^[5] discussed the relationship between core enterprise and business ecosystems through case study, and proposed two kinds of symbiotic strategy model: mutualism symbiosis and predation symbiosis. Similarly, we speculate that in the context of cross-border e-commerce, there will be a symbiotic ecosystem with similar dynamic balance.

In recent years, business ecosystem theory has been introduced into the field of e-commerce and an e-commerce ecosystem has been proposed. The e-commerce ecosystem is a new product of the integration of traditional business and Internet technology. It specifically refers to a series of closely-related organizations and individuals, transcending the constraints of time and space, using the Internet as a platform for competition and communication to complement each other and share resources ^{[6]-[7]}. The development of core e-commerce companies, the need for self-propagation and evolution of various species, the joining of favorable policies, and the attraction of a large number of value-added service providers have prompted e-commerce to evolve from a single site to a multi-species e-commerce ecosystem ^[6]. The characteristics of e-business ecosystem are (1) high system renewal, (2) absolute leadership of the core enterprise, (3) ambiguous system boundaries, (4) threatened by external environment. Scholars at home and abroad have conducted exploratory research, and a few studies have introduced network structures for quantitative analysis. Hu ^[6] examined members of e-commerce ecosystem, pointed out that e-commerce ecosystem is composed of leading species, key species, supporting species and parasitic species, and lastly divided the evolution process of e-commerce ecosystem into four phases: development, expansion, coordination and evolution. Combined with the complex networks theory, Gao et al. ^[7] analyzed e-commerce ecosystem's network structure and evolution process. Taking two e-commerce villages in China as an example, Leong et al. ^[8] studied the development and obstacles of the rural e-commerce ecosystem.

2.2 CBEC DEVELOPMENT INFLUENCING FACTORS

A considerable number of studies have been carried out to examine the factors affecting the development of cross-border e-commerce from the perspective of macro environment and the basic process of realizing CBEC. To sum up the literature, we divide the factors that affect the development of CBEC into three levels: macro, meso and micro (Table 1, 2 and 3). The macro perspective is the main way to analyze the factors at the national level, while mesoscopic views are mainly about business and industry perspectives, and microscopic aspects are

conducted from the viewpoint of consumers, especially some psychological factors. As shown in Table 1, a large amount of studies use PEST framework to analyze the national environment. The research on the mid-point of view is to study the main elements of the development of CBEC, such as cross-border logistics, customs clearance, cross-border payment, international marketing^[15]. At the micro level, trust is the most important research topic, which is also a classic research direction in e-commerce.

Table 1. Factors influencing the development of CBEC in macro perspectives

Author	Research Theme	Influencing Factors to CBEC
Farhoomand et al.	Barriers to GEC	PESTEL
Bingi et al.	GEC challenges	PEST
Javalgi & Ramsey	GEC	Political, legal, commercial, social and cultural, tech infrastructure
Gibbs et al.	GEC diffusion	Global, national environment, national policy
XiaoboXu	EC adoption	Information infrastructure and demographics
Cyr et al.	E-readiness	National cultural values and corruption
Zhu et al.	GEC adoption	Policy, legal, social, cultural environment, GDP per capita
Kshetri et al.	E-retailing	Broadband penetration, retail spending trends, traditional alternatives, economic freedom, GDP per capita, e-retail spending
Liu et al.	CBEC development	System, technology, market environment
Kawa & Zdrenka	Integrator in CBEC	Transportation costs, delivery time & quality, foreign language, payment currency & terms, free trade agreements, returns and standardization
Boyd et al.	Obstacles to international EC	Language, currency differences, customs and inspection fees, customs agreements, access
Kim et al.	CBEC demand	Geographical distance, delivery time and shipping costs, subjective distance, institutional distance
Cho & Lee	Overseas direct purchase	Logistics connectivity, customs efficiency, regulatory quality and globalization

Table 2. Factors influencing the development of CBEC in meso perspectives

Author	Research Theme	Influencing Factors to CBEC
Lewis & Cockrill	E-readiness	Financial resources, staff support and skills
Bingi et al.	GEC challenges	Access to tech, availability of appropriate skills, organizational core competencies, top management commitment
ZHAO et al.	CBEC Capability of manufacturing SMEs	Internet marketing, cross-border logistics, cross-border payments, electronic customs clearance
Gomez-Herrera et al.	CBEC trade flow	Distance, transportation costs, language, borders, legal system or colonial context, quality of governance
Liu et al.	CBEC development	Industry competition and cooperation, cross-border realization of intermediaries, business capabilities
Wang et al.	EC international logistics performance	Laws and regulations, cross-border payments, electronic clearance and technology application level
Deng et al.	Export sales	Firm tenure, product price; product diversity
Ai et al.	Cross-border logistics performance	Cross-border marketing, cross-border payments, customs clearance, laws and regulations

Table 3. Factors influencing the development of CBEC in micro perspectives

Author	Research Theme	Influencing Factors to CBEC
Boyd et al.	Obstacles to international EC	Consumer responsibility impact
Cyr et al.	E-readiness	Trust and website availability
Chen et al.	Customer loyalty in B2B EC	System, information, service, process and collaboration quality

Notes: GEC is on behalf of global e-commerce. PESTEL is on the basis of PEST, plus environmental and legal factors. CBEC is on behalf of cross-border e-commerce. SMEs is on half of small and medium-enterprises.

3. INFLUENCING FACTORS MODEL OF CBEC DEVELOPMENT AND INTERPRETATION

To investigate the factors of cross-border e-commerce development, some existing literatures on the influencing factors of CBEC were reviewed. The main purpose of this paper is to promote the CBEC development in China from the perspective of strategic management. PEST is an effective theoretical framework for strategic management research and a widely-used tool in business strategy analysis, which is suitable for macro-environmental analysis. In order to explore all the members of the cross-border e-commerce industry and its influencing factors, the theoretical model was developed to the PEST framework and e-commerce ecosystem theory in Figure 1 to solve the three research questions mentioned in Section 1.

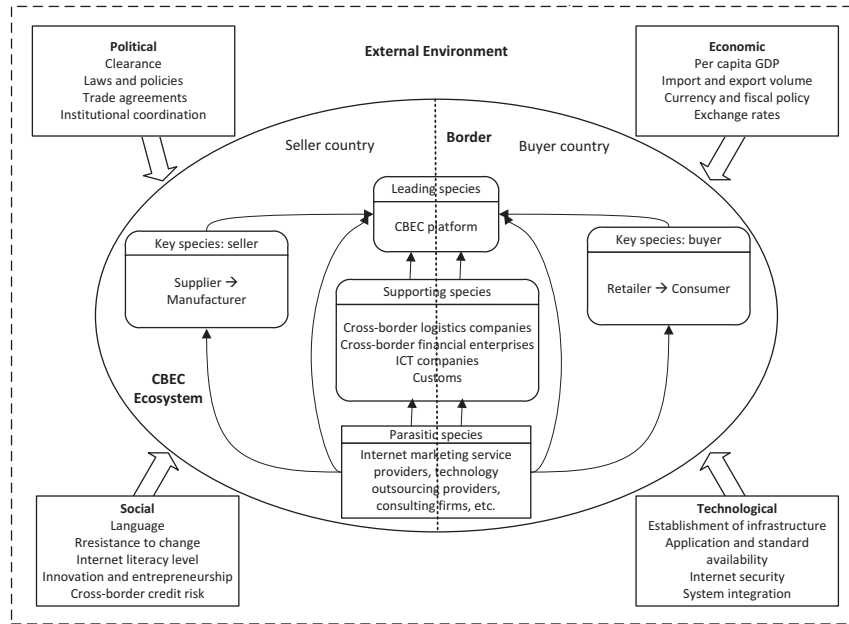


Figure 1. Theoretical model

3.1 CBEC ECOSYSTEM

Cross-border e-commerce is the result of the integration of e-commerce and international trade. Similar to the e-business ecosystem [6], there are also leading species, key species, supporting species and parasitic species in the CBEC ecosystem.

Leading species of CBEC ecosystem is the CBEC platform that connects the cross-border transactions of key species, provides sellers with a broad market, and buyers with product, payment, and logistics information. It is responsible for the allocation of resources of the entire system and at the core of the ecosystem. It can be B2B, B2C or C2C cross-border e-commerce platforms, such as Amazon, eBay or DHgate, etc. The birth of a CBEC platform marks the advent of a cross-border e-commerce era. The leading species is at the core of the ecosystem, and communicates information among species. Alibaba is a relatively typical example. As Jack Ma, the founder of Alibaba Group, said: "We are not operating a company, but an ecosystem, a new economy that has been built up with new technologies and new ideas and is still growing and evolving by hundreds of millions of consumers, retailers, manufacturers, service providers and investors all over the world."

Key species are companies or individuals involved in cross-border trade through CBEC platforms, i.e. buyers and sellers of cross-border online shopping, such as foreign trade companies and their suppliers, multi-national buyers and consumers, etc. Buyers have demand pulling power and sellers have market driving force. Both of them are the power source for CBEC ecosystem. They demand for lower prices, better services, more products, and more convenient purchases. The key species of CBEC come from domestic e-commerce and international trade, but some of them are not good at CBEC and need extra services for parasitic species and

shopping guidance for leading species. The development of CBEC provides development opportunities for the majority of SMEs, but also provides consumers with more shopping options.

Supporting species are organizations, responsible for realizing logistics and capital flow between buyers and sellers, such as cross-border logistics companies (including international express delivery), cross-border payment companies and customs, etc. Supporting populations existed before the emergence of cross-border e-commerce, but these organizations have evolved in the context of cross-border e-commerce. Cross-border logistics companies actively implement order tracking services. Besides, cross-border payment companies have realized the online flow of funds. And customs is actively implementing electronic customs clearance to achieve trade facilitation. They are crucial to whether cross-border e-commerce can be realized and whether it can be realized quickly, which is an important factor in determining the cross-border shopping experience of consumers.

Parasitic species are organizations that provide comprehensive support and value-added service to other species, such as translation firms, certification bodies, consulting firms, marketing firms and training companies, etc., which can facilitate the implementation of CBEC effectively. Parasitic species exist with the proliferation of other populations, and they are huge in number and variety. The CBEC platform has gathered all kinds of businesses, how to make sellers' products or services stand out among a large number of competitors, how do buyers find their own satisfied products in the vast amount of products and how does the platform expand its influence? They all lead to the emergence of parasitic species. A lot of value-added service providers provide various convenient and efficient third-party services for other species. It enhances consumers' cross-border shopping satisfaction.

Each species has its own specialization and plays an indispensable role in an effective cross-border e-commerce process. At the same time, there are also companies infiltrating each other and invading other species. In the Internet age, "cross-border" competition has become a norm. Any industry is at risk of being eroded or even replaced at any time. An innovation may subvert a whole industry. This potential threat, once translated into reality, means the subversion of the business model and the reshuffling of the market landscape [29]. However, there is no full-fledged leader in CBEC ecosystem at present. Alibaba, Amazon, Paypal and other companies are in full swing to open up the global market. The competition within CBEC ecosystem members still continues. For example, platform companies develop cross-border logistics and payment services, and cross-border logistics companies develop cross-border e-commerce services. Within the species, each company also has its own core competitiveness and can complement each other, but there is more intense peer competition. For example, C2C vertical electricity supplier, Meilishuo and Social electricity provider, Mogujie make a merge. The ecological members of CBEC closely relate to each other, evolve together, compete with each other and promote jointly.

3.2 EXTERNAL ENVIRONMENT OF CBEC

Through the summary of factors affecting CBEC in Section 2 and with reference to the importance order noted by Farhoomand et al. [9] from the technical, economic, political, social aspects, we make a summary about factors determining the development of CBEC, which is the causal condition proposed by Liu et al. [14].

Technically, the influencing factors of CBEC development include the establishment of infrastructure, the availability of applications and standards, Internet security and system integration. The establishment of infrastructure is a vital indicator of e-readiness [2]. The technology advancement of CBEC logistics, payment, information communication and data analysis can greatly improve the speed, safety, effectiveness and convenience of CBEC and indeed reduce a lot of operating costs [27]. Wang et al. [25] believed that the level of electronic clearance and application of technology was a mediating factor affecting the performance of

international logistics. According to the PayPal Cross-Border Consumer Research 2016, free postage and the security of cross-border payment are main factors influencing cross-border online shopping, which shows great importance of cross-border logistics and payment. Trinity of service: information flow, capital flow and logistics, provides great support to cross-border e-commerce development. Nevertheless, different countries have different standards, so it is difficult to integrate the system and quite insufficiency for comprehensive information service platform.

In terms of economy, factors affecting CBEC are GDP per capita, import and export volume, monetary and fiscal policies^[9], and exchange rates^[9]. GDP^[8] and income^[18] are considered to be control variables, and the demographics is always adopted together with GDP and income as GDP per capita^{[6]-[14]}, by which most of economic indicators are measured^{[9]-[7]}. It is generally believed that cross-border e-commerce is a new form of foreign trade transformation and upgrading. It is inferred that import and export volume, monetary and fiscal policies and exchange rates, which are the most important indicators in international trade, will also affect CBEC development. In the previous work, Farhoomand et al.^[9] noted the impact of monetary, fiscal policies and exchange rates on the development of cross-border e-commerce.

Politically, the factors influencing the development of CBEC include customs clearance, laws and policies, trade agreements and institutional coordination. Customs clearance is a critical part of the basic flow of CBEC. Its process and technology absolutely influence the speed of customs clearance and ultimately affect the realization of cross-border logistics^[27]. Zhao & Yang^[19] believed that electronic customs clearance was an essential factor influencing the CBEC capability of small and medium-sized manufacturing enterprises. Laws and policies have always been influential factors in macro environment analysis, and a major measurement for e-readiness and e-commerce development^[2]. According to laws and policies, Customs, Industry and Commerce, Tax and other departments carry out CBEC regulatory activities to maintain interests of market participants and promote the development of CBEC^[25]. Directly or indirectly, laws and policies affect CBEC logistics, payment, customs clearance and other activities in CBEC. Institutional theory^[18] and two market theories give full consideration to the institutional distance between two countries and legal differences between online and offline transactions, which can greatly affect the development of CBEC^{[25]-[28]}. Further, Zhu et al.^[12] studied specific indicators to measure the policy environment^[10] and the legal environment. Government preferential policy is the facilitating factor for the development of CBEC. Chen et al.^[1] also studied the supportive policies of the government, and found that as a business innovation, CBEC played an intermediary role among supporting innovation policies and business performance, especially its institutional innovation and business model innovation. The trade agreement, mainly concerned with tariffs and legal issues, is a factor to be considered in traditional international trade and also an essential element affecting the development of CBEC^[16]. Institutional coordination^[4], involving the law^{[12]-[27]}, taxation, regulation and other aspects, is not only about the coordination within the country itself, but also about the coordination among the same agencies in different countries.

Socially speaking, influencing factors include language, resistance to changes, Internet literacy level, innovation and entrepreneurship and cross-border credit risk. Linguistic differences are considered to be great obstacles to the development of CBEC^{[9]-[6]-[12]-[28]}. Countries that use English as their language have an enormous advantage in this respect, and four-fifths of Hong Kong respondents in an interview conducted by Farhoomand et al.^[9] supposed that boycotting changes is one of the most powerful obstacles to global economic cooperation. Furthermore, Gomez-Herrera et al.^[24] found that the impact of language are significant in their research. The level of Internet literacy is represented by literacy levels, IT skills levels and education levels^{[6]-[8]}. The country's attitude towards innovation and entrepreneurship affect the development of CBEC^[12], and innovation and entrepreneurship are relevant to the performance of CBEC enterprises^[1]. Cross-border credit

risk^[11] exists in traditional international trade, but in the context of CBEC, cheating costs are so lower that the credit risk could be greater. Cultural factors such as language difference, resistance to changes and cross-border credit risk can be recognized as huge obstacles or impediments to the development of CBEC.

4. CONCLUSIONS

In this paper, after analyzing the cross-border e-commerce ecosystem, and expounding in detail the external environment such as politics, economy, culture and technology that affect the development of CBEC, we propose a model of influencing factors of cross-border e-commerce. Through promoting factors such as preferential policies and other promotional efforts to push the development of cross-border e-commerce, and striving to overcome such obstacles as language, resistance to changes and cross-border credit risk, we strongly advise as many government departments as possible to make joint effort to change the external environment. It does good to attract new or former marginal business ecosystem members^[4].

This paper does not validate the model. In the future research, we recommend that researchers could consider conducting quantitative research to explore the dimensions raised in the model. We feel that the available methods are econometric models, multivariate statistical models or structural equation models. Empirical research on influencing factors is of great value to the development of CBEC for both theoretical research and social practice.

Future work can also take a comparative study of multiple markets to understand the role of cultural differences in the development of cross-border e-commerce. Most studies on the development of CBEC in China are limited to the domestic scope and lack of research across multiple countries.

ACKNOWLEDGEMENT

This research was supported by the National Social Science Foundation of China under Grant 16XGL014.

REFERENCES

- [1] Chen, N, & Yang, J. (2017). Mechanism of government policies in cross-border e-commerce on firm performance and implications on m-commerce. *International Journal of Mobile Communications*, 15(1):69.
- [2] Oreku, G S, Li Jianzhong, Kimeli Kimutai, & Mtenzi, F J. (2009). State of Tanzania e-readiness and e-commerce: overview. *Information Technology for Development*, 15(4):302-311.
- [3] Lewis R, Cockrill A. (2002). Going global-remaining local: the impact of e-commerce on small retail firms in Wales. *International Journal of Information Management*. 22(3):195 - 209
- [4] Moore J F. (1993). Predators and prey: a new ecology of competition. *Harvard business review*, 71(3):75.
- [5] Ding Ling, Wu Jinxi. (2017). Case Study of Core Enterprise and Business Ecosystems: Strategies of Mutualism Symbiosis and Predation Symbiosis. *Management Review*, 29(7):244-257 (in Chinese)
- [6] HU Gang-lan, LU Xiang-hua, HUANG Li-hua. (2009). An Essay on E-business Ecosystem and Its Evolutionary Path—With Focus on the Phenomenon of E-business Industrial Cluster in China. *Economic Management Journal*, 31 (06): 110-116(in Chinese)
- [7] Gao R, Zhang Z, Tian Z, et al. (2013). Modelling the emergence and evolution of e-business ecosystems from a network perspective. *Studies in Informatics & Control*, 22(4):339-348.
- [8] Leong C, Pan S L, Newell S, et al. (2016). The emergence of self-organizing e-commerce ecosystems in remote villages of china: A tale of digital empowerment for rural development. *MIS Quarterly*, 40(2):475-484.
- [9] Farhoomand A F, Tuunainen V K, Yee L W. (2000). Barriers to Global Electronic Commerce: A Cross-Country Study of Hong Kong and Finland. *Journal of Organizational Computing & Electronic Commerce*, 10(1):23-48.

- [10] Gibbs J, Kraemer K L, Dedrick J. (2003). Environment and Policy Factors Shaping Global E-Commerce Diffusion: A Cross-Country Comparison. *Information Society*, 19(1):5-18.
- [11] Xu Xiaobo. (2008). Effects of two national environmental factors on e-commerce functionality adoption: a cross-country case study of a global bank. *Enterprise Information Systems*, 2(3):325-339.
- [12] Zhu L, Thatcher S M B. (2010) National information ecology: A new institutional economics perspective on global e-commerce adoption. *Journal of Electronic Commerce Research*, 11(1):53-72.
- [13] Kshetri N, Bebenroth R, Williamson N C, et al. (2014). Cross-national heterogeneity in e-retail spending: a longitudinal analysis of economic, technological and political forces. *Electronic Commerce Research*, 14(4):1-25.
- [14] LIU Jia-guo, LIU Wei, LIU Xiao-qi, ZHAO Jin-lou. (2015). Research of Sino-Russian Cross-border E-commerce Development Based on Grounded Theory Method. (09):27-40 (in Chinese)
- [15] Kawa A, Zdrenka W. (2016) .CONCEPTION OF INTEGRATOR IN CROSS-BORDER E-COMMERCE, 12(121):63-73.
- [16] Boyd S L, Hobbs J E, Kerr W A. (2003). The impact of customs procedures on business to consumer e-commerce in food products. *Supply Chain Management*, 8(3):195-200.
- [17] Kim T Y, Dekker R, Heij C. (2017). Cross-Border Electronic Commerce: Distance Effects and Express Delivery in European Union Markets. *International Journal of Electronic Commerce*, 21(2):184-218.
- [18] Cho H, Lee J. (2017). Searching for Logistics and Regulatory Determinants Affecting Overseas Direct Purchase: An Empirical Cross-National Study. *Asian Journal of Shipping & Logistics*, 33(1):11-18.
- [19] ZHAO Zhi-tian, YANG Jian-zheng. (2014). Capability of Small and Medium-sized Manufacturing Enterprises Cross-border E-commerce: Identification, Validation and Comprehensive Evaluation. *Systems Engineering*, 32 (10): 53-62(in Chinese)
- [20] Bingi P, Mir A, Khamalah J N. (2000). The Challenges Facing Global E-Commerce. *Information Systems Management*, 17(4):22-30.
- [21] Javalgi R, Ramsey R. (2001). Strategic issues of e-commerce as an alternative global distribution system. *International Marketing Review*, 18(4):376-391.
- [22] Cyr D, Berthon P, Campbell C, et al. (2008). E-readiness and trust: macro and micro dualities for e-commerce in a global environment. *International Marketing Review*, 25(6):700-714(15).
- [23] Chen J V, Chen Y, Capistrano E P S. (2013). Process quality and collaboration quality on B2B e-commerce. *Industrial Management & Data Systems*, 113(6):908-926.
- [24] Gomez-Herrera E, Martens B, Turlea G. (2014). The drivers and impediments for cross-border e-commerce in the EU. *Information Economics & Policy*, 28(1):83-96.
- [25] Wang L, Yang J, Yin S. (2015). Electronic commerce international logistics performance influence factor analysis. *International Journal of Mobile Communications*, 13(5): 498-509.
- [26] Deng Z, Wang Z. (2016). Early-mover advantages at cross-border business-to-business e-commerce portals. *Journal of Business Research*, 69(12):6002-6011.
- [27] Ai W, Yang J, Wang L. (2016). Revelation of cross-border logistics performance for the manufacturing industry development. *International Journal of Mobile Communications*, 14(6):593.
- [28] Hegemann L, Fruchtmann R, Bonnekoh B, et al. (2010). E-Commerce on the Global Platform: Strategic Insights on the Localization-Standardization Perspective. *Journal of Electronic Commerce Research*, 11(1):6-13.
- [29] Kou Zong-lai. (2018). Financial Times Chinese. www.ftchinese.com/story/001076723,

New Weather Indices for China: Tool of Risk Control of International Supply Chain

Qing Zhu^{12}, Jiarui Li¹, Jian Chai³*

¹ International Business School, Shaanxi Normal University, Xi'an, 710000, China

² Institute of Cross-Process Perception and Control, Shaanxi Normal University, Xi'an, 710000, China

³ Jian Chai, Management and economic School, Xidian University, Xi'an, 710000, China

Abstract: China is at the core of the world's supply chain because of its focus on production and consumption. However, as weather can significantly affect supply chain operations, China plans to introduce weather derivatives to secure the multinational supply chain. Using historical records over the decade, weather derivatives could be an important tool for hedging risk and meeting the needs of Chinese market. In this paper, new weather indices for China financial markets are experimentally created through simulated machine learning to assess the ability of the weather indices to reduce risk. Through a simulation test from 2008 to 2017, the indices were found to successfully match 98% of the risk with the situation across two dimensions: *i*). changing Chinese weather data; and *ii*). a connection with US weather indices.

Keywords: ANNs, C-CDDs, DCC-GARCH model, weather derivatives

1. INTRODUCTION

While weather is generally predictable, its effects are often random. Therefore, the weather plays a significant role in production, especially for the energy industry ^[1] and other industries greatly affected by weather changes. Studies have shown that weather risks directly led to losses of around \$1 trillion in the US economy ^[2, 3]. Many enterprises also face dramatic changes in sales volumes because of weather variations, which can significantly affect viability and even hinder industrial development ^[4, 5].

As risk management is a vital part of every enterprise ^[6], it was inevitable that weather derivatives would become a trend. *Su (2010)* ^[7] stated that weather derivatives could be used to hedge the risk of extreme or prolonged weather conditions ^[8]. The first weather derivatives appeared in the US in 1997, and in 1998, the UK, Germany, Belgium, and Norway also introduced weather derivatives to hedge the risk of weather damage. In 1997, the trading volume was \$500 million, which rose to \$1 billion in 1998 and to \$3 billion in 1999, proving their viability in financial markets.

In the early years, weather derivatives were only traded in US over-the-counter (OTC) markets. However, in January, 2000, the London International Financial Futures Exchange (LIFFE) in England began to trade these derivatives online ^[9]. Exchange markets now include exchanges such as the Chicago Merchandise Exchange (CME) and the International Exchange in Atlanta. The types of weather derivative futures and options available include Cooling Degree Days (CDDs), Heating Degree Days (HDDs), and Cumulative Average Temperatures (CATs), with many corporations choosing to hedge risk for the best results.

Currently, the US, the UK, Japan, Australia, and some cities in Europe have introduced and are trading weather derivatives; however, China has not yet built a standardized weather derivatives trading platform. As China is at the core of the global supply chain, adverse weather conditions inevitably influence the development of the global economy; therefore, the risks caused brought by adverse weather need to be well-managed to ensure a stable global supply chain and reduce the negative influences. China also has a large well developed

* Corresponding author. Email: zhuqing@snnu.edu.cn

agricultural sector ^[10] that contributes significantly to GDP every year; however, production can be easily influenced by adverse weather and climate change ^[11]. Therefore, the introduction of weather derivatives is vital for securing the global supply chain and developing the domestic economy. If an enterprise chooses to purchase Chinese weather derivatives, they could better hedge their risks and maintain profitability, which would be beneficial to the development of the global economy and the international financial markets. Weather derivatives allow for the risks associated with the energy industry and agriculture to be well managed, and can ensure the global supply chain and international financial market stability.

Although China has not launched weather derivatives on the financial markets, there has been some associated weather derivative research. For example, *Hong et. al (2013)* ^[9] proposed a method using peer group analysis to set a prior price for new weather derivatives, and *Zong et. al (2016)* ^[12] proposed regional weather indices for China and demonstrated that weather derivatives were a practicable tool for efficiently hedging risk. Weather derivatives are fundamentally dependent on temperatures; according to data from the Chinese Government Network (2014), the 665 cities in mainland China had contract difficulties because of difficulties in obtaining temperature data ^[12].

As contract prices are based on monthly and quarterly cumulative weather indices, the fluctuations in the weather indices are very important. Therefore, in this article, more attention is paid to the design of Chinese weather indices and new feasible and stable Chinese weather indices proposed. Being different with the weather indices of other cities, the new weather indices we propose is not only decided by local weather but influenced by global weather derivatives market. Therefore, the new weather indices will have much more possibility to suit the global market and improve its viability and adaptation.

In this article, new weather indices for Chinese weather are proposed using a DCC-GARCH model and ANNs. In Section 2, the models used are introduced, in Section 3 the city selection process is described, and a new equation for the weather indices proposed. A simulation and discussion are presented in Section 4, and the conclusion is given in Section 5.

2. METHODOLOGY

2.1 Dynamic conditional correlations - Generalized autoregressive conditional heteroscedasticity

The dynamic conditional correlation-Generalized autoregressive conditional heteroscedasticity (DCC-GARCH) model is based on the Generalized autoregressive conditional heteroscedasticity (GARCH) model proposed by *Bollerslev (1986)* ^[13] and the generalized ARCH model by *Engle (1982)* ^[14]. The GARCH(p, q) model is described as follows:

$$\varepsilon_t = \sigma_t z_t \quad (1)$$

$$\sigma_t^2 = a_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (2)$$

where z_t represents the uniform and independent random variables, and σ_t denotes the conditional variances. The parameter p and q represent the order for the ARCH and GARCH models; when $p = 0$, it is considered an ARCH(q) model.

Weather derivatives are monthly or quarterly contracts based on an index ^[4]. As Chinese financial markets have been gradually integrated into the global financial system ^[15], and it is hoped that these new weather indices can be integrated into international weather derivatives markets, the new weather indices need to be correlated tightly with weather indices in other cities to ensure stability and feasibility. However, because the GARCH model was unable to represent the co-movement of two indices, we can hardly make a connection

between Chinese market and global markets. Therefore, in this article, we choose the DCC-GARCH model to calculate the dynamic conditional coefficients to make a tightly connection with markets trading weather derivatives and make sure the viability and adaptation of the new weather indices we proposed. In 2002, Engle (2002) ^[16] proposed the Dynamic conditional correlation-Generalized autoregressive conditional heteroscedasticity (DCC-GARCH) model to calculate the co-movement of two markets. In Engle's model:

$$H_t = D_t R_t D_t \quad (3)$$

where R_t is a $n \times n$ correlation matrix.

$$D_t = \text{diag}\{\sqrt{h_{ii,t}}\} \quad i = 1, 2, \dots, n \quad (4)$$

and

$$h_{ii,t} = \alpha_{i0} + \alpha_{i1}\varepsilon_{i,t-1}^2 + \alpha_{i2}J_{i,t-1}\varepsilon_{i,t-1}^2 + \beta_{i1}h_{ii,t-1} \quad (5)$$

where $J_{i,t-1} = 1$ if $\varepsilon_{i,t-1} < 0$, else $J_{i,t-1} = 0$.

$$Q_t = (1-a-b)S + au_{t-1}u_{t-1}' + bQ_{t-1} \quad (6)$$

where $u_t = \varepsilon_t / D_t$, the conditional correlation matrix of ε_t is derived from $R_t = E(u_t u_t' | I_{t-1})$, and S is a matrix of the location parameters. Another conditional correlation matrix R_t can be defined as

$$R_t = \text{diag}(\sqrt{q_{11,t}}, \dots, \sqrt{q_{nn,t}}) Q_t \text{diag}(\sqrt{q_{11,t}}, \dots, \sqrt{q_{nn,t}}) \quad (7)$$

The elements of R_t can be expressed as

$$\rho_{ij,t} = q_{ij,t} / \sqrt{q_{ii,t} q_{jj,t}} \quad (8)$$

which equals

$$\rho_{ij,t} = [(1-a-b)\bar{q}_{ij} + bq_{ij,t-1} + au_{i,t-1}u_{j,t-1}] \times \sqrt{[(1-a-b)\bar{q}_{ii} + bq_{ii,t-1} + au_{i,t-1}^2][(1-a-b)\bar{q}_{jj} + bq_{jj,t-1} + au_{j,t-1}^2]} \quad (9)$$

The dynamic correlation coefficients are nonlinear functions of the two parameters a and b from the DCC model ^[17]. In this article, the DCC-GARCH model is used to estimate the volatility of data and ensure the co-movement of the new weather indices with existing weather indices, in order to improve the feasibility and stability of the proposed weather indices.

2.2 Artificial neural networks(ANNs)

During data classification and prediction, ANNs simulate the learning processes of human brains ^[18] and are composed of three parts; an input layer, hidden layers, and an output layer. Each neuron has an activate function, with variables distributed to different neurons in the hidden layers based on different weights. The ANNs structure is shown in Figure 1.

The ANN is a nonlinear model that can efficiently learn data characteristics and encapsulate time dependency ^[19]. As weather contract prices are based on monthly or quarterly cumulative indices, the contract price can be set based on the proposed weather derivatives; therefore, the feasibility and stability of the weather indices are very important when introducing weather derivatives and developing weather indices. In recent years, there have been some methods proposed to research and predict weather derivative indices. Zapranis et. al(2009) ^[19] proposed a model that demonstrated that neural networks were better able to approximate any nonlinear process. Therefore, based on the characteristics of the weather derivatives indices, ANNs were selected to

simulate the proposed indices. Using machine learning, the stability and feasibility of the new weather indices were tested, the details of which are in Section 4.

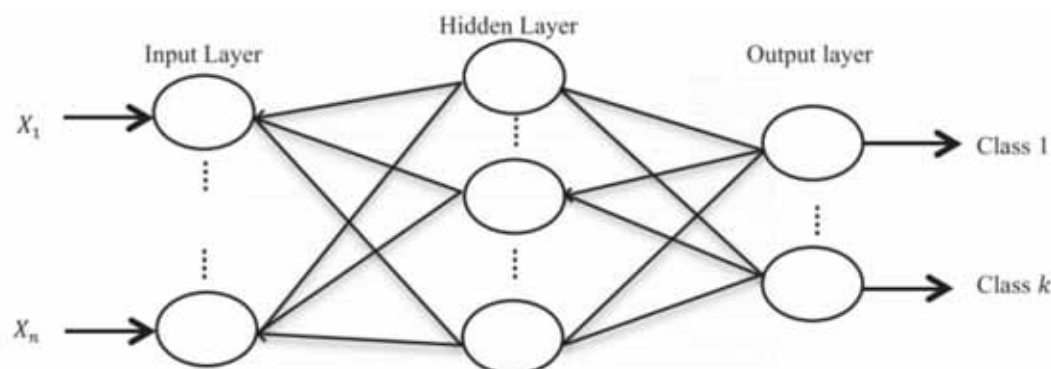


Figure 1. A three-layered artificial neural network

3. DATA SELECTION AND MODELING

3.1 Data selection

3.1.1 Selection of Chinese cities

As the new Chinese weather indices were developed based on existing CDDs, they could be the basis for the Chinese weather derivative development. Because of China's size, it is not possible to have weather derivatives for the whole country; therefore, it is necessary to develop regional Chinese weather indices. There are four dominant futures trading markets in China; Dalian, Shanghai, Shenzhen and Zhengzhou, of which Shanghai is the most important as it is the economic and financial center of China and an important harbor trading city. Therefore, Shanghai and its surrounding area were set as the research region for the design of the new weather indices for four main reasons. *i*). Shanghai is the economic and financial center of China, with a GDP of 2.5 trillion RMB in 2015 (from: <http://data.stats.gov.cn/index.htm>), and the cities near Shanghai such as Hangzhou and Nanjing also contribute around 1 trillion RMB GDP each year. This region that encompasses Shanghai, Jiangsu province and Zhejiang province contributes approximately 16.6% to Chinese GDP (from: <http://data.stats.gov.cn/index.htm>). *ii*) Shanghai and its surrounding areas annually face adverse weather and regularly suffer from typhoon and flood damage, which can result in enormous losses, especially for energy enterprises. *iii*) The weather conditions in Shanghai and its surrounding cities are generally stable, which is beneficial for derivatives trading.

Therefore, Shanghai was chosen as the center for the proposed weather derivatives, with the cities surrounding Shanghai selected for the regional temperatures. Therefore, three cities; Shanghai, Hangzhou and Nanjing; were selected for the regional temperatures. The correlation coefficients for the monthly average temperatures over 25 years for Shanghai, Nanjing and Hangzhou were 1, 0.994, 0.995, respectively, indicating a tight correlation between the three cities.

3.1.2 Selection of cities to trade weather derivatives

There are 24 cities in the US, 11 cities in Europe, 6 cities in Canada, 3 cities in Australia, and 3 cities in Japan trading weather contracts. As Australia is located in the southern hemisphere and has different weather conditions to China, geographic conditions in Australia were not considered. The geographic conditions of all other cities except Australia are shown in Table 1. As can be seen, there are similar geographic conditions in China and the US, and they both also have large land masses and similar latitudes. Further, because of the

number of cities trading weather derivatives in the US, the weather derivatives market is mature; therefore, our research was combined with the US weather indices. The main US weather indices are HDDs, which trade from November to April, and CDDs, which trade from April to October. In this article, CDDs were chosen as the basic weather indices, primarily because HDDs can be retrieved from the CDDs ^[20]. CDDs are calculated from the daily average temperatures, the calculation method for which is as follows:

$$CDDs = \sum_{t=T_1}^{T_2} \max[W_t - 18^\circ \text{C}, 0] \quad (10)$$

where T_1 and T_2 denote the beginning and the end of a month or a season, and W_t represents the average value of the maximum and minimum temperatures of the day; therefore, there is a tight correlation between the average temperature values and the CDDs values.

Table 1. Latitude and longitude of main cities

Country	City	Coordinate	City	Coordinate	City	Coordinate
US	Atlanta	33°46'N, 84°25'W	Detroit	42°23'N, 83°05'W	New York	40°44'N, 73°55'W
	Baltimore	39°17'N, 76°37'W	Houston	29°45'N, 95°23'W	Philadelphia	40°N, 75°09'W
	Boston	42°19'N, 71°05'W	Jacksonville	30°2'N, 81°4'W	Portland	45°31'N, 122°39'W
	Chicago	41°53'N, 87°37'W	Kansas	39°02'N, 94°33'W	Raleigh	35°47'N, 78°39'W
	Cincinnati	39°1'N, 84°3'W	Las Vegas	36°1'N, 115°1'W	Sacramento	38°34'N, 121°28'W
	Colorado Springs	38°51'N, 104°47'W	Little Rock	34°44'N, 92°19'W	Salt Lake City	40°46'N, 111°52'W
	Dallas	32°47'N, 96°47'W	Los Angeles	34°05'N, 118°22'W	Tucson	32°13'N, 110°58'W
	Des Moines	41°36'N, 93°38'W	Minneapolis	0°45'N, 93°15'W	Washington	38°53'N, 77°02'W
Europe	Amsterdam	52°21'N, 4°52'E	Barcelona	41°18'N, 2°06'E	Berlin	52°31'N, 13°2'E
	Essen	51°27'N, 7°00'E	London	51°3'N, 0°07'E	Madrid	40°26'N, 3°42'E
	Oslo	59°56'N, 10°41'E	Paris	48°51'N, 2°2'E	Prague	50°05'N, 14°25'E
	Rome	41°52'N, 12°37'E	Stockholm	59°23'N, 18°00'E		
Japan	Hiroshima	34°23'N, 132°27'E	Osaka	34°4'N, 135°30'E	Tokyo	35°41'N, 139°44'E
Canada	Calgary	51°05'N, 114°05'W	Edmonton	53°34'N, 113°25'W	Montreal	45°3'N, 73°35'W
	Toronto	43°4'N, 79°22'W	Vancouver	49°13'N, 123°06'W	Winnipeg	49°53'N, 97°1'W
China	Beijing	39°55'N, 116°23'E	Changchun	43°5'N, 125°2'E	Changsha	28°1'N, 113°E
	Chengdu	30°37'N, 104°06'E	Chongqing	29°31'N, 106°35'E	Foochow	26°01'N, 119°2'E
	Guangzhou	23°10'N, 113°18'E	Guiyang	26°35'N, 106°4'E	Haikou	20°03'N, 110°10'E
	Hangzhou	30°1'N, 120°07'E	Harbin	45°45'N, 126°41'E	Hefei	31°51'N, 117°16'E
	Huhehot	40°48'N, 111°38'E	Hong Kong	22°17'N, 114°08'E	Jinan	36°5'N, 117°E
	Kunming	25°04'N, 102°41'E	Lanzhou	36°01'N, 103°45'E	Lhasa	29°41'N, 91°1'E
	Macao	22°11'N, 113°33'E	Nanchang	28°38'N, 115°56'E	Nanjing	32°03'N, 118°46' E
	Nanning	23°N, 108°E	Shanghai	31°14'N, 121°27'E	Shenyang	41°48'N, 123°25'E
	Shijiazhuang	38°04'N, 114°28'E	Taipei	25°02'N, 121°38'E	Taiyuan	37°5'N, 112°3'E
	Tianjin	39°08'N, 117°12'E	Urumchi	30°35'N, 114°19'E	Xining	36°34'N, 101°49'E
	Xian	34°16'N, 108°54'E	Wuhan	30°35'N, 114°19'E	Yinchuan	38°28'N, 106°13'E
	Zhengzhou	34°35'N, 113°38'E				

Two American cities were chosen that had high daily average temperature correlation coefficients with Shanghai. Therefore, the new weather indices included two parts: regional China and part of the existing US weather indices. The correlation coefficients were calculated for the monthly average temperatures in US cities and Shanghai. The final American cities chosen were Las Vegas and Little Rock as these two cities were found have high monthly average temperature correlation coefficients of 0.9122 and 0.9054.

3.2 Weather indices modeling

In this article, new weather indices are proposed to introduce Chinese weather derivatives. As China is the third largest country in the world and has 665 cities, it would be difficult for Chinese financial markets to introduce weather derivatives for only one city. Therefore, it was decided that the best way to introduce weather derivatives was to select a part of regional China and consider the weather conditions in several vital cities^[11]. In this way, the scope of application for the proposed weather derivatives was broadened. Therefore, a representative city was chosen as the central city and two more cities near the center city were chosen as the regional cities. To ensure a tight correlation with the existing weather derivatives markets, the influence of international markets was also considered, with the two cities with the highest weather condition correlation coefficients with the center cities being chosen. Therefore, the proposed weather index was included a regional section and an existing sections. The equations for the new Chinese Cooling Degree Days indices(C-CDDs) were as follows:

$$C-CDDs = kCDDs_{regional} + (1-k)CDDs_{exsited} \quad (11)$$

$$CDDs_{regional} = \sum_{i=1}^I \max(\alpha_i T_i - 18^\circ \text{C}, 0) \quad (12)$$

$$CDDs_{exsited} = \sum_{j=1}^J \beta_j CDD_j \quad (13)$$

where i denoted the Chinese cities Hangzhou, Nanjing and Shanghai, and j represented the cities which already had weather derivatives; Las Vegas and Little Rock; and k , α_i and β_j were all parameters. Therefore, the equations above were used to develop the new weather indices. The calculation coefficients α_i and β_j for the regional temperatures were defined by:

$$\alpha_i = \frac{\alpha_i}{\sum_{i=1}^3 \alpha_i} \quad (14)$$

where i represented the researched cities, α_i denoted the correlation coefficients between city i and Shanghai, and the number 3 indicated Shanghai, Hangzhou, and Nanjing. The same method was used to calculate the coefficients β for the international CDDs. The coefficients were compared, Las Vegas and Little Rock selected, and the β value calculated using a similar method.

4. PARAMETER SETTLEMENT AND DISCUSSION

4.1 Parameter settlement

Two principles were followed to develop the proposed weather derivatives; i) to connect the Chinese weather derivatives indices with the US weather derivative indices to ensure feasibility of the new Chinese

weather indices and to ensure that there were highly dynamic correlation coefficients between the Chinese weather indices and US city weather indices using the DCC-GARCH model; and *ii*) as these weather indices described the Chinese weather condition, the domestic section needed improving, which required the k value in equation (11) to be improved.

For this research, 10 years of daily average temperatures from April to October were collected for each of cities chosen; Shanghai, Nanjing, Hangzhou as the Chinese cities, and Las Vegas and Little Rock, which had existing weather indices. The DCC-GARCH (I, I) model was used to simulate the CDDs for the Chinese cities and for the existing weather indices. As the daily average temperatures in Las Vegas were found to have the highest linear correlations with Shanghai, Las Vegas was selected as the matching city to analyze the DCC between C-CDDs and CDDs in Las Vegas (CDDs-LV) to set parameter k .

The data characteristics were then set, including the C-CDDs for different k values and the CDDs-LV. The characteristics for these data were found to be similar. As the ADF test found the data to be non-stationary, the 1st difference data was set as the data to be analyzed for the building of the DCC-GARCH model. The ADF test indicated that the 1st difference data was stationary and that the data for each year was almost normal with weekly skewness and kurtosis, and the McLeod Li test showed a significant ARCH effect. Based on these characteristics, the DCC-GARCH model was built to calculate the dynamic conditional correlation coefficients. The DCC (I, I) and the GARCH (I, I) model were both chosen.

It was found that the DCC data fluctuated around an approximate certain number. Therefore, the mean DCC value was calculated as the DCC for each year, the results for which are shown in Table 2.

Table 2. DCC for the different k values in ten years

DCC (mean value)	k value (0 - 1 pause: 0.1)										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
2008	0.72	0.72	0.72	0.71	0.71	0.69	0.66	0.61	0.47	0.27	0.01
2009	0.73	0.73	0.73	0.73	0.71	0.68	0.65	0.58	0.47	0.28	0.04
2010	0.62	0.62	0.62	0.62	0.61	0.60	0.56	0.49	0.37	0.22	0.06
2011	0.57	0.56	0.56	0.55	0.54	0.53	0.48	0.42	0.33	0.21	0.04
2012	0.64	0.65	0.65	0.65	0.64	0.63	0.61	0.55	0.45	0.28	0.09
2013	0.69	0.69	0.69	0.69	0.69	0.68	0.66	0.62	0.51	0.28	-0.02
2014	0.73	0.73	0.73	0.73	0.72	0.70	0.67	0.60	0.47	0.22	-0.10
2015	0.71	0.71	0.71	0.70	0.69	0.67	0.64	0.57	0.46	0.26	0.07
2016	0.73	0.74	0.73	0.73	0.72	0.70	0.64	0.58	0.47	0.30	0.07
2017	0.70	0.70	0.70	0.69	0.67	0.64	0.59	0.50	0.36	0.19	0.07
Mean	0.685	0.684	0.683	0.679	0.670	0.651	0.615	0.553	0.434	0.249	0.033

From Table 2, as k was below 5, this was unable to show the influence of the changes in Chinese temperature, with a k above 0.7 showing a poor dynamic correlation with the US weather indices. Therefore, $k = 0.6$ was set as a suitable value for the proposed principles. To gain a more accurate k value, a further calculation was conducted. Therefore, a k value around from 0.55 to 0.65 was assumed to calculate the DCC value, the results for which are shown in Table 3.

Table 3. DCC for different k values in ten years

DCC (mean value)	k value (0.55 - 0.65 pause: 0.01)										
	0.55	0.56	0.57	0.58	0.59	0.6	0.61	0.62	0.63	0.64	0.65
2008	0.62	0.62	0.61	0.6	0.6	0.59	0.58	0.58	0.57	0.56	0.55
2009	0.68	0.68	0.67	0.67	0.65	0.64	0.64	0.63	0.62	0.62	0.62
2010	0.66	0.65	0.65	0.65	0.64	0.64	0.63	0.63	0.62	0.62	0.61
2011	0.69	0.68	0.68	0.68	0.67	0.67	0.66	0.66	0.65	0.65	0.64
2012	0.67	0.66	0.66	0.66	0.66	0.66	0.65	0.65	0.65	0.64	0.64
2013	0.62	0.62	0.62	0.61	0.61	0.61	0.6	0.6	0.6	0.59	0.59
2014	0.51	0.5	0.5	0.49	0.49	0.48	0.48	0.47	0.47	0.46	0.46
2015	0.58	0.58	0.57	0.57	0.56	0.56	0.55	0.55	0.54	0.53	0.53
2016	0.66	0.67	0.67	0.66	0.66	0.65	0.65	0.64	0.64	0.63	0.62
2017	0.68	0.67	0.67	0.67	0.66	0.66	0.65	0.66	0.66	0.65	0.65
Mean	0.637	0.633	0.630	0.626	0.620	0.616	0.609	0.607	0.602	0.595	0.591

From Table 3, the DCC results for the k conditions from 0.55 to 0.65 were found to be similar. Because the aim was to gain as a high a value as possible, a DCC above 0.6 was considered too small. When $k = 0.6$, there was a high DCC, which was able to secure a large proportion of the Chinese regional data in equation (11). Based on the principles proposed above, the final equation for C-CDDs was determined:

$$C - CDDs = 0.6CDDs_{regional} + 0.4CDDs_{exited} \quad (15)$$

4.2 Test on American cities

To assure the feasibility of the proposed C-CDDs, we test the new weather indices proposed on some cities that have similar weather conditions with Shanghai, assuming these cities do not trade the weather derivatives contracts. Atlanta, which has similar weather conditions and high correlation coefficients of 0.9004 with Shanghai, was chosen as the center city to conduct the same analysis using the proposed equation (15). Little Rock and Raleigh were chosen as the matching cities, and Columbus and Mascon were chosen as the regional cities near to Atlanta that had similar weather conditions according to the monthly average temperatures over 25 years. The equation (15) was used to calculate the simulated daily C-CDDs in Atlanta and were compared with the real daily CDDs in Atlanta over ten years, the simulation results show they have similar fluctuation trends.

According to the RMSE data, the error in 2011 was the largest at 0.96, with the mean value of the RMSE over ten years being approximately 0.793. The RMSE in each year were all below 1, indicating that the simulation results were reasonable; therefore, it was believed that the same C-CDDs were feasible in the Shanghai market. The CDDs are cumulative cooling degree day indices, and the prior price settlement refers to the previous monthly cumulative CDDs. Therefore, it is also important to compare the cumulative CDDs in each month. The cumulative CDDs for each month over ten years were calculated and analyzed using equation (10) with the pause between T_1 and T_2 being one month. The mean values for each month were calculated and compared with the mean value of the cumulative CDDs in Atlanta. Then, absolute percentage error (APE) δ was utilized to determine the simulated and real value errors, the calculation for which was as follows:

$$\delta = \frac{|y - y^*|}{y^*} \quad (16)$$

where y represented the simulated cumulative CDDs and y^* represented the mean real value for the cumulative CDDs in Atlanta from 2008 to 2017 (ten years). The results are shown in Table 4.

Table 4. Mean monthly simulated CDDs and true value in Atlanta over ten years

Month	Monthly weather indices		
	Simulated values	True values	δ
Apr.	42.85	42.47	0.009
May.	124.7	124.46	0.002
Jun.	242.4	236.51	0.025
Jul.	280.25	266.13	0.053
Aug.	258.03	252	0.024
Sep.	164.72	165.86	0.007
Oct.	43.32	44.75	0.032

Table 5. Mean monthly simulated CDDs and true value in Las Vegas over ten years

Month	Monthly weather indices		
	Simulated values	True values	δ
Apr.	90.80	92.01	0.013
May.	212.04	222.96	0.049
Jun.	398.39	424.85	0.062
Jul.	461.36	512.57	0.100
Aug.	434.97	471.96	0.078
Sep.	312.64	337.14	0.073
Oct.	129.45	129.53	0.001

Table 6. Mean monthly simulated CDDs and true value in Little Rock over ten years

Month	Monthly weather indices		
	Simulated values	True values	δ
Apr.	36.31	36.04	0.007
May.	111.12	111.71	0.005
Jun.	240.19	251.76	0.046
Jul.	274.63	282.405	0.028
Aug.	254.70	264.89	0.038
Sep.	150.89	151.33	0.003
Oct.	37.23	36.1	0.031

From Table 4, it can be seen that in April, May and September, there were smaller errors than in other months and in July, the error was the largest at 5.3%. Besides testing Atlanta city, we also choose Las Vegas and Little Rock, which have similar weather conditions, to make sure the viability of proposed weather indices. We do the same experiments as we test on Atlanta and the results are shown in Table 5 and 6. It can be seen that, for

Las Vegas, in April and September, there were smaller errors than in other months and in July, the error was the largest at 10%. As for Little Rock, the error is much smaller in April, May and September and the largest is 4.6% in June. As the result of RMSE, the Las Vegas over ten years is approximately 1.505 and Little Rock is 0.90.

These results indicated that the simulated C-CDDs performed well and proved that it was feasible for regions in China such as Shanghai and its surrounding cities to introduce C-CDDs to hedge enterprise risk.

5. SIMULATION AND PROVE

To further analyze the C-CDDs and CDDs correlations in Las Vegas (CDDs-LV), the DCC-GARCH (I, I) model was used to simulate these two weather indices. The data range was set from 2008 to 2017 as a continuous sequence, with the data still being stationary 1st difference data. The two data sets had weak skewness, kurtosis, and normality and had a significant ARCH effect. Based on these characteristics, C-CDDs and CDDs-LV were simulated using the DCC (I, I) and GARCH (I, I) models to determine the parameters for these two weather indices. For the C-CDDs the parameter $\alpha = 0.103$, $\beta = 0.842$ and for the CDDs-LV $\alpha = 0.286$, and $\beta = 0.413$. The parameter for DCC a was 0.032, and for b was 0.820, which indicated that the results were reasonable as the weather indices had a tight correlation, were stable and had strong continuity; therefore the feasibility of proposed C-CDDs was demonstrated.

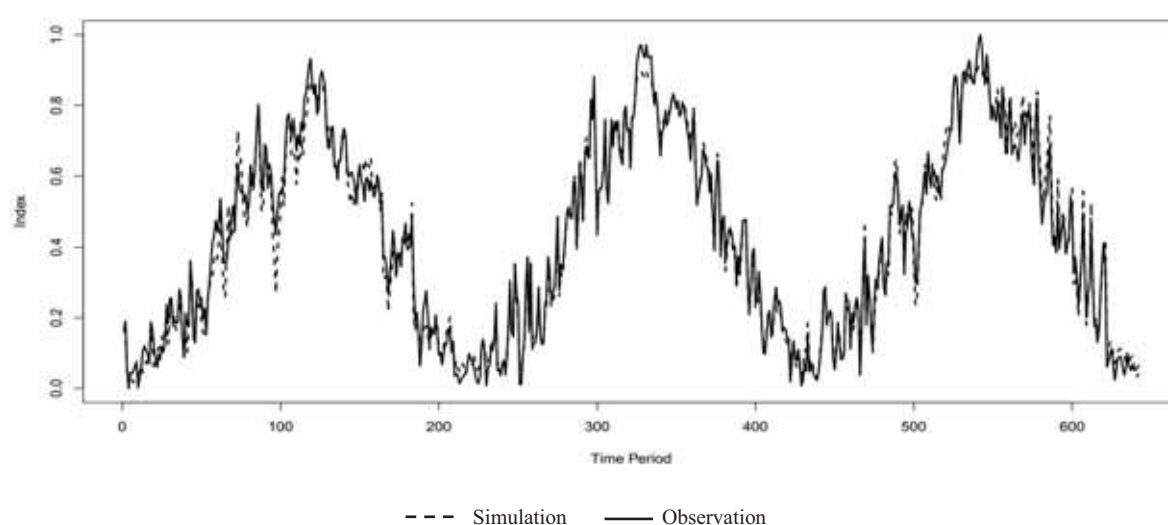


Figure 2: The simulation results of C-CDDs using ANNs

The ANNs was employed to simulate the C-CDDs in Shanghai, the temperatures in Hangzhou, Nanjing and Shanghai and the CDDs in Las Vegas and Little Rock as inputs, with the C-CDDs calculated using equation (15) being the outputs. The previous seven years data was used as the training sets, and the other data was the test set. The simulation for the test set is shown in Figure 2. The RMSE for the simulation was 0.025, and the R-square was 0.978. The simulation results were almost perfect, with a reasonable simulation degree. Only the temperatures in Hangzhou, Nanjing and Shanghai were then used as the inputs for the simulation, with the influence of the US weather indices excluded, as shown in Figure 3; the RMSE and R-square were 0.132 and 0.770. Compared with the previous simulation which included Hangzhou, Nanjing and Shanghai and the CDDs from Las Vegas and Little Rock as the inputs, the simulation degree was found to be worse and the RMSE was

much higher.

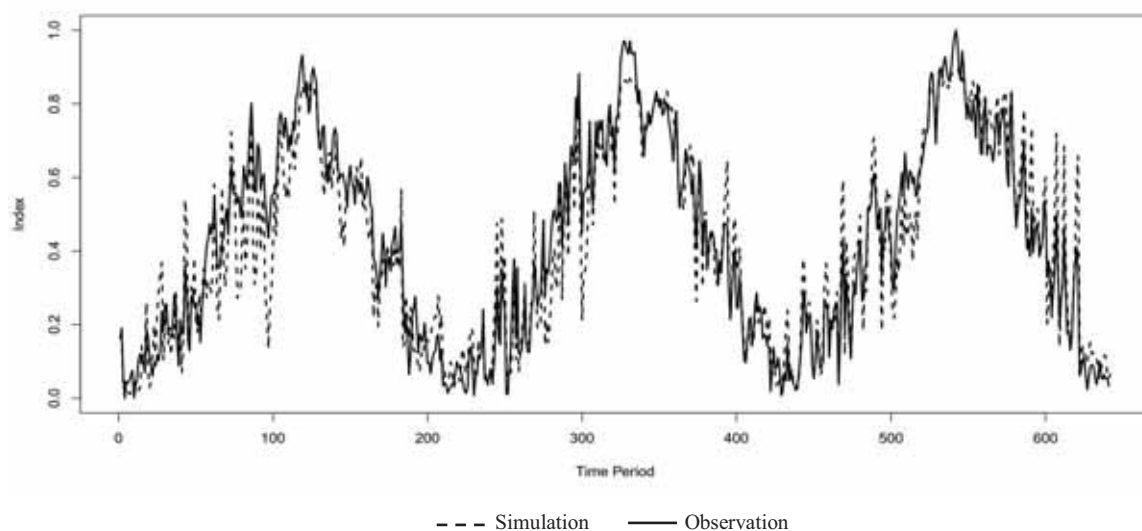


Figure 3: C-CDDs Simulation results using ANNs without US components

As the new weather indices were not predictable and lacked a unified market logic, the poor ANNs simulation results indicated there were design inefficiencies. The 100% fitness also indicated an over fitting as there was no natural variability or volatility. However, in general, the simulation showed that the proposed weather indices had the ability to express the market logic, and were accepted by the machine learning. The exclusion of the US weather indices illuminated some characteristics of the new weather indices. The influence of the US weather indices improved the simulation accuracy of the new weather indices by 21% with the other 77% being influenced by the regional weather in China.

The results of the deep machine learning showed that it was able to part predict the new weather indices and because the ANN performed well when learning the new weather indices, the new weather indices were considered stable and feasible. Therefore, the C-CDDs could be used to determine the prior prices for the weather derivative contracts.

6. CONCLUSION AND FUTURE WORKS

In this paper, a new weather index was proposed that combined Chinese regional weather conditions with US cities that already employed weather derivatives. Two principles were followed when creating the new weather indices and the influence of existing weather indices was added to improve feasibility. From the experiments, it was shown that the proposed weather indices were reasonable and feasible. The Atlanta, Las Vegas and Little Rock weather indices were included in a further assessment to test the suitability of the proposed weather indices for the Chinese derivatives markets, and the DCC-GARCH and ANN models were employed to demonstrate the feasibility, for which the error in the simulated results was found to be acceptable.

China's position as the core global supply chain country significantly influences international trade. Because of China's large land mass, there are varying weather conditions across the country, which can affect production; therefore, it would be sensible for enterprises dealing with China to hedge their risk by trading on the Chinese financial markets. While weather derivatives are being used in Japan, as there are only a few types, they do not cover all situations; therefore, it would be more convenient for surrounding countries to trade

weather derivatives on the Chinese financial markets based on the variability in Chinese weather conditions. In future work, we plan to further develop the Chinese weather derivative markets to assist enterprises and especially energy firms efficiently hedge risk to avoid losses. The creation of the new weather indices may provide a well reference for Chinese financial markets to develop the Chinese weather derivative market, which can help managers well to hedge the risk brought by terrible weather. The policy maker could set the prior prices for the weather derivatives contracts, according to the weather indices we proposed, to help start trading weather derivatives in Chinese financial markets. The development of regional temperature indices could also provide inspiration to other large cities that do not have weather derivatives.

The proposed Chinese weather derivative indices were connected with the US market; however, European indices were not considered in this paper. However, with the expansion of "The Belt and Road" initiative, it is necessary to deepen cooperation with European countries, improve the development of the international financial markets, and ensure communication between the two regions.

ACKNOWLEDGEMENT

This research was supported by the Science and Technology Department of Shaanxi Province Government under Grant 2011KRM108, The Chinese Ministry of Science under Grant 2012GXS2D027, Shaanxi Land Construction Group under Grant SXDC201208.

REFERENCES

- [1] Gersema, G., Wozabal, D. (2017). An equilibrium pricing model for wind power futures. Social Science Electronic Publishing.
- [2] Challis, Bright forecast for profits, reactions.
- [3] H. M, Hedging the force of nature.
- [4] A. Zapranis, A. Alexandridis (2010). Modeling and forecasting cumulative average temperature and heating degree day indices for weather derivative pricing. *Neural Computing & Applications*, 20(6): 787–801.
- [5] Cramer, S., Kampouridis, M., Freitas, A. A., & Alexandridis, A. K. (2017). An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications*, 85: 169-181.
- [6] S. Liu, L. Wang, W. Huang (2017). Effects of process and outcome controls on business process outsourcing performance: Moderating roles of vendor and client capability risks. *European Journal of Operational Research*, 260(3): 1115–1128.
- [7] Su, L. (2011). A Rule Extraction Based Approach in Predicting Derivative Use for Financial Risk Hedging in Construction Companies. *International Conference on Information Management, Innovation Management and Industrial Engineering IEEE*, 6510-6514.
- [8] Buchholz, M., & Musshoff, O. (2014). The role of weather derivatives and portfolio effects in agricultural water management. *Agricultural Water Management*, 146(146): 34-44.
- [9] Hong, S. J., & Sohn, S. Y. (2013). Peer group analysis for introducing weather derivatives for a city. *Expert Systems with Applications*, 40(14): 5680-5687.
- [10] Benth, J. Å., Benth, F. E., & Jalinskas, P. (2007). A spatial-temporal model for temperature with seasonal variance. *Journal of Applied Statistics*, 34(7): 823-841.
- [11] Turvey, C. G., & Kong, R. (2010). Weather risk and the viability of weather insurance in china's gansu, shaanxi, and henan provinces. *China Agricultural Economic Review*, 2(1): 5-24.
- [12] Zong, L., Ender, M. (2016). Spatially-aggregated temperature derivatives: agricultural risk management in China. *International Journal of Financial Studies*, 4(3).
- [13] Tim Bollerslev. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):

- 307-327.
- [14] Robert F. Engle. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4): 987-1007.
 - [15] Hou, Y., Li, S. (2016). Information transmission between u.s. and china index futures markets: an asymmetric dcc garch approach . *Economic Modelling*, 52: 884-897.
 - [16] R. F. Engle. (2002). Dynamic conditional correlation: A simple class of multivariate garch models. *Business and Economic Statistics*, 20 (3): 339-350.
 - [17] Y. Tsukuda, J. Shimada, T. Miyakoshi. (2017). Bond market integration in east asia: Multivariate garch with dynamic conditional correlations approach. *International Review of Economics & Finance* 51: 193–213.
 - [18] X. Zhong, D. Enke. (2017). A comprehensive cluster and classification mining procedure for daily stock market return forecasting. *Neurocomputing*.
 - [19] A. Zapranis, A. Alexandridis. (2009). Weather derivatives pricing: Modeling the seasonal residual variance of an ornsteinuhlenbeck temperature process with neural networks. *Neurocomputing*, 73(13): 37–48.
 - [20] A. K. Alexandridis, M. Kampouridis, S. Cramer. (2017). A comparison of wavelet networks and genetic programming in the context of temperature derivatives, *International Journal of Forecasting*, 33(1): 21–47.

An Exploration of Cross-border E-commerce Consumer Feedbacks:

An LDA Approach

Jian Mou^{1}, Gang Ren², Chunxiu Qin¹, Kerry Kurcz³*

¹ School of Economics and Management, Xidian University, Xi'an, Shaanxi, China 710126

² College of Business Administration, Pusan National University

³ Department of Information & Decision Science, University of Illinois at Chicago

Abstract: Cross-border e-commerce (CBEC) has become an important channel to help a firm to go into the international market in China. The recent influx in development of CBEC has caused a simultaneous influx in accumulation of text data such as consumer feedback. To better understand consumer feedback, we collected data from a leading CBEC firm in China to explore the topics of feedback posted directly by their customers. We employed the Latent Dirichlet Allocation (LDA) model to explore the potential topics focused on most by consumers. We found 35 primary topics are mentioned by both sellers and buyers. For the seller's perspective, we found that the topics such as commission, product audit, communication between seller and buyer, order management and traffic are most important. On the buyer's side, we found that return and refund, product tracking, product description, shipping time, and seller performance are the most mentioned topics. This study will help contribute to the understanding of how consumer feedback will help firms in many ways, including but not limited to recovering service and product failures, audit internal functions, and improve product quality.

Keywords: Cross-border e-commerce, Latent Dirichlet Allocation, text mining, consumer feedback, LDA

1. INTRODUCTION

Cross-border e-commerce (CBEC) has become an important channel to help a firm to go into the international market. This phenomenon has been highly recognized, especially in China. Many CBEC firms have emerged in the past decade, including both AliExpress (launched 2010) and Lightbox (2012). The recent influx in development of CBEC has caused a simultaneous influx in accumulation of big data. Among the vast majority of big data, text data is particularly useful. Namely, text data containing consumer feedback has been noted as a significant resource. Successful techniques can identify consumers' feelings, both before and after the purchase stage. Indeed, text mining can help businesses to better understand consumer behavior, helping to expand a firm's brand into an international market.

To begin, text mining is a lot like it sounds. Computer algorithms can sort through a body of text for certain key words and phrases. Consumer feedbacks published publicly can influence other consumer's purchase decisions^[16]. As expected, the negative sticks more than the positive. Dissatisfied consumers are more likely to post negative feedback, which may threat a firm's reputation for a longer period of time than positive feedback^{[7][14]}. Consequently, managing consumer feedback is important for CBEC firms.

A majority of the existing papers studying consumer feedback behavior focuses on the context of social media^[6]. This is because social media is a revolutionarily crucial communication channel to share individuals' opinion on a brand or company. The feedback published on social media, including Facebook, Yelp, and Trip Advisor, is easy to see among any user with an account. Hence, CBEC firms benefit immensely from managing their brand on social media. Often, unsatisfied consumers want to disseminate their unsatisfied shopping experiences from some firm by voicing their intention to refrain from continuing to shop, and to caution other consumers not to buy from that firm as well. Other consumers may remain loyal consumers despite experiencing

* Corresponding author. Email: jian.mou@xidian.edu.cn (Jian Mou)

unsatisfactory service, but they still let the firm know about their bad experience by expressing their feelings publicly, hoping the firm will thus attempt to fix the service failure. On the other hand, some consumers prefer their reviews not be published for anyone to see (e.g., some embarrassing or private products). Such consumers may send their feedback to the firm directly. Therefore, the feedback published on social media may not properly represent all customers' perspectives. As a firm, both public and private feedback are important to obtain in order to fully understand their consumer's feelings. Therefore, it is also important for a firm to collect both types of feedback using social media as well as their own feedback collection system(s).

Previously, the research of consumer feedback that generated on social media has been studied in the tourism and hospitality area^{[7][26][34]}, e-commerce setting^{[16][28]}, and social media^[14]. There remains a lack of study focusing on consumer feedback in CBEC context. CBEC is different with domestic e-commerce. This is because the nature of CBEC results in more complex transactions (i.e., the complex return and refund process). In CBEC, additional uncertainties such as asymmetric product information, privacy concerns, after-sales quality uncertainty, confiscation risk and delivery risk, all of which may detract from the use of CBEC^{[21][27]}. In our study, we used consumer feedback from a leading CBEC firm. On its official website, there are several incentives for customers to give feedback. For example, multiple open questions (i.e., *what are your suggestions for our firm?*) lead to links which allow customers to post their opinions. This offers an up-to-date view of consumers' suggestions and feedback. Only the accumulated feedbacks are analyzed by a firm. This may gain more insight to recover the service failures or to improve consumer satisfaction.

Previously, consumer-generated feedback has been studied in the purchase decision process, the reliability of the feedback, and the true impact of the feedback^[34]. Few studies employ an LDA method to discover the topics which emerged in CBEC context. To better understand consumer's feedback, we collaborated with a leading CBEC firm to explore the topics of feedbacks that posted by their customers. To do so, we employed the Latent Dirichlet Allocation (LDA) model to explore the potential topics, or concerns, raised by consumers. Specially, we aim to answer the following research question:

Which types of feedback are mostly commonly voiced via a firm's private feedback collection system?

This study is organized as follows: in next section, the context of CBEC has been introduced, and the term of consumer feedback has been reviewed. Thereafter, to resolve the research question, the research method of LDA has been outlined. Then, the research results are discussed. Finally, drawing on the research results, the conclusion section has been spread out.

2. LITERATURE REVIEW

2.1 Cross-border e-commerce

Chinese e-commerce revolutionary Belt on One Initiative (previously One Belt On Road) brings more opportunities to increasing world economics. Especially for China, it can gain more access to foreign markets by exporting their products and services overseas^[18]. It also brings more opportunity for the development of CBEC. According to China Daily (30-10-2017), Chinese CBEC transaction reached 6.7 trillion yuan (\$1.01 trillion) in 2016 and it is predicted that by 2018, the value of China's import and export transaction via e-commerce might hit 8.8 trillion yuan^[32].

CBEC is a new mode of import and export trade via e-commerce platform. It allows the firm to avoid high competition in their local markets and to seek more markets in abroad^[9]. The term of CBEC derives from the combination of cross-border shopping and electronic commerce. Cross-border shopping refers to the activity wherein a consumer purchases goods abroad, which may need a consumer cross an international boundary into another country^[38]. CBEC refers to "an online transaction of buying and selling products and services using

information communication technologies (ICTs), where buyers and sellers are located in two different countries and therefore, different jurisdictions^[1]. Due to the definition of CBEC, it includes both import and export transactions. In China, the term of “haitao¹” has been used frequently to refer Chinese consumers using an e-commerce website for the purchase of goods from abroad^[25]. In our study, we only consider one specific type which refers to Chinese firms sell products of services via a CBEC platform.

Different from traditional cross-border (CB) shopping, the characteristics of CBEC are *multilateral*, *direct*, *small amount* and *high frequency*^[36]. *Multilateral* in this context specifically refers to the fact that CBEC can integrate different countries or firms' services, such as logistics service and payment services together to perform a transaction. *Direct* refers to CBEC context allows a consumer to buy product directly from a foreign country, while offline cross-border purchase may purchase a large number of quantity products. Further, CBEC are usually in a small amount with a high frequency purchase^[36]. CBEC logistics also represents more process linkages, varied channels and wider span of geography^[12].

CBEC is more complex than domestic electronic commerce due to the uncertainty among international market environments and the shipping of product across countries^[38]. Meanwhile, cross-border shopping features the economic advantages, quality, service and variety of product selection^[38]. The advantages of shopping abroad additionally include more competitive prices, warranties, access to latest products and a pleasant shopping environment^[33]. Cross-border shopping may also motivate value perceptions, such as relaxation, pleasure, or dissatisfied with their local shopping environment e.g., currency rates, and government policies^[33].

Although CBEC brings potential benefits, consumers still concern the legal enforcement, culture barriers and high shipping costs^[9]. Further, import duty has also been recognized as a barrier for CBEC, this is because consumers may face an additional cost when purchasing a product from a foreign country^[23]. Turban et al.^[35] also emphasized that CBEC website should provide multiple languages, currencies, payment systems and input devices. To reduce the complex checkout process, it should eliminate the detailed user profiles and offer free shipping and rewards to encourage repeat traffic.

Cross-border trade research has been gradually attracting attention in recent years. For example, just this year, Huang and Chang^[19] empirically examined the determinants of a consumer purchase on a CBEC platform. They found that any consumer transaction on a CBEC platform is mostly influenced by trust beliefs and value perceptions. The determinants of cross-border shopping have been classified as market-based factors and consumer-based factors^[22]. Market factors relate to price level, shopping environment, shopping facilities, product variety and product/service quality. Consumer factors focus on consumer demographics and motivational characteristics, such as pleasure bargaining and social experience^[22]. Singh et al.^[31] state that localization also plays an important role in CBEC setting, which may affect consumers' usability, accessibility and interactivity. Utilizing this knowledge can help a firm increase the relationship between their products and consumer. For instance, language serves as a cultural indicator, and can help shape ideas for a target consumer^{[23][24]}. In addition, Liu and Hong^[25] emphasized the importance of online payment facilities and cost-efficiency of parcel delivery systems. Further, Guo et al.^[10] studied the impacts of chargeback fraud on CBEC seller's intention to use CBEC platform.

2.2 Consumer feedback

Given consumer feedbacks include both positive and negative, we define consumer feedback as the favorable and unfavorable experiences and a consequent recommendation of a product or service to an online platform^[26]. The benefits of managing consumer feedback have been summarized as impacting on customer

¹ Haitao means Chinese consumers purchase products from abroad.

retention rates, deflecting the spread of damaging word-of-mouth, promoting more positive word-of-mouth, increasing customer perception of quality, improving marketing intelligence, and promoting a positive company image^[15].

Consumer feedback affects both the seller and the buyer side. Since consumer feedback reflects consumers' online shopping experience, it can help other consumers to evaluate a seller's reputation on the buyer side, and therefore potentially increase (or decrease) sale performance for the seller^[13]. On the other hand, it also can help sellers to better improve their service and product quality^[11]. Consumer reviews are more focused on online stores and their goods as well as services. Reviews regarding a specific purchase are more likely to occur just after that purchase. General consumer feedback that are more focused on the suggestions for the platform and its services may occur at any time.

Wu and Huang^[28] argue that while consumer complaint behavior can help online vendors develop innovative products and services, it also can help consumers to solve transaction issues. However, as a new firm, it may also hinder consumers' ability to share their negative experience on social media^[15]. Given the high cost of finding a new customer, firms are fully aware that it is imperative to maintain customer satisfaction whenever and wherever possible. Therefore, it is important to redress the problems to satisfying the complained customers.

When a consumer experiences a problem with a product, three types of consequent actions have been summarized by Harrison-Walker^[15], which are listed below:

Disassociating themselves with the company and/or spreading negative word-of-mouth

posting a complaint directly to the company, and taking indirect public action on a social media.

Negative word-of-mouth and feedbacks than are posted on social media have been widely studied in marketing and information systems fields^{[3] [5] [14] [17] [20] [30]}.

However, feedback that have been posted directly to the company has been paid little attention. This study aims to discover the topics from the feedbacks that have been posted by the customers to the CBEC firm directly.

Online service reviews have been emphasized its two primary functions. The first is to assist the decision-making of service consumer; the second, to assist service providers in service quality improvement^[11]. The context of this study focuses on online feedback to assist service providers in service quality improvement. This is because customer service teams need to be particularly sensitive to customer needs and expectations when they attempt to recover from service failures. Prior studies have merely explored data from online review websites, e-commerce website and tourism website are often examined the numeric ratings rather than extensive exploring the textual comment data. The current study extends the current literature by focusing on the topics and tones of the textual feedbacks given by cross-border consumers, and discussing the benefits for both the buyer and seller side.

Because both public and private feedback plays an important role in studying consumer satisfaction, organizations must manage consumer feedback appropriately^[29]. Currently, firms usually collect consumer feedback constantly via customer satisfaction surveys, online reviews, e-mails, as well as in-person, for both praises and complaints^[34]. Consumer feedback can also serve as a benchmark, internal audit for a firm^[34]. Moreover, by studying consumer feedback can help an organization to resolve the problems caused by service failures^[2].

3. RESEARCH METHOD

3.1 Topic Extraction Using Latent Dirichlet Allocation (LDA)

Topic modeling has become an increasingly important method for extracting useful topics from vast amounts of textual data. Among all topic modeling methods, Latent Dirichlet Allocation (LDA)^[4] is one of the most widely used^[25]. It helps to capture both explicit and implicit topics, including the relative words for each topic. The entire process is based on Bayesian networks.

Take a collection of seller or buyer reviews, $R = \{r_1, r_2, \dots, r_{|R|}\}$. Each review, call it \mathbf{r} , is represented as a collection of words $\mathbf{r} = \{w_{1r}, w_{2r}, \dots, w_{Nr}\}$, here $r_{|R|}$ denotes the $|R|$ th review in the collection R , and w_{Nr} denotes the N th word in review \mathbf{r} . The latent topics are denoted by $K = \{\Phi_1, \Phi_2, \dots, \Phi_{|K|}\}$, where $|K|$ is the topic number.

The underlying principle of the LDA model is that it allocates words from multiple, separate reviews to a new, single document, assigning a probability to each word. Consequently, each new document represents a topic consisting of some highly-related co-occurring words (i.e., top words). The generative process for each new document is described as follows:

- (1) For each review, choose the topic distribution $\theta_r \sim \text{Dirichlet}(\alpha)$
- (2) For each word w_{Nd} in the review \mathbf{r}
 - (a) Choose a topic $z_{Nr} \sim \text{Multinomial}(\theta_r)$
 - (b) Choose a word w_{Nr} from $p(w_{Nr} | z_{Nr}, \beta)$ (i.e., $\Phi_{r,N}$)

In the generative process, two Dirichlet priors α and β are endowed to determine the document-topic distribution θ and the topic-word distribution Φ , respectively, where we assume both θ and Φ follow multinomial distribution. We estimate the parameters for θ and Φ using the Variational Inference^[4] and Gibbs sampling^[8] methods. Given α and β for a single review, the joint distribution of θ , z , and w is computed as follows:

$$p(\theta, z | w, \alpha, \beta) = \frac{p(\theta, z, w | \alpha, \beta)}{p(w | \alpha, \beta)} \quad (1)$$

where:

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta) \quad (2)$$

$p(w | \alpha, \beta)$ in Equation (1) denotes the weigh, or probability, of each extracted word. The probability represents relative importance of one word in one topic. The higher the probability, the more relevant the word.

$$p(w | \alpha, \beta) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \int \left(\prod_{i=1}^k \theta_i^{\alpha_i - 1} \right) \left(\prod_{n=1}^N \sum_{i=1}^K \prod_{j=1}^V (\theta_i \beta_{i,j})^{w_n^j} \right) d\theta \quad (3)$$

Once we have the probabilities per word, it is often still difficult to determine the topic number K from a given collection for the LDA model^[37]. To obtain an appropriate topic number, Blei et al.^[4] suggests calculating the perplexity for a held-out test set to evaluate the performance of the trained models. For example, if given the topic number collection $K = \{1, 2, 3, 4, 5\}$, we could select one appropriate topic number by calculating the perplexity for the five trained models, wherein each topic number from the K s can be used to train a single LDA model. The lower perplexity, the better generalized performance will be obtained for the trained model. The entire process is to find the lowest perplexity from the given topic number set K s, and use this to determine the topic number K from a given collection. The perplexity is given by

$$\text{Perplexity} = \exp \left\{ - \frac{\sum_{r=1}^{|R|} \log p(w_{Nr})}{\sum_{r=1}^{|R|} N_r} \right\} \quad (4)$$

It is typical to calculate a multiple-fold validation method and use an average perplexity.

3.2 Data Analysis and Results

In this study, we adopted the data from a large CBEC firm in China. The data ranges between January 2016 to December 2016, containing 21,852 total posts from both buyers and sellers. The seller feedback makes up 11,816 posts, and the buyer feedback, 10,036 posts.

We employed a five-fold validation method to calculate an average perplexity for each trained model. We observed the change of perplexity by setting the number of topics ranging from 5 to 100, with an interval of 5. It is sufficient to capture the number of topics from the feedback in real life. We found that it obtains the smallest average perplexity for both seller and buyer data when the number of topics approximate to 35. Therefore, the appropriate number of topics is 35 for the seller and buyer data. Note the similarity in shape; it is safe to assume there is a drop before the perplexity begins to increase as the number of topics also increases. However, it occasionally obtained the identical number of topics for the seller and buyer in our study. Because these 35 topics can be summarized into several broader topics (i.e., dimension) for the buyer and seller respectively, it does not make any sense with the identical number of topics. Rather than the identical number of topics, it is more important to exploit the common topic they pay attention to.

Table 1: Top 5 topics out of 35 topics (seller dataset)

Topic	Dimension	Top 10 words
1	佣金 (Commission) 0.925	佣金(Commission) 0.467 太高(Too High) 0.174 手续费(Handling Charge) 0.052 收取(Charge) 0.049 利润(Profits) 0.044 比例(Proportion) 0.039 过高(Too High) 0.035 降低(Decrease) 0.033 支付(Pay) 0.020 接受(Accept) 0.013
2	产品审核(Product Audit) 0.877	审核 (Audit) 0.516 速度 (Velocity) 0.137 太慢 (Too Slow) 0.107 快点 (Quicker) 0.026 快速 (Fast) 0.025 比较慢 (Slowish) 0.022 加快 (Accelerate) 0.019 通不过 0.009 快些 (Quicker) 0.008 用品 (Article) 0.008
3	买卖家沟通 (Communication) 0.863	客户 (Customer) 0.541 沟通 (Communication) 0.108 朋友 (Friend) 0.081 介绍 (Introduction) 0.062 评价 (Evaluate) 0.019 喜欢 (Like) 0.016 不上 (Offline) 0.012 机制 (Mechanism) 0.010 第二天 (The Next Day) 0.007 延迟 (Delay) 0.007
4	订单管理 (Order Management) 0.821	订单 (Order) 0.461 发货 (Shipping) 0.094 取消 (Cancel) 0.053 成交 (Make a Deal) 0.038 付款 (Payment) 0.037 单号 (Order Number) 0.033 下单 (Place The Order) 0.032 错误 (Error) 0.028 风控 (Risk Control) 0.027 地址 (Address) 0.015
5	流量 (Traffic) 0.803	流量 (Traffic) 0.377 太少 (Too Little) 0.148 曝光 (Exposure) 0.084 整体 (Overall) 0.035 提高 (Improve) 0.034 有待 (Remain) 0.030 引流 (Drainage) 0.027 浏览量 (Page View) 0.024 提升 (Promote) 0.024 太低 (Too Low) 0.021

To move toward a better understanding of seller and buyer feedbacks, we extracted the top words over the top topics utilizing the LDA model. Table 1 reports the top 5 topics and each respective top 10 words (and phrases) from the seller perspective. The top 5 topics are commission, product audit, communication between seller and buyer, order management and traffic. Each topic's top ten discussed words can see detailed in Table 1. Note that it does not yet associate each top 10 words or phrases with some degree of quality (i.e., good or bad). We also presented the relative weight of each word that calculated in Equation (3). A sum of the weights of the top 10 words, represents the relative importance of each topic. We found that the most important topic for the sellers was "commission" (0.925), which refers to the commission that the cross-border platform charges when a transaction occurs. In addition, many sellers pointed out that the product audit processing time needs to decrease. Further, the communication between buyer and seller needs more ease of use and user friendly platforms. The other topics, interestingly, are related with these primary issues. For instance, the sellers complained that order

management is difficult to use. Indeed, the function of order management needs improvement.

Further, we also considered the LDA results for buyer feedback when setting the number of topics as 7. The extracted topics include the shipping, goods, ease of use, return and refund, product description, Customer Relationship Management (CRM), and seller performance. By comparing the relative importance of topics, we found that buyers were more concerned about the shipping time with a total weight of 0.210. A too-long duration of shipping may cause the consumers' complaints, while a fast shipping, conversely, may increase the favorable impression to the product. Interestingly, the positive topic (e.g., Good) has been found from the buyer's dataset. On one hand, CBEC buyers are posting some complaints to the firm, on other hand, CBEC buyers are also praise firms for their excellent product and services. This can significantly help CBEC firms to better understand their problems and benefits for current buyers.

Table 2 depicts the top 5 topics and key words of each topic by setting the number of topics as 35 according to the perplexity. The top 5 topics are return and refund, product tracking, product description, Shipping time, and Seller performance. Give the complex nature of CBEC, the process of return and refund becomes critical for the CBEC success, where the more simplistic, the better. In addition, the difficulty of tracking products is another topic for CBEC transaction. Further, product description, shipping, and seller performance are also recognized as principal issues for CBEC.

Table 2: Top five topics out of 35 topics (buyer dataset)

Topic	Dimension	Top 10 words
1	Return and Refund 0.815	item 0.340 refund 0.248 return 0.059 purchased 0.038 asked 0.033 full 0.031 reason 0.018 request 0.018 receive 0.018 lied 0.012
2	Tracking 0.787	tracking 0.246 number 0.194 information 0.118 provide 0.066 post 0.038 package 0.036 correct 0.030 shipped 0.025 update 0.020 stated 0.015
3	Product Description 0.771	product 0.409 description 0.106 pictures 0.096 photo 0.048 details 0.023 actual 0.019 advertised 0.018 include 0.017 match 0.015 image 0.018
4	Shipping Time 0.761	time 0.280 long 0.121 delivery 0.088 ship 0.075 takes 0.060 survey 0.041 delivered 0.031 arrive 0.031 date 0.020 promised 0.015
5	Seller Performance 0.742	make 0.317 purchase 0.221 customers 0.056 clear 0.031 coupon 0.029 easier 0.026 merchandise 0.019 part 0.014 mistake 0.014 returns 0.013

4. DISCUSSION AND CONCLUSION

CBEC has become an important channel to help local firms to increase abroad market share. With the development of CBEC, there has been accumulated many consumer feedbacks. Through managing consumer feedback, CBEC firms can increase customer retention rates, deflect the spread of damaging word-of-mouth, promote more positive word-of-mouth, increase customer perception of quality, improve marketing intelligence, and promote a positive company image. Given aforementioned advantages, CBEC firms are eager to know their consumer feedback. To better understand the types of feedback are mostly commonly voiced, we employed an LDA technique to track the topics from a large CBEC firm's private feedback collection system. Our study makes both practical and theoretical contributions.

Our study makes the practical contributions for both CBEC platforms and business consulting firms. Through analyzing consumer feedback, the CBEC platforms can be more easily to understand their firm's disadvantage and advantage, and therefore help them to build a better consumer relationship. The firms may choose the most important issues and resolve such issues to significantly improve consumer's satisfaction. In addition, CBEC firms should handle consumer feedbacks appropriately to correct the failed service and thereby transform it into a satisfactory encounter. For example, CBEC may develop a better consumer feedback system to contact with non-satisfied consumers to improve the level of consumer's post-purchase satisfaction. This can be used for CBEC firm to better mining their consumer's suggestions, and for complaints websites to deeper management their text data. Business consulting firms may track different CBEC platform's consumer feedbacks. As a benchmark to guide the newly entered CBEC to better management their consumer relationship.

Indeed, it is important to understand both the seller and buyer feedback for both platform developers and managers. A better understanding the feedback, is beneficial for the platform developers to understand the difficulties, barriers for users, both buyers and sellers. For example, many sellers have difficulties in the store design (including of that online), order management, and product function setting. In addition, the platform managers can also learn the sellers' complaints (e.g., complaints on the expensive commission, slow product auditing, and dispute arbitrament). A better understanding the buyer feedback, is beneficial for the sellers to understand the buyers' complaints, including product defects, speed of delivery. Moreover, by analyzing the buyers' complaints, the platforms can effectively detect products with poor quality, and fake products, as well as the fraudulent retailers. Rather than the numerical rating, the text mining process acts as a kind of regulation mechanism.

Our study also makes several theoretical contributions. Based on the theoretical underpinning of the consumer complaint behavior, we explored the 35 topics from consumer feedbacks are important for cross-border buyers. Importantly, for a seller perspective, we found that store related issues were ranked as the most important issue that is mentioned by CBEC sellers, while for a buyer perspective, we found that shipping related issue is the most important for international buyers. The factors we determined which may significantly improve consumer's satisfaction, and therefore develop the theoretical of consumer complain behavior and satisfy consumers.

ACKNOWLEDGEMENT

This study supported by the Fundamental Research Funds for the Central Universities of No. JBX180604; the Chinese Ministry of Education Humanities and Social Sciences project of No. 16YJC630038; and the National Natural Science Foundation of China of No. 71573199.

REFERENCES

Given the page limitations, we only listed a part of our citations. Remains citations may request from the corresponding authors.

- [1] Abdallah F. (2014). Postal service innovation: New value propositions to enable international e-commerce. *Postal Services in the Digital Age* M. Finger et al. (Eds.) IOS Press.
- [2] Bach S B, Kim S. (2012). Online consumer complaint behaviors: the dynamics of service failures, consumers' word of mouth, and organization-consumer relationships. *International Journal of Strategic Communication*, 6(1): 59-76.
- [3] Beser A, Lackes R, Siepermann M. (2017). Silence is golden-when firms should react to negative word of mouth. in *Proceeding of the 25th European Conference on Information Systems (ECIS)*, Guimarães, Portugal, June 5-10, 2017: 2136-2152.
- [4] Blei D M, Ng A Y, Jordan M I. (2003). Latent Dirichlet Allocation. *Journal of machine Learning research*, 3(1): 993-1022.
- [5] De Keyzer F, Dens N, De Pelsmacker P. (2017). Don't be so emotional! How tone of voice and service type affect the relationship between message valence and consumer responses to WOM in social media. *Online Information Review*, 41(7): 905-920.
- [6] Fernandes T, Fernandes F. (2017). Social Media and Tourism: The Case of E-Complaints on TripAdvisor. In *Marketing at the Confluence between Entertainment and Analytics* 825-829. Springer, Cham.
- [7] Gerdes Jr J, Stringam B B, Brookshire, R.G. (2008). An integrative approach to assess qualitative and quantitative consumer feedback. *Electronic Commerce Research*, 8(4): 217-234.
- [8] Griffiths T L, Steyvers M. (2004). Finding scientific topics. in *proceedings of the National Academy of Science*, 101:5228–5235.

- [9] Guo Y, Bao Y, Stuart B J, Le-Nguyen K. (2017). To sell or not to sell: Exploring sellers' trust and risk of chargeback fraud in cross-border electronic commerce. Forthcoming in *Information Systems Journal*.
- [10] Guo Y, Barnes S J, Jia Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59: 467-483.
- [11] James T L, Calderon E D V, Cook D F. (2017). Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. *Expert Systems with Applications*, 71: 479-492.
- [12] Jiao Z L. (2016). Modes and Development Characteristics of China's Cross-border E-commerce Logistics. Springer-Verlag Berlin Heidelberg, B. Liu et al. (eds), *Contemporary Logistics in China*, Current Chinese Economic Report Series.
- [13] Yao E, Fang R, Dineen B R, Yao X. (2009). Effects of customer feedback level and (in) consistency on new product acceptance in the click-and-mortar context. *Journal of Business Research*, 62(12): 1281-1288.
- [14] Istanbuluoglu D. (2017). Complaint handling on social media: The impact of multiple response times on consumer satisfaction. *Computers in Human Behavior*, 74: 72-82.

Understanding IS Success Model and Valence Framework in Sellers'

Acceptance of Cross-border E-commerce

*Yi Cui¹, Jian Mou^{*1}, Jason Cohen², Yanping Liu¹*

¹ School of Economics and Management, Xidian University, Xi'an, Shaanxi, China 710126

² School of Economic and Business Sciences, University of the Witwatersrand

Abstract: Cross-border e-commerce becomes more and more popular and general. The foci of researches in e-commerce have moved from domestic towards to global market. Yet, most of extant literatures are from buyer's perspective, whereas sellers are also important in the success of cross-border e-commerce. In this study, we are aiming to identify the elements of the success of cross-border e-commerce and the relationship with trust and intention to use from seller's perspective. To do so, we apply a mixed method to accomplish this research. We have identified the key factors which sellers are concerned about, and why they engage in cross-border e-commerce. In addition, we have developed new dimensions with associated items for system quality, service quality, perceived benefit and perceived cost in the context of cross-border e-commerce. The theoretical contributions and practical contributions have been discussed lastly.

Keywords: cross-border e-commerce, trust, ISS model, valence framework, mixed method

1. INTRODUCTION

Along with the development of e-commerce and economic globalization, cross-border e-commerce (CBEC) has successfully blossomed in the recent years. Buyers and sellers in online transactions are not limited to a domestic e-marketplace but in a broader and more general global market. According to Alibaba Group's report, the global B2C cross-border e-commerce transactions was \$230 billion in 2014, and this volume will increase to \$1 trillion by 2020 ^[1]. The development of cross-border e-commerce in European Union is relatively early. The main reasons of high utilization rate comparing to other non-EU countries is that EU countries have greater internet penetration, availability of credit cards, investment, availability of venture capital, education level, and spillover effects from neighboring countries ^[2]. Therefore, CBEC holds the greatest potential for the growth of e-commerce in the EU and globally ^[3].

In addition, China has become another major market for CBEC with a compound annual growth rate of 30% each year since 2012, and the volume of its CBEC is approximate 20% of total volume of foreign trade ^[4]. It has been predicted that China will transcend the USA, UK, German, France, Japan and become the biggest cross-border market by 2020; moreover, Asia will be the CBEC center with 40% of total revenues by 2025 ^[5].

The masses of rise in the number of e-marketplaces such as Amazon, eBay, and Alibaba have triggered the interests of both practitioners and scholars ^[6]. However, many cross-border e-marketplaces have failed in recent years due to the poor performance ^[7], for example, Metao.com found in 2013 and failed in 2016. What mechanisms underline such failing issues have posed numerous questions for further investigation. Even though CBEC provides more choices with lower price and offers opportunities to both developing and developed countries to get benefits from global transactions, it still faces much more barriers than domestic e-commerce due to its special attributes ^[8]. These barriers include cultural difference, language translation, legal issues, geographic issues, localization, payments in global e-commerce trades, customs clearance problem, and logistic factors ^[9]. Beyond these common issues in CBEC, the quality of e-marketplace itself is another substantial in the

* Corresponding author. Email: jian.mou@xidian.edu.cn (Jian Mou)

success of global online trade. This also associates with the trust building process. The “trust” has been empirically studied as a crucial factor of intention to use through signaling theory in B2C^[10] and B2B^[11]. Therefore, understanding how trust influences the increasing CBEC adaption is crucial for uncovering the success of CBEC. Extant researches on CBEC are mainly focusing on the different determinants of successful transaction between buyers and sellers, such as reputation, word of mouth and so forth which lack the reasons in depth for the success of CBEC from seller's perspective. There are some studies focusing on e-marketplaces success, however, the measurement items are mainly from previous study. We agree that adapting previous measurement to study a new context is appropriate. However, given the complex nature of CBEC, we believe that there will be some new insights associating with system quality and service quality. In addition, there lacks a valid scale for both ISS model and valence framework from sellers' perspective, especially in the context of CBEC. To investigate potential components, we therefore conducted a mixed method study to explore their dimensions. Therefore, we wish to bridge this gap by finding the basic components of system quality, service quality, perceived benefits and perceived costs of conducting CBEC from the seller's perspective. As discussed above, China holds the greatest potential to become the largest cross-border e-marketplace in the world, so we apply ISS model and valence framework as our theoretical frameworks, and we collected data from a leading cross-border e-marketplace in China.

This paper is organized as follow: firstly, the theoretical backgrounds for our research have been introduced; secondly, the detail of our exploratory study and its findings are described; thirdly, the confirmatory study including our research model, hypothesis development, and method are displayed; finally, we conclude the contributions of this research and future works.

2. LITERATURE REVIEW AND THEORETICAL BACKGROUND

2.1 Information System Success Model

The original ISS model was revealed in 1992 through a systemic review of researches between 1981 to 1987^[12]. In this model, six dimensions are identified namely system quality, information quality, use, user satisfaction, individual impact, and organizational impact^[13]. Afterwards, to meet the needs of researchers in information system and adapt the e-commerce environment, DeLone and McLean^[14] have updated their ISS model with three modifications. Firstly, service quality as a new dimension was added along with system quality and information quality to make the whole model more comprehensive in evaluating the overall quality of e-commerce success. Secondly, system use was separated into intension to use and use which provide a non-mandatory system use option. Finally, they combined individual and organizational impacts into net benefits to make the model more parsimonious^[15]. The final model could be described as that system quality, information quality, and service quality positively affect intention to use and user's satisfaction which will lead to actual usage, both user's satisfaction and actual usage can influence net benefits positively^[14]. As suggested by Petter et al.^[12], the ISS model is applicable in various of contexts, this model has been successfully applied empirically in e-commerce including stickiness intention^[15], e-learning system^[16], repurchase intention^[17] and cloud office^[18]. This study focuses on the determinants of the success of e-marketplace from seller's perspective, therefore we adopt the updated ISS model as a part of our research model to evaluate the efficiency of e-marketplace.

2.2 Valence framework

The valance framework is also applied to inform the development of our research model. The valence framework is derived from economics and psychology literature, and articulated through summarizing studies

on consumers' purchasing behavior^[19]. This framework was proposed by Peter and Tarpey^[20], they considered that perceived risk/cost and perceived benefit were the two fundamental aspects of individual decision-making. The valence framework has been empirically proved to be a valid model for e-commerce environment. For instance, Kim et al.^[19] introduced trust and satisfaction into the valence framework to study consumers' behavior in e-commerce; Mou et al.^[21] covered trust beliefs and behavior intention with the valence framework in e-health services. Lu et al.^[22] incorporated payment trust into their study about mobile payment. More recently, valence framework has been employed as a theoretical background to study CBEC consumer behavior^[23]. The extended framework proposed by Kim et al.^[19] is the most relevant model to our research. The ISS model is primarily dealing with the characteristics of information system itself to see the impact of usage and evaluate the system, whereas the valence framework is mainly focusing on the perception of user from both cost and benefit aspects to make a decision. Hence, we apply both models in our research, on one hand to avoid any unexpected negligence, on the other hand to make our result robust from both e-marketplace itself and seller aspects.

3. EXPLORATORY STUDY

3.1 Research method

Following Venkatesh et al.^[24], we carried out a mixed-method study by combining both qualitative and quantitative approaches.

We first conducted interviews to identify the appropriate components of perceived benefits, perceived costs, system quality, and service quality. Because the IS success model and valence framework were two separate evaluation models, we conducted two separate sets of interviews with the first focused on the IS success model and the second focused on the valence framework. We were worried about the perfunctory responses due to the long interview duration time if we explored both frameworks at the same time.

The first set of interviews was focused on the valence framework. Although the dimensions of benefits and costs have been explored previously in e-commerce setting^[22], they are more often focused on a buyer's perspective. Given the complex transaction nature of CBEC, the potential benefits and costs may be different from the seller's perspective. Therefore, qualitative interviews were adopted to identify the potential benefits and costs of conducting CBEC business from the seller's perspective.

We conducted interviews with 14 randomly selected sellers who have been selling products through the cross-border e-marketplace for at least 3 years. They all use multiple cross-border e-marketplaces to conduct their business which means that they have enough knowledge and experience to answer our questions. Therefore, the results are considered to be reliable and valid. The interviews were conducted by the authors of this study one-on-one in an informal environment for approximate 10 minutes. The interviewees were asked to answer two open questions associated with our research topic, that were, what benefits you may perceive when you engage in CBEC and what cost or risk you may perceive when you engage in CBEC.

Another set of interviews was conducted with 14 other sellers to identify the key factors associated with IS success and seller's satisfaction. In particular, the predictors of behavior intention have been considered as system quality and service quality, and questions probed seller's complaints about the CBEC platform.

In the next step, the records and transcripts from interviews were open and axial coded following Corbin and Strauss^[25]. The open coding process was conducted by one author, and the concepts extracted from the transcripts are identified. Afterward, these concepts were grouped into categories which reflected the commonalities to reduce the number of concepts. We then grouped these concepts into categories which reflected the commonalities to reduce the number of concepts.

3.2 Data analysis and results

As a result of the first set of interviews, 22 concepts and 9 categories were identified from open coding process. Similar concepts were grouped into the same category which were future classified according to their properties (Table 1). In our study, there are 5 categories for perceived benefit and 4 categories for perceived cost respectively.

Table 1 Coding results for perceived benefit and perceived risk.

Domain	Category	Concepts	Frequency	Percentage
Financial benefit	Profits	High profit	14	100%
	Sales volume	High order volume, High sales volume, High customer volume, Stable customers	12	86%
Product benefit	Brand	Proprietary brand development	4	29%
Strategic benefit	Trend	Long-term development, Development tendency	3	21%
Marketing benefit	Competition	More market, Less competition	4	21%
Financial cost	Monetary loss	Chargeback, Costly rent, Sales return	7	50%
Logistic cost	Logistic issue	Long duration of logistics, Costly logistics, High packet loss probability, Costly customs clearance	7	50%
Marketing cost	Market trends	Unpredictable foreign markets, Difficult inventory control	6	43%
Product cost	Patent dispute	Patent infringement	6	43%

The result of coding data from the second set of interviews is shown in Table 2. 5 system quality and 1 service quality categories were identified. The system quality factors related to ease of uploading and managing products on the site, while the service quality factors related to training and security, among other issues.

Table 2 Summary of factors influencing satisfaction of usage.

Dimension	Category	Comments
System quality	Products upload	Rigid uploading template, Screen stuck, Bulk uploading problem, Slow uploading, Adding draft saving function, Limited categories and specifications, Table editing deficiency
System quality	Product management	Auto push function for overdue product, Auto push function for safe stock
System quality	Order management	Adding remark function, Adding order sort function by dispute time, Adding fuzzy search function, Adding detail export and detail search function
System quality	Logistics	Limited logistics options
Service quality	E-marketplace service	Few train, Bad customer service
System quality	Others	Adding sub-account authority, Push function and control mechanism for malicious buyers.

4. CONFIRMATORY STUDY

4.1 Research model and hypothesis development

The research model of this study is based on the ISS model and the valance framework (Figure 1). According to the results of exploratory study, we utilized categories from table 1 and table 2 as the items of system quality, service quality, perceived benefit, and perceived cost to generate figure 1.

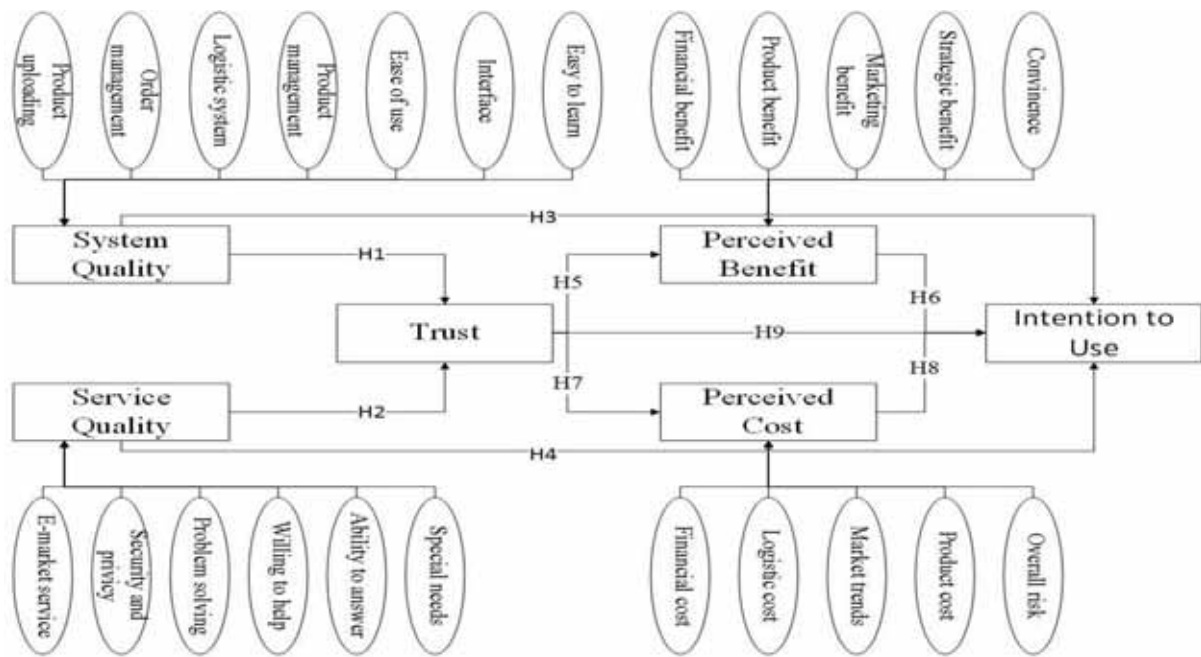


Figure 1 The research model.

Trust in the context of online shopping refers to one's subjective belief that entity on the internet will fulfill its obligations ^[19]. Trust is generally built based on the information from small signals, symbols, or cues provided by the trusted party. Different from shopping in traditional stores, in online marketplace, people cannot physically interact with sellers, therefore, the trust is built by perceived website quality. According to Wang et al.'s study ^[15] on group-buying websites showed the significant relationships of system quality, and service quality with the users' trust beliefs and usage intention. Numerous e-commerce studies have indicated that system quality and service quality can positively influence users' perceived value, meanwhile, users' perceived value can facilitate the formation of users' trust. Therefore, we hypothesize that:

H1. System quality positively influences sellers' trust in e-marketplace.

H2. Service quality positively influences sellers' trust in e-marketplace.

Intention of usage in this study is regard to seller's willing of use a certain e-marketplace. The updated ISS model proposed by DeLone and McLean ^[14] showed that information-related, system-related, and service-related quality could positively influence the intention of use. This model and assertion were confirmed by many researches and also meta-analysis of ISS model. In our case of cross-border e-marketplace, we also believe that website quality (i.e. system quality and service quality) have positive effects on the intention of use a certain e-marketplace. We hypothesize that:

H3. System quality positively influences sellers' intention of using a certain CBEC platform.

H4. Service quality positively influences sellers' intention of using a certain CBEC platform.

Perceived benefit in this study refers to seller's subjective perceptions about the potential values of selling their goods in a certain e-marketplace. As claimed by Hadaya ^[26], market efficiency was increased through market aggregation which provided opportunities for both buyers and sellers with lower transaction cost. Therefore, increase both buyer's and seller's intention to use is significant. Empirical evidence suggests that there is a positive relationship between trust and a variety of benefits. Kim et al. ^[19] suggested that customers could save their cost and comparison process or even increase their productivity when transacted with trusted sellers. Thus, all these benefits would be perceived by trusting customers, and these customers also believed that trusted sellers would fulfill their obligations. Moreover, as suggested by Kim et al. ^[27], online customers were more likely to make a transaction in trusted website with their perception of benefit. Similarly, sellers in CBEC

believe that the trusted platform can fulfill its obligations for sellers and perceived more benefits compared to alternative platforms. They can increase their market reach, penetrate international markets and build their brands while lowering costs of transacting with international buyers on a trusted platform. Consequently, they are willing to do their business via trust platform because of more benefits they perceived. So, we hypothesize that:

H5. Seller's trust positively influences perceived benefit in CBEC platform.

H6. Perceived benefit positively influences sellers' intention of using a certain CBEC platform.

Perceived cost refers to sellers' subjective perceptions about the potential uncertainties or negative values of selling their goods in a certain e-marketplace. It has been investigated several potential uncertainties within e-commerce including financial costs, product costs, information costs^[27]. Moreover, uncertainties have also been emphasized as a realistic issue in CBEC^[23]. When e-commerce goes toward globalization, it must face much more barriers such as cultural difference, language translation, legal issues, geographic issues and financial issues^[7]. Hence, sellers are more sensitive to perceived cost with e-marketplace on the solution of some issues, for example, credit card charge back, security, customs clearance and return cost. Because of these issues are not existed in a brick-and-mortar retail store like Wal-Mart where sellers can directly interact with buyers, so e-commerce seller's perception of cost or experience of using an e-marketplace depends more on seller's trust. Previous studies have suggested that trust reduce the sense of risk^{[19] [30]}. For buyers, as trust increase, they may act more risk-taking behavior, and engage in a risky relationship with the vendor^[19]. Similarly, if sellers trust a CBEC platform, they may perceive less cost and interact more with this platform rather than alternatives. Further, perceived cost has also been found that negatively influence a customer's online decision^[28]. Unpredictable demand, hidden delivery costs, the potential for product return and chargeback, and buyer protection that disadvantages sellers will detract from sellers' trust and engagement on cross-border platforms. Thus, we hypothesize that:

H7. Seller's trust negatively influences perceived cost in CBEC platform.

H8. Perceived cost negatively influences sellers' intention of using a certain CBEC platform.

Sellers' trust built by whether their perception of net valance or website quality itself can enhance the intention behaviors. On one hand, trust can amplify potential benefits and increase the tolerance of perceived costs^[27], on the other hand, trust act as a mediator between website quality and intention behaviors^[15]. Empirical evidence shows that trust significantly influence behavioral intention in e-commerce setting^{[29] [30]}. Whether buyers or sellers, no matter where their trust are built (quality perceptions), they wish to transact with trusted parties to save their cost and increase their benefit. Therefore, we hypothesize that:

H9. Trust positively influences sellers' intention of using a certain CBEC platform.

4.2 Operationalizing and testing the research model

We have identified all new items needed for the following study through qualitative research. All six constructs in our research model are measured with multiple items and in a 5-point Likert-scale ranging from one (strongly disagree) to five (strongly agree). Most of items of system quality, service quality, perceived benefit and perceived cost are from qualitative study of this research, whereas other items will be adapted from prior validated scales. As our questionnaire is originally in Chinese, we will conduct a back-translation procedure to ensure translation validity to English.

In the next step, we will collaborate with a Chinese CBEC platform and distribute our questionnaire to active sellers randomly via firm's e-mail system in three steps. There will be two rounds of data collection in our study and we will select all sellers as our targets who have not participated in the platform satisfaction survey within the previous three months period. In the first round, we are aiming to test our self-developed items from

the result of qualitative study by a pilot test. Afterwards, we will make some changes according to the result of pilot test to constitute our final version survey. We will administer the main survey over a 10 days period to allow for sufficient participation in the survey. The participation of sellers will be totally voluntary without any loss if they refuse to participate and their anonymity will be ensured. We will adopt structural equation modeling (SEM) by AMOS to test our hypothesis (structure model).

5. CONCLUSIONS

In the current study, we have proposed an integrated research model covering technical and perceptive aspects of CBEC from sellers' perspectives. Through the qualitative part of this research, we have identified the key factors which sellers are concerned about, and why they engage in CBEC. We have made a theoretical contribution by developing a comprehensive model for CBEC by combining two popular theories in e-commerce. Secondly, to operationalize the model, we have developed new dimensions with associated items for system quality, service quality, perceived benefit and perceived cost in the context of CBEC. Thirdly, we will test our hypothesis and research model using structural equation modeling. Practically, results of this study may help cross-border platforms better understand the needs and concerns from the seller's perspective who are generally under-represented in the literature but are crucial for the success of CBEC business.

ACKNOWLEDGEMENT

This study supported by the Fundamental Research Funds for the Central Universities of No. JBX180604; and the Chinese Ministry of Education Humanities and Social Sciences project of No. 16YJC630038; and the National Natural Science Foundation of China of No. 71573199..

REFERENCES

- [1] Najberg A, Erickson J. (2015). Cross-border e-commerce to reach\$1 trillion in 2020. <http://www.alizila.com/cross-border-e-commerce-to-reach-1-trillion-in-2020/>.
- [2] Ho S C, Kauffman R J, Liang T P. (2007). A growth theory perspective on b2c e-commerce growth in Europe: An exploratory study. *Electronic Commerce Research & Applications*, 6(3): 237-259.
- [3] Herrera E G, Martens B, Turlea G. (2013). The drivers and impediments for cross-border e-commerce in the EU. *Information Economics & Policy*, 28(1): 83-96.
- [4] Mooney T. (2016). E-commerce, FTZ buoy Tianjin recovery. https://www.joc.com/port-news/asian-ports/e-commerce-ftz-buoy-tianjin-recovery_20160415.html
- [5] Cho H, Lee J. (2017). Searching for logistics and regulatory determinants affecting overseas direct purchase: an empirical cross-national study. *Asian Journal of Shipping & Logistics*, 33(1): 11-18.
- [6] Alt R, Zimmermann H D. (2016). Electronic markets on electronic markets in education. *Electronic Markets*, 26(4): 311-314.
- [7] Laudon K C, Traver C G. (2015). *E-commerce: Business, Technology, Society* (11th global ed.). Pearson education.
- [8] Terzi N. (2011). The impact of e-commerce on international trade and employment. *Procedia Social and Behavioral Sciences*, 24: 745–753
- [9] Turban E, Whiteside J, King D, Outland J. (2017). Implementation issues: From globalization to justification, privacy, and regulation. *Introduction to Electronic Commerce and Social Commerce*, 383-413.
- [10] Guo Y, Bao Y, Barnes S J, Le K. (2018). To sell or not to sell: exploring sellers' trust and risk of chargeback fraud in cross-border electronic commerce. *Information Systems Journal*, 28(2), 359-383.
- [11] Koh T K, Fichman M, Kraut R E. (2012). Trust across borders: buyer-supplier trust in global business-to-business e-commerce. *Journal of the Association for Information Systems*, 13(11), 886-922.

- [12] Petter S, DeLone W, McLean E. (2008). Measuring information systems success: Models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, 17(3): 236-263.
- [13] DeLone W H, McLean E R. (1992). Information systems success: the quest for the dependent variable. *Journal of Management Information Systems*, 3(4): 60-95.
- [14] DeLone W H, McLean E R. (2003). Model of information system success: A ten-year update. *Journal of Management Information Systems*, 19(4): 9–30.
- [15] Wang W T, Wang Y S, Liu E R. (2016). The stickiness intention of group-buying websites: the integration of the commitment-trust theory and e-commerce success model. *Information & Management*, 53(5): 625-642.
- [16] Chen H J. (2010). Linking employees' e-learning system use to their overall job outcomes: an empirical study based on the is success model. *Computers & Education*, 55(4): 1628-1639.
- [17] Fang Y H, Chiu C M, Wang E T G. (2011). Understanding customers' satisfaction and repurchase intentions: an integration of is success model, trust, and justice. *Internet Research*, 21(4): 479-503.
- [18] Yoon S J, Kim M Y. (2015). A study on the use activation of the cloud office with focus on DeLone and McLean is success model. *Journal of Economic History*, 14(2): 289-314.
- [19] Kim D J, Ferrin D L, Rao H R. (2009). Trust and satisfaction, two stepping stones for successful e-commerce relationships: A longitudinal exploration. *Information Systems Research*, 20(2): 237-257.
- [20] Peter J P, Tarpey L X. (1975). A comparative analysis of three consumer decision strategies. *Journal of Consumer Research*, 2: 29-37.
- [21] Mou J, Shin D H, Cohen J. (2016). Health beliefs and the valence framework in health information seeking behaviors. *Information Technology & People*, 29(4): 876-900.
- [22] Lu Y, Yang S, Chau P Y K, Cao Y. (2011). Dynamics between the trust transfer process and intention to use mobile payment services: A cross-environment perspective. *Information & Management*, 48: 393-403.
- [23] Mou J, Cohen J, Dou Y, Zhang B. (2017). Predicting buyers' repurchase intentions in cross-border e-commerce: A valence framework perspective. In proceedings of the 25th European Conference on Information Systems (ECIS).
- [24] Venkatesh V, Brown S A, Bala H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, 37(1): 21-54.
- [25] Corbin J, Strauss A. (2008). *Basics of Qualitative Research*, 3rd ed., SAGE Publications, Thousand Oaks, CA.
- [26] Hadaya P. (2008). Determinants and performance outcome of SMEs' use of vertical B2B e-marketplaces to sell products. *Electronic Markets*, 18(3): 260-274.
- [27] Kim D J, Ferrin D L, Rao H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems*, 44(2): 544-564.
- [28] Antony S, Lin Z, Xu B. (2006). Determinants of escrow service adoption in consumer-to-consumer online auction market: An experimental study. *Decision Support Systems*, 42(3): 1889-1900.
- [29] Bhattacharjee A. (2002). Individual trust in online firms: Scale development and initial test. *Journal of Management Information Systems*, 19(1): 211-241.
- [30] Gefen D. (2002). Reflections on the dimensions of trust and trustworthiness among online consumers. *AcmSigmis Database*, 33(3): 38-53.

Policy Recommendations for Promoting the Development of Cross-Border E-Commerce between China and Central Asian Countries

Tingting Xiao¹, Shizhong Ai², Weili Zhang³

¹²³School of Economics and Management, Xidian University, Xi'an, 710000, China¹

Abstract: As the core area of the entire Belt and Road, Central Asian countries' prosperity has a direct bearing on the smooth implementation of the China's Belt and Road Initiative (BRI). The trade and economic relations between China and Central Asia are developing entirely within the worldwide economic globalization trends. In this research, we analysis the several problems that exist in the development of cross-border e-commerce between China and the Central Asian five countries at the first. And then we put forward four countermeasures for the Chinese government and enterprises investing abroad to promote the cross-border e-commerce transactions between China and the countries along the Belt and Road.

Keywords: Central Asian countries, the Belt and Road, cross-border e-commerce, policy recommendations

1. INTRODUCTION

China's Belt and Road Initiative is a magnificent national strategy for seeking development in China global layout at the new century. The initiative is a political strategy, and followed by an economic policy. With the global e-commerce booming, the development of the Belt and Road cannot be separated from the "Internet Silk Road" construction. According to statistics from *Comprehensive Department of the Ministry of Commerce*, the size of China's import and export trade in 2016 reached 24.3 trillion CNY, down 0.9% ^[1]. Data from *the China Electronic Commerce Research Center* shows that in 2016 China's cross-border e-commerce market rose by 24% year-on-year and reached 6.7 trillion CNY. The proportion of cross-border e-commerce in total turnover has risen from 6% in 2010 to 28% in 2016, and is estimated to hit 38% by 2020 ^[2]. In other words, cross-border e-commerce has become an important driving force for China's foreign trade growth and boosting the implementation of China's Belt and Road Initiative. For countries along the Belt and Road, cross-border e-commerce will increase new vitality into their multilateral security cooperation and provide an opportunity for their economy growing by leaps and bounds.

The Silk Road economic belt has very obvious characteristics in terrain. That is both ends are connected with the prosperous European economic circle and the developed East Asian economic zone respectively. However, in the middle part of the economic belt, there is an economic backward zone between northwestern China and Central Asia ^[3].

Central Asia, as described in this paper, refers to a narrowly defined geographical concept and only includes five countries: Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan and Turkmenistan. The five Central Asian countries, located in the center of Eurasia, are the strongholds of the Belt and Road, guarding the land access in Asia and Europe ^[4]. This region is adjacent to the northwestern of China. And the total length of the common border between China and Kazakhstan, Kyrgyzstan and Tajikistan together is up to 3,309 kilometers ^[5]. From an economic point of view, Central Asia is an economic depression area in Belt and Road and the overall economic development lags behind. In view of the domestic political environment, all five Central Asian countries are in a period of social and economic transition. The government regard domestic economic construction as the primary task at present and formulate series of development strategic plans for their economic development and improving people's livelihood. However, the randomness of the authorities'

* Corresponding author. Email: shizhai@mail.xidian.edu.cn(Shizhong Ai), tingtingxiao@stu.xidian.edu.cn(Tingting Xiao)

intervention in economic policies has greatly fluctuated the investment environment ^[4]. In terms of resource endowment, as one of the world's three major energy centers, Central Asian natural gas reserves reached 11.8 trillion cubic meters, accounting for 6.2% of the world. Proven reserves of 4.28 billion tons of oil, accounting for 2.1% of the world total reserves ^[6]. Non-ferrous and rare metals storage capacities are also substantial. Uranium storage is the highest in the world. In addition, this area also has good conditions for developing oasis agriculture and animal husbandry. And cotton production has unique advantages. Therefore, Central Asia is called Strategic Resources Base in 21st century ^[5]. In trade cooperation perspective, China has become the most important cooperative partner of Central Asian countries in trade and investment now since the establishment of the *Shanghai Cooperation Organization* in 2001. According to the data from *Economic and Commercial Counselors' Office of the Chinese Embassies in Central Asia*, China is the largest trade partner of Kyrgyzstan ^[7], ^[19], Turkmenistan ^[8], ^[20] and Uzbekistan ^[9]. And it is Tajikistan ^[21] and Kazakhstan's ^[22] second largest trade partner, also is Kazakhstan's ^[23] largest source of imports and the largest source of direct investment in other four Central Asian countries except Kazakhstan ^[7], ^[8], ^[11], ^[24], ^[25].

To sum up, the implementation of BRI has helped Central Asian countries' economy to thrive, that will be of great significance to China's geopolitical security and energy security. The rest of this paper is organized as follows: In section two, we investigated three major problems that exist between China and Central Asian countries in cross-border e-commerce transactions through the integration of online data and interview with cross-border e-commerce participants. Then, we put forward some policy suggestions to the government and enterprises for promoting cross-border e-commerce development better in section 3. Section 4 concludes the paper.

2. PROBLEMS IN DEVELOPING CROSS-BORDER E-COMMERCE WITH CENTRAL ASIAN COUNTRIES

We concluded that there are three major challenges of China in the process of cross-border e-commerce trade cooperation with Central Asian countries by reading the literatures published by scholars and related reports. Specifically, Central Asian national information development level is relatively low, its infrastructure construction such as logistics and transportation are lagging behind and also its e-commerce market is still not perfect.

2.1 The level of informatization development in Central Asian countries is generally low.

Telecommunications network connectivity is the basis of the "Internet Silk Road" construction. but this foundation is weak in the countries along the Belt and Road, especially in the five Central Asian countries. For example, in terms of communications network, only Kazakhstan has better domestic communications infrastructure and a higher Internet penetration rate. From the *Internet World Stats'* data, we found that the number of Internet users in Tajikistan accounted for 20.5% of the total population as of June 30, 2017^[18]. This proportion is only higher than Turkmenistan and Afghanistan in Asia. Besides the Tajikistan's Internet speed is slow and the price is expensive, 2M bandwidth Internet access costs about 1,000 US dollars per month ^[10]. Turkmenistan's Internet penetration rate is about 18.0%. There are about 20,000 network users in total now and most of them are group users such as state organs, public institutions and diplomatic organizations ^[8]. In addition, although Kyrgyz informatization development is more advanced than other Central Asian countries, it is more difficult to install fiber-optic cables in mountainous country as here. So more than 70% of Internet users are gathered in the capital Bishkek ^[7]. With regard to electricity, as a result of power shortage of four to five billion kilowatt hours per year, the Tajikistan's government imposed power-cut measures in autumn and winter. The power will be supplied nine hours a day in some areas. In winter 2015-2016, most cities supply electricity

each day for only six to ten hours. Many remote rural residents even only use electricity for two or three hours a day^[10].

All of the above data shows that the overall level of informatization construction in the five Central Asian countries is relatively low. That is reflected in two aspects: the imperfect information infrastructure construction and the insufficient information services development.

2.2 Transport and postal infrastructure in Central Asian countries are underdeveloped.

The main problem in the transportation field of Kazakhstan is the facilities overuse. For railways, locomotives and wagons suffered heavy wear and tear with a loss rate as high as 72%. Almost a third of freight wagons and a half of locomotives have been used for more than twenty-five years. In the aspect of highways, the overall infrastructure is backward. Roads have been not maintained for a long time. Only about 37% of the national highways and 9% of the local highways are in good condition, and most of them are Grade III highways. Freight car's loss rate is about 60%. its service life is more than twelve years^[12].

In Tajikistan, the territory of the mountainous area is up to 93%. Complicated terrain makes road construction more difficult. Statistics show that the total length of the roads in Tajikistan is 13,700 km, but almost all of which were constructed during the Soviet era^[10]. Due to the lack of funds, the Tajik government failed to make the necessary investment in road maintenance. Coupled with the successive years of war and the frequent destruction of natural disasters such as landslides and debris flows, the road was severely damaged and difficult to access. There is only one land border crossing between China and Tajikistan, and it is located in a plateau with an average altitude above 4,000 meters. The customs clearance time is only six months each year because of the bad weather and geography conditions. Furthermore, the port's cargo volume is small and not fully utilized. Therefore, the major bulk commodities trade between China and Tajikistan needs to bypass the other countries such as Kyrgyzstan, Uzbekistan and Kazakhstan. It will take more time, cost more money and be more vulnerable to neighboring countries^[10].

Turkmenistan's domestic postal infrastructure is also developing slowly. There are only about one hundred post offices in whole country and courier parcels are usually not delivered on time^[8].

The main problem in Kyrgyz railway network is the severe aging of infrastructure, the average aging rate of the facilities is about 50% to 55%. Among them, the electricity system equipment aging rate is 50%, signal and communication system aging rate is 63%, road maintenance equipment aging rate hit 71%, sleeper aging rate is about 50%. There are more than 30 percent of the railway tracks are overload and over 86 percent of diesel locomotives exceed their service life^[13].

From the above, the long absence of road maintenance and the lack of funding for construction have strongly restricted the timeliness of logistics in the cross-border e-commerce activities between China and Central Asia, and then affected the Central Asian countries foreign trade.

2.3 The development of cross-border e-commerce market in Central Asian countries is not perfect.

From the perspective of the cross-border e-commerce regional market, *China Cross-Border E-Commerce (Export B2B) Development Report 2016* data displays that China's cross-border e-commerce products are mainly exported to European and American area represented by the United States and Canada. Also to the Oceania region, which represented by Australia and New Zealand. While the global emerging markets such as ASEAN, Russia and India all are booming in an all-round way, the cross-border e-commerce market in Central Asia, especially in the five Central Asian countries, continued to weaken.

From the view of product categories in cross-border e-commerce transactions, the monitoring data of *China Electronic Commerce Research Center* shows that the China's major product categories of cross-border

e-commerce exports in 2015 are shown in Table 1. According to the statistics of *China Customs*, the main commodities imported by Central Asia from China are mechanical and electrical products, boiler machinery, rubber products, nuclear reactors, etc. Followed by clothing, shoes and leggings, travel goods and so on. What needs to be clarified is that Central Asian major demand (mechanical and electrical products, nuclear reactors etc.) is not traded through cross-border e-commerce and do not be included in categories of cross-border e-commerce exports. So as for the product category, the demand and supply did not match well.

Table 1. Distribution of China's cross-border e-commerce exports in 2015

Categories	Proportion	Categories	Proportion
3C electronic product	37.7%	clothing	10.2%
outdoor product	7.5%	health and beauty product	7.4%
jewelry	6%	household and horticultural product	4.7%
shoes and bag	4.5%	baby product	3.6%
auto parts	3.1%	lighting	2.8%
Security Monitoring	2.2%	others	10.3%

From the aspects of the Internet corporate development in Central Asia, by March 2017, they officially launched Qoovee.com. It's the first unified cross-border B2B e-commerce platform in the Commonwealth of Independent States (CIS) with only 14,000 registered Users and not yet formed a scale^[14]. At present, the top ten websites of the Central Asian user visits are all foreign companies' websites, and there is not even a local website. This demonstrates that the Central Asian domestic Internet industry development tends to be hollow^[15] and cross-border e-commerce market lacks motivation.

3. POLICY RECOMMENDATIONS FOR PROMOTING SYNERGETIC DEVELOPMENT OF CROSS-BORDER E-COMMERCE BETWEEN CHINA AND CENTRAL ASIAN COUNTRIES.

In the last part, we've explored the bottleneck factors which restricting the coordinated development of cross-border e-commerce between China and Central Asian countries. Then, we will propose the following four policy-related suggestions to Chinese government and enterprises.

3.1 Cooperation with Central Asian countries to carry out infrastructure construction.

As mentioned earlier, the inadequacy of infrastructure construction such as communications networks, postal services, transportation or border crossings is an important factor that restricts the trade between China and the Central Asian countries in the process of cross-border e-commerce cooperation. In addition to focusing on our cooperation partners in which with better infrastructures and avoiding wasting our cross-border e-commerce resources in areas where infrastructure construction is particularly poor. What's more, we should give full play to the lead role of the government and take the form of "government cooperative investment, enterprise construction" to participate in the Central Asian countries' infrastructure rebuilding. Helping them to improve their backward infrastructure and foster cross-border e-commerce market. The establishment of the Asian Infrastructure Investment Bank (AIIB) has also alleviated some of the pressure of the enterprises investment risks and financing difficulties in infrastructure. With the support of the AIIB and Silk Road Fund, Chinese government should speed up the evolution of infrastructure and industrial cooperation with Central Asian countries and promote the integration of information applications. Up to now, the Silk Road Fund has signed 15 projects, committed to a cumulative investment of 6 billion U.S. dollars. In addition, it has established a special fund for China-Kazakhstan capacity cooperation with an investment of 2 billion U.S. dollars^[26].

Gradually, China and Central Asia will work together to establish a network of infrastructure that radiates the major Asian regions and lay the foundation for the cross-border e-commerce exchange and other forms of bilateral trade.

It should pay more attention to the technical supports beyond the investment in equipment and funds. For example, Chinese enterprises should regularly formulate training programs with Central Asian countries' transport professionals and managers to promote mutual communication. Jointly organize training courses on road design concepts, road management and maintenance. This will serve as an intellectual guarantee for promoting the establishment of a unified traffic standard between China and Central Asia. Furthermore, it's also a feasible measure to establish a platform for cross-border transportation information exchange. Chinese enterprises can provide the road planning and advisory services to Central Asian countries through this platform. The online information exchange platform would contribute to reduce the cost of communication between the two parties.

Another point that must be mentioned is that the relevant departments should make investment risk assessment and provide risk pre-warning service to Chinese enterprises that are constructing construction projects abroad. At the same time, Chinese enterprises themselves must strictly abide by the local laws and regulations, strengthen communication with local government to minimize risks ^[4].

We have made some achievements after the government took measures as above policy recommendations. For example, under the framework of the strategic cooperation between China and Kazakhstan, the two sides have reached the intention of 51 capacity cooperation projects, mainly involving transportation infrastructure construction, manufacturing and other fields. The total investment exceeds 26 billion U.S. dollars. Also, the Chinese government has signed agreements or plans with other four Central Asian countries on promoting trade contacts and infrastructure construction. In 2016, there were more than 1,200 columns of China-Europe freight trains transiting Kazakhstan and the amount of rail transport between the two countries reached more than 8.2 million tons. All these have greatly enriched the connotation of China-Kazakhstan economic and trade cooperation ^[22].

3.2 Encourage capable China's Internet companies to "go global" and foster cross-border e-commerce market.

China's Internet industry has enjoyed a good momentum of development. *National Informatization Development Evaluation Report (2016)* data shows that China's national informatization development index ranked 25th among the eighty-eight major countries that participate in the evaluation. The Kazakhstan's index ranked 47th and Kyrgyzstan ranked 63rd. The remaining three Central Asian countries are not on the list. The index evaluates the level of information technology development in a country or region from the five aspects: network infrastructure readiness, industrial and technological innovation, informatization application benefits, cybersecurity assurance and sustainable development. That report also shows that China's Internet companies have obvious comparative advantages in terms of industrial and technological innovation and informatization application benefits. And these two aspects are expected to achieve complementary development with countries along the Belt and Road.

China's Internet companies also have become increasingly competitive in the international arena. Data from *State Administration for Industry and Commerce* indicates that the number of newly registered enterprises in China's information transmission, software and information technology services industry grew from 74,141 in 2013 to 240,413 in 2015, an increase of more than three times. Besides, according to *China Securities Regulatory Commission* statistics, in 2014-2015, the market scale of China Netcom enterprises had increased more than double from 1.493 trillion CNY to 3.184 trillion CNY. And in 2015, four Chinese Internet companies,

namely Alibaba, Tencent, Baidu and JDcom, ranked among the top ten in global market capitalization. Twelve Chinese Internet companies ranked the top thirty in global market capitalization ^[17]. According to *Alexa's* statistics in February 2016, AliExpress, the Alibaba's international trading platform, was ranked in the top 20 websites in the traffic rankings of Kazakhstan and Uzbekistan. AliExpress is the most popular shopping site in the CIS countries except Kyrgyzstan, surpassing Amazon and eBay ^[15]. As a result, Chinese internet companies have a promising future in the Central Asian market.

Nowadays, some of China's Internet companies have the ability to "go global" already. During their cooperation with the Central Asian countries in cross-border e-commerce business, they'll export the advanced experience of the cross-border e-commerce industry innovation and its efficient application in information technology to help them foster a group of competitive cross-border e-commerce enterprises like China's transportation companies and equipment manufacturers, thus stimulating the development of the Central Asian countries' cross-border e-commerce market.

3.3 The processing and marketing of products should be tailored to Muslim purchasing preferences.

China should take the Central Asian countries' religion into more consideration when conduct the cross-border e-commerce cooperation with them. As is well-known that Central Asia is a multi-religious area dominated by Islam, and more than 80% of the inhabitants are Muslims. Islam has a profound mass foundation a very important influence on the political process, economic development and individual's life in the five Central Asian countries ^[16].

It is very important to fully respect the religious beliefs of Central Asia and to consider the consumption preferences of Muslims for China's cross-border e-commerce enterprises. From the past trade with Central Asian countries, we found that the most popular cross-border e-commerce products in Central Asia are clothing, leggings, handicrafts as well as processed agricultural products, etc. Chinese exporters should conduct adequate market research to understand the needs and preferences of Muslim, so as to determine the product's style, color, material and packaging. In particular, for the processing of agricultural products, such as fruit juice, the production process standards in the target country of export should also be considered. At the same time, we must pay attention to cross-cultural marketing issues.

3.4 Using the Belt and Road information platform to promote the exchange of information, capital and talents among regions.

Chinese enterprises should take advantage of various exchanges platforms to collect demand information, display products and cooperate with each other. And further expand the popularity of China's cross-border e-commerce products and services in Central Asian countries. With these exchange platforms, the Chinese government could actively propel the implementation of the China's Belt and Road Initiative and promote the flow of information, capital and talents across the cross-border e-commerce cooperation countries. At the same time, these information platforms could provide some opportunities to enhance the dialogue between China and Central Asia in the cultural field and deepen mutual understanding.

The forms of information platform mentioned above are various ^[4], such as forums, expositions, fairs, festivals and so on. The influential forums include:

- Euro-Asia Economic Forum, establish a permanent venue in Xi'an
- Continental Bridge Forum (International Law Enforcement Cooperation Forum on Secure Corridor of the New Eurasian Land Bridge), set up a permanent venue in Lianyungang.
- Asia-Pacific Trade Facilitation Forum
- Silk Road Mayor Forum

Major festivals and expositions include:

- Silk Road International Festival, is settled permanently in Shaanxi Province.
- Silk Road International Fair
- China-Eurasia Expo, Urumqi.
- Lanzhou Investment and Trade Fair
- Western China International Fair (WCIF)

The 2017 Euro-Asia Economic Forum held in Xi'an has yielded very fruitful results. The organizers held a total of 35 events in four major sections: dialogues, exhibitions, investment negotiations and cultural exchanges. The exhibition area covers an area of 30,000 square meters, 246 domestic and foreign exhibitors, 1,492 international standard booths, attracting 13,000 professional visitors. During the exhibition, 130 intent cooperation projects were signed on-site, more than 500 orders were traded and an expected turnover is about 1.12 billion CNY.

4. CONCLUSIONS

Taking the five Central Asian countries as an example, this study points out that there are mainly three aspects constraints in the process of China's cross-border e-commerce trade with the countries along the Belt and Road by analyzing the related data. Firstly, the level of informatization development in the Central Asian countries is generally low, the degree of network facility readiness and popularity are not high. Secondly, Central Asian countries lag behind in infrastructure construction such as road traffic and postal ports. Thirdly, the development of domestic cross-border e-commerce market in the five Central Asian countries is not perfect. Then, we propose four suggestions for the Chinese government and enterprises against the above three problems and demonstrate some of the efforts that have been achieved. First, the government should encourage the use of state investment and enterprises construction to participate in the infrastructure improvement in the five Central Asian countries. Second, outstanding Chinese Internet enterprises should be encouraged to "go global" and give their own advanced experiences to local enterprises in Central Asia to foster cross-border e-commerce market. Third, Chinese e-commerce exporters should cater to the religious beliefs and preferences of the Central Asia nation in such aspects as product production, packaging and marketing. Fourth, the Chinese government and foreign trade enterprises should make full use of the Belt and Road cooperation platform to promote the flow of information, capital and talents among the Central Asian countries.

ACKNOWLEDGEMENT

This research was supported by both the National Natural Science Foundation of China under Grant 71771184 and the Xi'an Science and Technology Project under Grant SF1502(1).

REFERENCES

- [1] Ministry of Commerce of the People's Republic of China Comprehensive Department. (2017-01-16). December 2016 Import and Export Summary. <http://zhs.mofcom.gov.cn/article/aa/201701/20170102501769.shtml> (in Chinese)
- [2] China Industrial Information. (2017-12-22). Analysis on the Export Situation of China Cross-border E-commerce Industry in 2017. <http://www.chyxx.com/industry/201712/595670.html> (in Chinese)
- [3] Zhu Xianping, Zou Xiangyang. (2006). Idea of New Silk Road Economy Development Zone between China and Central Asia. *Northeast Asia Forum*, 15(5):3-6 (in Chinese)
- [4] Han Lu. (2017). The Silk Road Economic Belt Advances in Central Asia: Achievements and Prospects. *International Studies*, (3):108-124 (in Chinese)

- [5] Zhu Ruixue. (1995-01-25). Research on the Economic and Trade Cooperation between China and the Central Asian countries under the Background of “Silk Road Economic Belt”. PhD Thesis. Liaoning: Dongbei University of Finance and Economics (in Chinese)
- [6] Personal library. (2017-06-21). Distribution and Exploitation of Superior Mineral Resources in Five Central Asian Countries. http://www.360doc.com/content/17/0621/00/4019270_665031589.shtml (in Chinese)
- [7] Ministry of Commerce, Institute of International Trade and Economic Cooperation. (2017). Foreign Investment Cooperation Country / Region Guidelines–Kyrgyzstan. <http://fec.mofcom.gov.cn/article/gbdqzn/> (in Chinese)
- [8] Ministry of Commerce, Institute of International Trade and Economic Cooperation. (2017). Foreign Investment Cooperation Country / Region Guidelines–Turkmenistan. <http://fec.mofcom.gov.cn/article/gbdqzn/> (in Chinese)
- [9] Ministry of Commerce, Institute of International Trade and Economic Cooperation. (2017). Foreign Investment Cooperation Country / Region Guidelines–Uzbekistan. <http://fec.mofcom.gov.cn/article/gbdqzn/> (in Chinese)
- [10] Ministry of Commerce, Institute of International Trade and Economic Cooperation. (2017). Foreign Investment Cooperation Country / Region Guidelines–Tajikistan. <http://fec.mofcom.gov.cn/article/gbdqzn/> (in Chinese)
- [11] Ministry of Commerce, Institute of International Trade and Economic Cooperation. (2017). Foreign Investment Cooperation Country / Region Guidelines–Kazakhstan. <http://fec.mofcom.gov.cn/article/gbdqzn/> (in Chinese)
- [12] Ministry of Commerce of the People's Republic of China. (2013-03-26). Kazakhstan Transport Development Overview. <http://www.mofcom.gov.cn/article/i/dxfw/jlyd/201303/20130300067264.shtml> (in Chinese)
- [13] Han Dong. (2015). Research on Promoting Trade and Investment Facilitation between China and Five Central Asian Countries. PhD Thesis. Xinjiang: Xinjiang University (in Chinese)
- [14] CIETNET. (2017-03-20). Cross-border E-commerce Platform Qoovee.com in Uzbekistan Began Operation. http://www.ec.com.cn/article/hwds/201703/15394_1.html (in Chinese)
- [15] Qiu Jiahui. (2016). The Current Situation of Internet in “One Belt, One Road” Countries. MsD Thesis. Zhejiang: Zhejiang University of Media and Communications (in Chinese)
- [16] Chen Genxia. (2017). Study on Industrial Manufactured Goods Trade Potential between Chinese and the Five Central Asian Countries under the Belt and Road Policy Background. MsD Thesis. Jiangsu: Nanjing University (in Chinese)
- [17] CNNIC. (2016). National Informatization Development Evaluation Report (2016). http://www.cnnic.cn/hlwfzyj/hlwxzbg/hlwjbg/201611/t20161118_56109.htm (in Chinese)
- [18] Internet World Stats. (2017-06). Asia Internet Use, Population Data and Facebook Statistics. <http://www.internetworldstats.com/stats3.htm>
- [19] Economic and Commercial Counselor's Office of the Embassy of the People's Republic of China in Kyrgyzstan. (2017-11-02). January-September 2017 Kyrgyzstan Overview of Socio-Economic Development. <http://kg.mofcom.gov.cn/article/ztdy/201711/20171102664311.shtml> (in Chinese)
- [20] Economic and Commercial Counselor's Office of the Embassy of the People's Republic of China in Turkmenistan. (2016-09-21). Trade Exchanges between Turkmenistan and Three Major Partners. <http://tm.mofcom.gov.cn/article/ztdy/201609/20160901397990.shtml> (in Chinese)
- [21] Economic and Commercial Counselor's Office of the Embassy of the People's Republic of China in Tajikistan. (2015-11-04). Bilateral trade between China and Tajikistan in the first three quarters of 2015. <http://tj.mofcom.gov.cn/article/ztdy/201511/20151101155184.shtml> (in Chinese)
- [22] Economic and Commercial Counselor's Office of the Embassy of the People's Republic of China in Kazakhstan. (2016-02-19). In 2015, the Trade Volume between China and Kazakhstan Reached 10.567 Billion U.S. Dollars. <http://kz.mofcom.gov.cn/article/zxhz/hzjj/201602/20160201258632.shtml> (in Chinese)
- [23] CEW. (2017). January-March 2017 China-Kazakhstan Bilateral Trade Overview. <http://www.cewgroup.cn/2017nian1-3yuehasakesitanhuowumaoyi-jizhonghashuangbianmaoyigaikuang/>

- [24] Economic and Commercial Counselor's Office of the Embassy of the People's Republic of China in Kyrgyzstan. (2018-01-05). January-September 2017, China is Kyrgyzstan's largest source of foreign direct investment. <http://kg.mofcom.gov.cn/article/jmxw/201801/20180102694640.shtml> (in Chinese)
- [25] Economic and Commercial Counselor's Office of the Embassy of the People's Republic of China in Tajikistan. (2018-01-05). China's Investment in Tajikistan Accounts for 47.3% of Total Foreign Investment in Tajikistan. <http://tj.mofcom.gov.cn/article/jmxw/201801/20180102694653.shtml> (in Chinese)
- [26] Phoenix Finance. (2017-05-14). Ten major events witnessed the construction of the Belt and Road. http://finance.ifeng.com/a/20170514/15373335_0.shtml (in Chinese)

Research on The Optimization Strategy of Cross-border B2B Supply Chain with Service Cost Information Sharing

Dan Fang^{1,2}, Wensheng Yang¹, Jun Ma³, Yiliu Tu^{3}*

¹School of Economics and Management, Nanjing University of Science and Technology, China

²School of Economics and Management, Changzhou Institute of Technology, China

³Schulich School of Engineering, University of Calgary, Canada

Abstract: The development of cross-border e-commerce highlights the importance of service integration and information sharing. This paper considers a cross-border B2B supply chain which consists of a local manufacturer and a foreign trade service integrator. The service integrator holds the service cost structure as private information. The optimal decisions and maximum expected profits of the manufacturer and service integrator are analyzed under two scenarios: no information sharing versus information sharing. The paper finds that information sharing always benefit the manufacturer but not for the service integrator. In the meanwhile, the value of information is increased with the manufacturer's forecast uncertainty about the service integrator's service cost. The whole supply chain can get pareto improvement through the Nash bargaining mechanism.

Keywords: Cross-border e-commerce, Supply chain management, Information sharing, Service decision

1. INTRODUCTION

Cross-border e-commerce usually refers to transactions, payments and logistics in different countries through e-commerce ^[1]. As a new type of e-commerce, cross-border e-commerce has been widely spread and developed rapidly under the national strategy background of "One Belt and One Road (B&R)" in China ^[2]. According to the data from China's cross-border e-commerce research center in May 2017, China's cross-border e-commerce import and export reached 6.6 trillion yuan in 2016, Increased by 38.7% over the same period in 2015, and B2B accounted for 88.7% of total transactions in China's cross-border e-commerce.

China's B2B cross-border e-commerce SME (small- and medium-sized enterprises) trade occurs mainly through foreign trade service integrators ^[3]. These all-in-one integrators (e.g., Onetouch.alibaba.com, Globalsources.com, DHgate.com) with integrated services for international logistics, cross-border payments, internet financing, multilingual services, etc. ^[4]. These integrated services can not only enhance the platform experience, but also bring new customers to the SME. For example, DHgate.com is rich in supporting services to 10 million enterprises and individual buyers from nearly 230 countries and regions provided by specialist partners from logistics (UPS, DHL, FedEx, TNT, etc.), to payments (MasterCard, VISA, American Express, etc.), to technology innovation, to internet financing (China Construction Bank, etc.).

Due to the dynamics of resources, customers and services, these service integrators often keep some private information in order to seek more profits, which causes the information asymmetry problem ^[5]. The existence of asymmetric information often leads to the loss of efficiency and brings difficulties to the coordination of the supply chain system. How to collaborate with these service integrators under information asymmetry become a key issue restricting the development of china's cross-border e-commerce.

Cost information asymmetry is ubiquitous in distributed supply chain systems ^[6]. Kumar and Pugazhendhi ^[7] indicated that information sharing is one of the important tools of coordination among firms in a supply chain,

* Corresponding author. Email: paultu@ucalgary.ca (Yiliu Tu), dan.fang1@ucalgary.ca (Dan Fang)

and allow enterprises to refine their strategies of supply chain to maximize their profits. Ketzenberg et al.^[8] reported that the value of information sharing can range from 0 to 30% profit increase.

Information sharing has been extensively studied in previous literature. The early research focuses on the incentives to share information with horizontal competition in oligopoly (Gal-Or^[9]; Li^[10]). Recent literature has been studied different types of information (e.g., demand, cost, inventory) sharing under complex competition. Some papers focused on demand information sharing. For example, Zhang^[11] took the leakage effect into account and suggested that the retailers will not share private demand information with the supplier in Cournot duopoly. Li and Zhang^[12] examined how the level of confidentiality influences firms information sharing decisions. Ha et al.^[13] focus on vertical information sharing under two competing supply chains consider production diseconomies. Jiang and Hao^[14] examined both vertical and horizontal information sharing under different channel structures. Cost-information sharing has also been studied by several scholars. For example, Yao et al.^[15] explored cost information sharing in a supply chain by considering value-added costs as retailer's private information, they indicated that the retailers are not always better off with information sharing. Zhou and Zhu^[16] investigated the incentive for cost-information sharing under Cournot competition. Kostamis and Duenyas^[17] demonstrated the value of both cost and demand information sharing. Zhao et al.^[18] considered a supply chain with two service providers competing for one client. Setak et al.^[19] demonstrated that the manufacturer with cooperative advertising program can encourage the retailers to share their cost information.

In this paper, we study information asymmetry in a cross-border B2B supply chain composed of a local manufacturer and a synthesized service integrator, the actual service cost is known only to services integrator herself unless she shares it with the manufacturer. We aim to address the following problems:

- a) How do supply chain members decide under complete and uncomplete information scenario?
- b) Can service cost information sharing improve supply chain profitability?
- c) Is it possible to induce service integrator to sharing information?

The remainder of this paper is organized as follows. In Section 2, we present the problem and the assumptions of the model. In Section 3, we present the decision under no information sharing case. In Section 4, we present the decision under complete information. In Section 5, we analyze the model and discuss a profit distribution mechanism. In Section 6, we conclude the paper and give the future research directions.

2. MODEL DESCRIPTION

We consider a cross-border B2B supply chain composed of a local manufacturer and a cross-border service integrator (SI). The SI provide service package at price p_s to foreign buyers, the service package includes pre-sales consulting, supplier selection, supplier qualification certification, logistics and route selection, financing loans, export tax rebates, etc. The manufacturer produces a single product at a unit cost c_m and sells his product to foreign buyers at a price p_m through SI, and need to pay SI a membership fee K . The assumption reflects the practice of cross-border service platforms (e.g., Globalsource.com, Alibaba.com, Made-in-China.com, etc.) who use the membership mechanisms. We assume that the membership fee K is fixed and exogenous to our model. We note that in this paper, the fixed membership fee K does not have any effect on the results, then K is not included in our model analysis and discussion.

Let D_m, D_{si} denote the consumer demand from the manufacturer and SI, respectively. We utilize a linear demand function which is widely used in supply chain literature (Tsay and Agrawal^[19]; Dan et al.^[5]). The corresponding demand functions to the manufacturer and the SI are described as follows:

$$D_{si} = a - b_1 p_s + \gamma_1 s \quad (1)$$

$$D_m = a - b_1 p_s - b_2 p_m + \gamma_1 s + \gamma_2 s \quad (2)$$

Where $a > 0$ corresponds to market potential; $b_1 > 0$ is the coefficient of service price elasticity of D_s ; $b_2 > 0$ is the coefficient of price elasticity of D_m . The parameters γ_1 is the SI's services sensitivities of the service demand and γ_2 are the free-ride services sensitivities to the manufacturer ($\gamma_1 > \gamma_2 > 0$). s is the service package level determined by SI. We use the service package to represent the all forms of SI's services together which is widely used in previous literature (Li et al. ^[20]; Dan et al. ^[21]). The conditions $D_{si} \geq 0$, $D_m \geq 0$ are specified to ensure that both demand non-negative.

The SI's service cost function is generally assumed to be convex with a commonly adopted form ^{[5], [15]} as below:

$$C_{si} = \frac{1}{2} \eta s^2 \quad (3)$$

Where η is an efficiency parameter for the SI's cost, $\frac{dC_{si}}{ds} > 0$, $\frac{d^2C_{si}}{ds^2} > 0$. this cost parameter is assumed to be the private information, known only to the SI. Many streams of research find that firms may protect their advantage by hiding their cost information, and they are reluctant to reveal their private information for the risk of information leakage ^{[11], [12]}. From the manufacturer's point of view, this parameter is a random variable with a known probability distribution $f(\eta)$.

With the above notation, the SI's and the manufacturer's profit function are given by

$$\Pi_M = (p_m - c_m) D_m \quad (4)$$

$$\Pi_{SI} = p_s D_{si} - \frac{1}{2} \eta s^2 \quad (5)$$

There is no leader in this supply chain. The manufacturer and SI make price decisions independently of each other. Thus, the SI determines its price p_s and service level s to maximize its profit Π_{SI} . Uninformed manufacturer determines its price p_m to maximize its profit Π_M .

For ease of reference, we add super script ND, SD on the equilibrium outcomes, where ND indicates that SI neither share signals nor cooperate in decision making, SD indicates that SI share signals and independently make decisions. We will contrast the optimal policies of both channel partners in the two cases: incomplete information when the manufacturer didn't get the exact cost information (Section 3), and information sharing when the manufacturer knows η exactly (Section 4).

3. THE CASE OF NO INFORMATION SHARING (ND)

In this section, we solve the game-theoretic problem for the case of no information sharing. The actual service cost is known only to SI herself unless she shares it with the manufacturer. The manufacturer does not know the exact value of η , but knows its distribution. According to Yao et al. ^[15], we assume that η is uniformly distributed, i.e., $\eta \sim U[\bar{\eta} - \varepsilon, \bar{\eta} + \varepsilon]$. $\bar{\eta}$ and ε are the average and deviation of the service cost efficiency, respectively.

The manufacturer and the SI make their decisions to maximize their own profit respectively. The model in no information sharing case is presented as follows:

$$\left\{ \begin{array}{l} \max_{p_m} E(\Pi_M^{ND}) = \int_{\bar{\eta}-\varepsilon}^{\bar{\eta}+\varepsilon} (a - b_1 p_s + \gamma_1 s - b_2 p_m + \gamma_2 s) (p_m - c_m) f(\eta) d\eta \\ \quad = \frac{1}{2\varepsilon} \int_{\bar{\eta}-\varepsilon}^{\bar{\eta}+\varepsilon} (a - b_1 p_s + \gamma_1 s - b_2 p_m + \gamma_2 s) (p_m - c_m) d\eta \\ \max_{p_s, s} \Pi_{SI}^{ND} = p_s D_{si} - \frac{1}{2} \eta s^2 \\ \text{s.t. } D_{si} > 0, D_m > 0 \end{array} \right. \quad (6)$$

Proposition1: Under the no information sharing case, the manufacturer's optimal prices, SI's optimal service price and optimal service level can be given as:

$$p_m^{ND} = \frac{a(\gamma_1^2 + 2\gamma_1\gamma_2) \ln \left[\frac{2b_1(\bar{\eta} + \varepsilon) - \gamma_1^2}{2b_1(\bar{\eta} - \varepsilon) - \gamma_1^2} \right] + 4b_1(2c_m b_2 + a)\varepsilon}{16b_1 b_2 \varepsilon} \quad (7)$$

$$p_s^{ND} = \frac{\eta a}{2b_1 \eta - \gamma_1^2} \quad (8)$$

$$s^{ND} = \frac{\gamma_1 a}{2b_1 \eta - \gamma_1^2} \quad (9)$$

$$\text{Where } 2b_1(\bar{\eta} - \varepsilon) - \gamma_1^2 > 0$$

Proof:

Suppose the condition $2b_1(\bar{\eta} - \varepsilon) - \gamma_1^2 > 0$ satisfy, since $\bar{\eta} - \varepsilon < \eta < \bar{\eta} + \varepsilon$, which indicates that $2b_1\eta - \gamma_1^2 > 0$.

The hessian matrix $H = \begin{bmatrix} -2b_1 & \gamma_1 \\ \gamma_1 & -\eta \end{bmatrix}$ is negative definite when the conditions $2b_1\eta - \gamma_1^2 > 0$ satisfy, so Π_{SI}^{ND} is jointly concave with p_s, s .

The optimal decisions for SI are obtained by solving the first-order conditions:

$$\begin{cases} \frac{\partial}{\partial s} \Pi_{SI}^{ND} = -\eta s + p_s \gamma_1 = 0 \\ \frac{\partial}{\partial p_s} \Pi_{SI}^{ND} = -2b_1 p_s + \gamma_1 s + a = 0 \end{cases} \quad (10)$$

Further substituting p_s^{ND} and s^{ND} into the manufacturers' profit functions in Eqs. (6), we get $\frac{d\Pi_M^{ND}}{dp_m^2} = -2b_2 < 0$, by solving the first-order conditions, we can obtain the manufacturers' optimal prices p_m^{ND} .

Corollary 1. The optimal profits of the manufacturer and SI are:

$$\Pi_M^{ND} = \frac{(a(\gamma_1^2 h + 2\gamma_1\gamma_2 h + 4b_1\varepsilon) - 8b_1 b_2 c_m \varepsilon)g}{256b_1^2 b_2 (b_1 \eta - \frac{\gamma_1^2}{2})\varepsilon^2} \quad (11)$$

$$\Pi_{SI}^{ND} = \frac{\eta a^2}{4b_1 \eta - 2\gamma_1^2} \quad (12)$$

$$\text{Where } g = a[\frac{h\gamma_1^4}{2} + h\gamma_1^3\gamma_2 - b_1(h\eta - 2\varepsilon)\gamma_1^2 - 2b_1\gamma_1\gamma_2(h\eta - 4\varepsilon) + 4b_1^2\eta\varepsilon] + 4b_1 b_2 c_m \varepsilon(2b_1\eta - \gamma_1^2)$$

$$h = \ln \left[\frac{2b_1(\bar{\eta} + \varepsilon) - \gamma_1^2}{2b_1(\bar{\eta} - \varepsilon) - \gamma_1^2} \right]$$

4. THE CASE OF INFORMATION SHARING (SD)

In this section, we solve the game-theoretic problem for the case where the manufacturer has complete information about the SI's service cost parameter η . The model in information sharing case is presented as follows:

$$\begin{cases} \max_{p_m} E(\Pi_M^{SD}) = (a - b_1 p_s + \gamma_1 s - b_2 p_m + \gamma_2 s)(p_m - c_m) \\ \max_{p_s, s} \Pi_{SI}^{SD} = p_s D_{si} - \frac{1}{2} \eta s^2 \\ \text{s.t. } D_{si} > 0, D_m > 0 \end{cases} \quad (13)$$

Proposition2: Under the information sharing case, the manufacturer's optimal prices, the SI's optimal service price and optimal service level can be given as:

$$p_m^{SD} = \frac{a(b_1\eta + \gamma_1\gamma_2) + c_m b_2(2b_1\eta - \gamma_1^2)}{2b_2(2b_1\eta - \gamma_1^2)} \quad (14)$$

$$p_s^{SD} = \frac{\eta a}{2b_1\eta - \gamma_1^2} \quad (15)$$

$$s^{SD} = \frac{\gamma_1 a}{2b_1\eta - \gamma_1^2} \quad (16)$$

Proof:

Following the same procedure as what is done in the proof of Proposition1, we have the result.

Corollary 2. The optimal profits of the manufacturer and SI are:

$$\Pi_M^{SD} = \frac{(a(b_1\eta + \gamma_1\gamma_2) + c_m b_2(\gamma_1^2 - 2b_1\eta))^2}{4b_2(2b_1\eta - \gamma_1^2)^2} \quad (17)$$

$$\Pi_{SI}^{SD} = \frac{\eta a^2}{4b_1\eta - 2\gamma_1^2} \quad (18)$$

Above, we have obtained closed form solutions for all decision variables and the respective profits of both parties in terms of the market parameters. The difference in profits in these two cases will give us the value of the information to the manufacturer. We will discuss these models in more detail in the next section.

5. DISCUSSION

5.1 The value of information sharing

In this subsection, we study the interesting question of how the profit of each channel partner changes from the incomplete information case to the complete information case.

Proposition 3. (i)The manufacturer's profit will always increase when the information about the SI's service cost structure is shared with him. The amount of increase $\Delta\Pi_m$ is given by:

$$\Delta\Pi_M = \frac{a^2\gamma_1^2(\gamma_1 + 2\gamma_2)^2((2b_1\eta - \gamma_1^2)h - 4b_1\varepsilon)^2}{256b_1^2b_2(2b_1\eta - \gamma_1^2)^2\varepsilon^2} > 0 \quad (19)$$

(ii) The SI does not benefit from the information sharing, her profit is the same as SD and ND scenario,

$$\Delta\Pi_{SI} = \Pi_{SI}^{SD} - \Pi_{SI}^{ND} = 0.$$

(iii) Service cost information sharing can improve the whole supply chain profitability, $\Delta\Pi = \Delta\Pi_M > 0$.

Proof:

By the equilibrium results given in Corollary1 and Corollary 2, we have

$$\Delta\Pi_M = \Pi_M^{SD} - \Pi_M^{ND} = \frac{a^2(2b_1\eta h - h\gamma_1^2 - 4b_1\varepsilon)^2\gamma_1^2(\gamma_1 + 2\gamma_2)}{256b_1^2b_2(2b_1\eta - \gamma_1^2)^2\varepsilon^2}, \text{ It is easy to verify that } \Delta\Pi_M > 0.$$

Obviously, $\Delta\Pi_{SI} = \Pi_{SI}^{SD} - \Pi_{SI}^{ND} = 0$, then the amount of total supply chain profit increase is $\Delta\Pi = \Delta\Pi_M > 0$. Hence the claim.

To examine the effects of service uncertainty ε on the supply members optimal decisions, we take the first-order partial derivatives of $\Delta\Pi$ with respect to ε , and obtain the following proposition.

Proposition 4. The amount of manufacturer's profit increase $\Delta\Pi_m$ is increased with the manufacturer's

estimate uncertainty ε .

Proof:

We rewrite the equation (19) as follows: $\Delta \Pi_M = \left[\frac{a^2 \gamma_1^2 (\gamma_1 + 2\gamma_2)^2}{256 b_1^2 b_2 (2b_1 \eta - \gamma_1^2)^2} \right] \left[\frac{(2b_1 \eta - \gamma_1^2)h - 4b_1 \varepsilon}{\varepsilon} \right]^2$.

let $g = (2b_1 \bar{\eta} - \gamma_1^2)$, $k = (2b_1 \eta - \gamma_1^2)$, $l = \left(\frac{(2b_1 \eta - \gamma_1^2)h - 4b_1 \varepsilon}{\varepsilon} \right)^2 = \left(\frac{kh - 4b_1 \varepsilon}{\varepsilon} \right)^2$, we can obtain $\frac{\partial \Delta \Pi}{\partial \varepsilon}$ by solve $\frac{\partial l}{\partial \varepsilon}$.

$$\frac{\partial l}{\partial \varepsilon} = \frac{2k}{\varepsilon^2} \left[\frac{kh - 4b_1 \varepsilon}{\varepsilon} \right] \left[\frac{(4b_1^2 \varepsilon^2 - g^2)h + 4b_1 g \varepsilon}{(g + 2b_1 \varepsilon)(g - 2b_1 \varepsilon)} \right].$$

$$\text{When } \varepsilon = 0, h = \ln \left[\frac{2b_1(\bar{\eta} + \varepsilon) - \gamma_1^2}{2b_1(\bar{\eta} - \varepsilon) - \gamma_1^2} \right] = \ln \left[\frac{g + 2b_1 \varepsilon}{g - 2b_1 \varepsilon} \right] = \frac{4b_1 \varepsilon}{g} = 0$$

$$\text{When } 0 < \varepsilon < \bar{\eta}, h = \ln \left[\frac{2b_1(\bar{\eta} + \varepsilon) - \gamma_1^2}{2b_1(\bar{\eta} - \varepsilon) - \gamma_1^2} \right] > \frac{4b_1 \varepsilon}{g} > 0, \ln \left[\frac{2b_1(\eta + \varepsilon) - \gamma_1^2}{2b_1(\eta - \varepsilon) - \gamma_1^2} \right] > \frac{4b_1 \varepsilon}{k} > 0$$

$$\text{It can be shown that } \frac{kh - 4b_1 \varepsilon}{\varepsilon} > \frac{k4b_1 \varepsilon - 4b_1 \varepsilon}{\varepsilon} > 0, \frac{(4b_1^2 \varepsilon^2 - g^2)h + 4b_1 g \varepsilon}{(g + 2b_1 \varepsilon)(g - 2b_1 \varepsilon)} = \left[\frac{4b_1 g \varepsilon}{(g + 2b_1 \varepsilon)(g - 2b_1 \varepsilon)} - h \right] > 0, \text{ thus } \frac{\partial l}{\partial \varepsilon} > 0.$$

$$\text{Therefore } \frac{\partial \Delta \Pi}{\partial \varepsilon} > 0.$$

Proposition 4 shows that the more uncertain the manufacturer's estimate about the SI's service cost structure, the higher the value of information. It would be profitable for the manufacturer if he can encourage the SI to share her service cost information, $\Delta \Pi_M$ would be the maximum amount the manufacturer would be willing to spend for the information. Notice that the SI's profit is the same as SD and ND scenario because her optimal service price and service level only depend on her own cost structure, so the SI has no incentives to share her service cost information to the manufacturer. The willingness of SI to share her cost information can be achieved by a profit allocation mechanism. Next section we will use a Nash bargaining model to solve this problem.

5.2 Nash bargaining solution

In this subsection, we want to figure out how to align benefits between members and create motivation to induce service integrator to sharing information.

Let $\Delta \Pi_M^*$, $\Delta \Pi_{SI}^*$ represent the portion of the extra-profit the manufacturer and the SI received at the information sharing scenario, respectively. In order to ensure the success of the information sharing, an optimal profit scheme is acceptable to both the manufacturer and the SI only if $\Delta \Pi_M^* > 0$, $\Delta \Pi_{SI}^* > 0$.

According to Nash^[23], We assume that both the manufacturer and the retailer have the following utility functions:

$$U_M = (\Delta \Pi_M^*)^{\varphi_1} \quad (20)$$

$$U_{SI} = (\Delta \Pi_{SI}^*)^{\varphi_2} \quad (21)$$

Where $\Delta \Pi_M^* + \Delta \Pi_{SI}^* = \Delta \Pi$. φ_1, φ_2 are the bargaining power for manufacturer and SI, respectively. Bargaining power is determined by negotiation skills, risk tolerance and bargaining tactics, etc.

The optimal bargaining profit scheme is obtained by maximizing the system utility function:

$$\begin{cases} \text{Max } U_M U_{SI} = (\Delta \Pi_M^*)^{\varphi_1} (\Delta \Pi_{SI}^*)^{\varphi_2} \\ \text{s. t. } \Delta \Pi_M^* + \Delta \Pi_{SI}^* = \Delta \Pi \end{cases} \quad (22)$$

The optimal solution of this problem is $\Delta \Pi_M^* = \frac{\varphi_1}{\varphi_1 + \varphi_2} \Delta \Pi$, $\Delta \Pi_{SI}^* = \frac{\varphi_2}{\varphi_1 + \varphi_2} \Delta \Pi$.

Where $\Delta\Pi = \frac{a^2\gamma_1^2(\gamma_1+2\gamma_2)^2((2b_1\eta-\gamma_1^2)h-4b_1\epsilon)^2}{256b_1^2b_2(2b_1\eta-\gamma_1^2)^2\epsilon^2}$. Thus, both the manufacturer and the SI can gain from information sharing, which leads to a win-win solution for the whole cross-border B2B supply chain.

6. CONCLUSIONS

In this paper, we present a cross-border B2B supply chain composed of a local manufacturer and a service integrator who face a demand sensitive to product price, service price, and service level. We assume only services integrator has private information about her service cost to reflect the information asymmetry between supply members. The manufacturer is assumed to be the Stackelberg leader, and two scenarios are analyzed: no information sharing case and the complete information case. Our results indicate that information sharing always has a positive impact on the manufacturer's performance but not for the service integrator. The service integrator would be motivated to share its private information with the manufacturer under some profits allocation mechanisms.

Our research can be extended in many two directions. First, we can figure out what kind of contracts can coordinate the supply chain. Second, we can consider the competition among several manufacturers and service integrators.

ACKNOWLEDGEMENT

This research was supported by the MOE (Ministry of Education in China) Project of Humanities and Social Sciences under Grant15YJA630087.

REFERENCES

- [1] Yue Hongfei, Wu Jiani, Yao Lin (2017). Analysis on the development and bottlenecks of China's cross-border e-commerce. *Studies in Asian Social Science*, 4(2): 26-35.(in chinese)
- [2] Xue, Wanxin; Li, Dandan; and Pei, Yilei (2016). The development and current of cross-border e-commerce. *WHICEB 2016 Proceedings*, 53.
- [3] Tao Tao, Li Guangqian (2014). Platform evolution, pattern discrimination and development orientation of cross-border e-commerce, *Reform, China and Globalization*, (9): 63-73. (in Chinese)
- [4] Research group (2017). Analysis on the pathway and risk avoidance of cross border B2B E-commerce platform. *China Business and Market*, 31(11):54-60 (in chinese)
- [5] Dan bin, Lou Yun, Han Xiaopeng, Zhang xumei. (2017). Pricing and coordination strategy in product-service supply chain with service demand affected by sales. *Journal of Systems & Management*, 26(2):381-389. (in chinese)
- [6] Agrell P J, Lindroth R, Norman A(2004). Risk, information and incentives in telecom supply chains. *International Journal of Production Economics*, 90(1):1-16.
- [7] Kumar R S, Pugazhendhi S (2012). Information sharing in supply chains: an overview. *Process Engineer* 38:2147–2154.
- [8] Ketzenberg M E, Geismar N, Metters R, Vander Laan E (2013). The value of information for managing retail inventory remotely. *Production and Operations Management*, 22(4):811–825.
- [9] Gal-Or, E. (1985). Information sharing in oligopoly. *Econometrica*, 53(2): 329–343.
- [10] Li, L. (1985). Cournot oligopoly with information sharing. *RAND Journal of Economic*, 16(4): 521–536.
- [11] Zhang H (2002). Vertical information exchange in a supply chain with duopoly. *Production and Operations Management*, 11(4):531–546.
- [12] Li L, Zhang H (2008). Confidentiality and information sharing in supply chain coordination. *Management*

- science,54(8):1467–1481.
- [13] Ha A Y, Tong S, Zhang H (2011). Sharing demand information in competing supply chains with production diseconomies. *Management Science*, 57(3):566–581.
- [14] Jiang L, Hao Z Y(2016). Incentive-driven information dissemination in two-tier supply shains[J]. *Manufacturing & service operations management*, 18(3): 393–413
- [15] Yao D-Q, Yue X, Liu J (2008) Vertical cost information sharing in a supply chain with value-adding retailers. *Omega* 36:838–851
- [16] Zhou Z Z , Zhu, K X(2010). The effects of information transparency on suppliers, manufacturers, and consumers in online markets. *Marketing Science*, 29 (6): 1125–1137.
- [17] Kostamis D, Duenyas I(2011). Purchasing under asymmetric demand and cost information: when is more private information better? *Operations Research*59(4): 914–928.
- [18] Zhao X, Xue L, Zhang F. (2014). Outsourcing competition and information sharing with asymmetrically informed suppliers. *Production and Operations Management*, 23(10): 1706–1718.
- [19] Setak M, Kafshian Ahar H, Alaei S (2017) Coordination of information sharing and cooperative advertising in a decentralized supply chain with competing retailers considering free riding behavior. *Journal of Industrial and Systems Engineering*, 10:151–168.
- [20] Tsay AA, Agrawal N (2004). Channel conflict and coordination in the e-commerce age. *Production and Operations Management* ,13(1):93-110
- [21] Li X, Li Y, Cai X, Shan J(2015). Service channel choice for supply chain: who is better off by undertaking the service? *Production and Operations Management*, 25 (3): 513–534.
- [22] Dan B, Zhang S G, Zhou M S (2017). Strategies for warranty service in a dual-channel supply chain with value-added service competition. *International Journal of Production Research*, DOI: 10.1080/00207543.2017.1377355.
- [23] Nash J F(1950).The bargaining problem.*Econometrica*,18:155–62.

Empirical Research on the Relationship between Foreign Trade Growth and Cross-Border Electronic Commerce in Fujian Province

Feifei Huang^{1, 2}, Lu Chen²

¹ College of Management, Xiamen University, Xiamen, China

² College of Management, Jimei University, Xiamen, China

Abstract: Based on the of domestic and foreign scholars' researches, foreign trade growth has a booming effect on economic development, otherwise, the role of cross-border electronic commerce developing in international trade business and economic still has not been proved by the actual data available. This essay selects quarterly data of foreign trade and cross-border electronic commerce in Fujian Province who has the advantages of foreign trade, to makeup the VEC model and research the relationship between the growth of foreign trade and cross-border e-commerce, the results show that: there is the long-term interactive relationship between foreign trade growth and cross-border electronic commerce. Cross-border electronic commerce in Fujian province is only starting, needing more positive and long-term policies to promote.

Keywords: Cross-Border e-Commerce, Foreign Trade, Empirical Research, VEC

1. INTRODUCTION

China's foreign trade industry rise rapidly since having joined the world trade organization, which greatly promoted the economic development, but in recent years, the increase of foreign trade has slowed or even stagnated, seeking the transformation and upgrading for industry is a vital issue in the current foreign trade in 2013, cross-border electronic commerce was recognized and concerned widely, for its low cost, fast speed, fewer steps and other specific advantages, it is considered to be a new foreign trade mode, which can reduce costs effectively, create more opportunities of trade and employment. Cross-border electronic commerce has the operational and practical significance in reforming China's industrial structure, taking economic transformation, up grading the foreign trade and our country's economic growth ^[1]. As a result, our country and local governments have instituted many policies on cross-border electronic commerce to support it, but the applicability of these policies needs to be further discussed.

The former studies are mainly about relationship between foreign trade and economic growth, and their actual data, the research on a this new cross-border electronic commerce model is insufficient, at present, a lot of researches have still focused on the promoting role of cross-border electronic commerce to foreign trade, not on the empirical research, this essay provide a theoretical basis for drawing up the policies by choosing the relevant cross-border electronic commerce and foreign trade data in Fujian province, which having the advantages of foreign trade, and also by using the VAR model analysis the interaction mechanism between foreign trade and cross-border electronic commerce.

2. COMPREHENSIVE REVIEWS

Libin E, Yongwen Huang (2014) compared cross-border e-commerce with traditional foreign trade industry by the following five aspects of trade steps, industrial chain, product cost, production and policy guidance, showing that the development of cross-border electronic commerce has better opportunities ^[2], but also face the adjustment, our governments should give more policy supporting. Cheng Chang (2015) analyzed the promoting role of cross-border electronic commerce on the trade increasing from the two aspects of qualitative and quantitative analysis, showing that cross-border electronic commerce will not only promote the development of

foreign trade, but also will benefit from the new opportunities brought by the growth of foreign trade demand in the long time^[3]. Yu (2016) also found that cross-border electronic commerce has played an important role in promoting China's foreign trade by using case analysis^[4]. On the other hand, Ji(2016) analyzed whether the cross-border electronic commerce has a subtle influence on China's foreign trade or not by the Transaction cost theory and Agglomerative economies theory, combined with the Alibaba, and showing that cross-border electronic commerce has not only created the perfect shopping experience but improved China's foreign trade^[5]. Wenjing Yu (2016) found cross-border electronic commerce has a profound effect on China's foreign trade using case analysis method from two aspects of business mode and value chain reconstruction^[6]. Most scholars agree that electronic commerce can promote the development of cross-border electronic commerce, but few scholars have done special researches on whether the trade development can promote cross-border electronic commerce, when and how to draw up the relevant policies. Fengyi Li (2016) analyzed the role of cross-border electronic commerce investment by econometric analysis, and then drew the conclusion that cross-border electronic commerce has the long-term promoting role to foreign trade^[7]. Feng Zuo (2016) analyzed the effect of cross-border electronic commerce on economic growth in Guangdong Province using cointegration, and found the long-term equilibrium relationship between them^[8].

According to the present documents, we can find that the researches on cross-border electronic commerce mainly depended on qualitative analysis, the few quantitative study without focusing on the long-term relationship between cross-border electronic commerce and the trade economic data, and did not research its short-term force, additionally, there has been a large number of study on the relationship between foreign trade and economic growth, the relationship between cross-border electronic commerce and economic growth on the country, there was not many quantitative researches on the relationship between cross-border electronic commerce and foreign trade, therefore, this essay takes cross-border electronic commerce as the research object, trying to make the quantitative analysis the relationship of cross-border electronic commerce and foreign trade.

3. THEORETICAL CONSTRUCTION AND EMPIRICAL ANALYSIS OF VAR MODEL

3.1 Theoretical model ideas seem similar to Alexander's synthesis of form.

The VAR model is an extension of the time series autoregressive model (AR), in which each endogenous variable in the system is constructed as a function of the lagged values of all endogenous variables in the system, describing the endogenous variables of n during the same sample period as a linear function in the past values.^[9] The VAR model constructed in this paper contains three endogenous variables: the total amount of foreign trade in Fujian Province (billion yuan), the amount of cross-border e-commerce transactions in Fujian Province (billion yuan), the number of netizens in Fujian Province people(million people).

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + u_t; \quad Y_t = \begin{bmatrix} LnT \\ LnE \\ LnN \end{bmatrix} \quad (1)$$

3.2 Data selection and unit root test

This paper collects 24 quarterly data of Fujian Province from 2011 to 2016, of which the total amount of foreign trade comes from Bureau of Statistics of Fujian Province^[10], the Cross-border e-commerce, Netizen data are from China E-Commerce Research Center, and metering software chooses STATA and EVIEWS. To remove the heteroscedasticity, we should make the logarithm from original data and begin stationarity tests. The results are as follows: At 5% level, the sequence is integrated of second order.

Table 1. ADF unit root test results

Variable [↗]	ADF Statistic [↗]	1% Critical Value [↗]	5% Critical Value [↗]	10% Critical Value [↗]	P [↗]	Result [↗]
LnT [↗]	-2.2756 [↗]	-3.7696 [↗]	-3.0049 [↗]	-2.6422 [↗]	0.1879 [↗]	unstationarity [↗]
DLnT (-1) [↗]	-1.5565 [↗]	-3.7696 [↗]	-3.00489 [↗]	-2.6422 [↗]	0.4871 [↗]	unstationarity [↗]
DLnT (-2) [↗]	-5.6068 [↗]	-3.7880 [↗]	-3.01236 [↗]	-2.6461 [↗]	0.0002 [↗]	stationarity [↗]
LnE [↗]	-1.5077 [↗]	-3.7529 [↗]	-2.9981 [↗]	-2.6388 [↗]	0.5119 [↗]	unstationarity [↗]
DLnE(-1) [↗]	-2.0869 [↗]	-3.7696 [↗]	-3.0049 [↗]	-2.6422 [↗]	0.2512 [↗]	unstationarity [↗]
DLnE(-2) [↗]	-4.7056 [↗]	-3.7880 [↗]	-3.0123 [↗]	-2.6461 [↗]	0.0013 [↗]	stationarity [↗]
LnN [↗]	-1.63429 [↗]	-3.7880 [↗]	-3.0124 [↗]	-2.6461 [↗]	0.4483 [↗]	unstationarity [↗]
DLnN(-1) [↗]	-3.3591 [↗]	-3.7880 [↗]	-3.0124 [↗]	-2.6412 [↗]	0.0248 [↗]	stationarity [↗]
DLnN(-2) [↗]	-3.4582 [↗]	-3.7880 [↗]	-3.0124 [↗]	-2.6462 [↗]	0.0202 [↗]	stationarity [↗]

3.3 VAR model of the order

According to the information criteria, select LR, AIC and SIC, the comparison results are as follows, select the order of 5.

Table 2. The best lag period to determine the results

滞后阶数 [↗]	LOGL [↗]	LR [↗]	FPE [↗]	AIC [↗]	SC [↗]	HQ [↗]
0 [↗]	52.52874 [↗]	NA [↗]	1.09E-06 [↗]	-5.213551 [↗]	-5.064429 [↗]	-5.188314 [↗]
1 [↗]	173.1107 [↗]	190.3926 [↗]	8.82E-12 [↗]	-16.95902 [↗]	-16.36253 [↗]	-16.85807 [↗]
2 [↗]	181.9553 [↗]	11.17214 [↗]	9.81E-12 [↗]	-16.94266 [↗]	-15.89881 [↗]	-16.766 [↗]
3 [↗]	189.9744 [↗]	7.597058 [↗]	1.39E-11 [↗]	-16.83941 [↗]	-15.34819 [↗]	-16.58704 [↗]
4 [↗]	224.52 [↗]	21.81824 [↗]	1.66E-12 [↗]	-19.52842 [↗]	-17.58983 [↗]	-19.20033 [↗]
5 [↗]	362.1716 [↗]	43.46893* [↗]	8.84e-18* [↗]	-33.07069* [↗]	-30.68474* [↗]	-32.66690* [↗]

3.4 Cointegration test and VEC model

In this paper, we choose Johansen co-integration test for Cointegration test of multiple variables ^[11]. The results show that there are two co-integration:

$$\begin{aligned} \text{LnT} &= -0.3827\text{LnE} \\ \text{LnT} &= 0.3860\text{L} \end{aligned} \quad (2)$$

VEC model is based on the VAR model with a cointegration relationship constraint model. It focuses on the long-term balanced relationship between variables and reflect the dynamic influence between variables.

$$\Delta Y_{t-1} = \begin{pmatrix} -0.52 \\ 0.71 \\ -0.53 \end{pmatrix} \text{Co int EQ}_{t-1} + \begin{pmatrix} 0.39 & 0.13 & 0.16 \\ 0.04 & -0.01 & -0.28 \\ 0.54 & 0.73 & 0.90 \end{pmatrix} \Delta Y_{t-1} + \dots \Delta Y_{t-4} + u_t; \quad (3)$$

$$\Delta Y_{t-1} = \begin{pmatrix} 0.05 \\ -0.12 \\ 0.11 \end{pmatrix} \text{Co int EQ}_{t-1} + \begin{pmatrix} 0.42 & 0.17 & 0.17 \\ 0.01 & 0.31 & -0.08 \\ 0.52 & -0.39 & 0.05 \end{pmatrix} \Delta Y_{t-1} + \dots \Delta Y_{t-4} + u_t; \quad (4)$$

The judgment coefficient of the equation R2 = 0.8956, the revised judgment coefficient R2 = 0.8260; AIC = -16.01388, SC = -14.37248, they are all negative number, indicating that the correction model is good. The AR root test was used to test the stability of the model. The test results showed that the reciprocal of all the characteristic roots 's absolute value is less than 1 and all the distributed points were located in the unit round, the model was stable.

3.5 Impulse Response Function IRF

Using the impulse response function and variance decomposition, we analyze the dynamic behavior of the VEC model. Impulse response function (IRF) is used to measure the impact of a standard deviation impact on the current and future values of all endogenous variables in the VEC model. Fig. 1 depicts the IRFs of $\ln T$, $\ln E$, and $\ln N$.

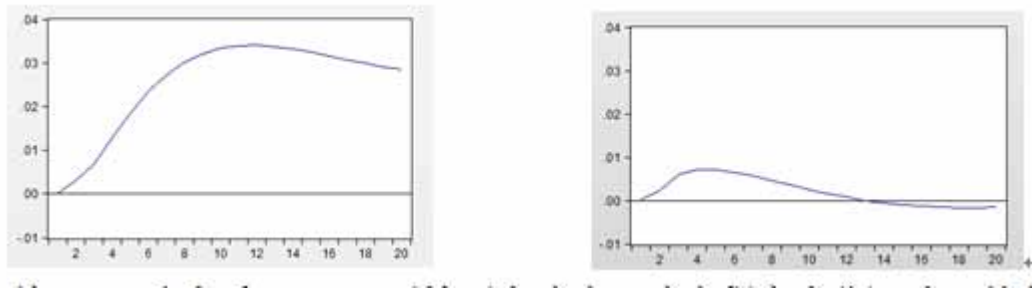


Figure 1. Impulse response of $\ln T$ to $\ln E$, and impulse response of $\ln T$ to $\ln N$, respectively.

When $\ln E$ lashed to $\ln T$, $\ln T$ began to respond positively, the response was very rapid, and showed the largest response at the twelfth period, the later stage slowly slow down into the stable period. This finding suggests that the impact of the Cross-border e-commerce transactions in Fujian Province can induce a rapid increase on the Fujian Province's total import and export trade in a short period of time. The gradual slowdown in the impact of cross-border e-commerce in Fujian Province may due to the fact that the cross-border e-commerce still has some problems, lack of stamina. When $\ln T$ lashed to $\ln E$, there will be a up and down effect to $\ln T$, and in the 13th, it's around at 0. eventually, the negative response began to converge and into the stable period. The growth rate is small, indicating that the number of netizens in Fujian Province has a cyclical fluctuations to the foreign trade in Fujian Province in the short term, but in the long run, the wave gradually disappear, which indicate that the impact is not deep.

Fig. 2 depicts the IRFs of $\ln E$, $\ln T$, and $\ln N$.

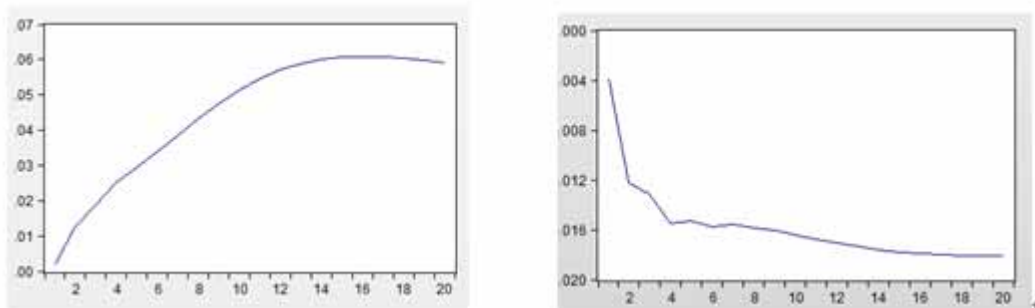


Figure 2. Impulse response of $\ln E$ to $\ln T$, and impulse response of $\ln E$ to $\ln N$, respectively.

When $\ln T$ make an impact to $\ln E$, $\ln E$ affected significantly and began to rise, reaching the maximum around the 15th, began to decline gradually. This shows that the impact on the volume of foreign trade transactions in Fujian Province can lead to the rapid growth of cross-border e-commerce transactions in Fujian Province over a certain period of time. As seen from Figure 3-5, $\ln N$ makes impact on $\ln E$, $\ln E$ showed a downward trend overall, and began into the eriod of stability only after the 12th; This shows that the relationship between the number of Internet users in Fujian province and Fujian's foreign trade is not obvious, which is consistent with the actual situation.

3.7 Variance decomposition

By Variance decomposition, we can analyze the contribution rate of each variable's lag period to the volatility

of the target variable, reflecting the relative importance of the each endogenous variable's random impact in the VAR system.

3.7.1 Variance contribution of variables to the foreign trade

Variance decomposition results in Fig 3 support and complement the empirical findings from the impulse response functions. According to the results of variance contribution, we can see that the foreign trade is affected by its own fluctuations only in the first period, and the second period begins to show the variance contribution of the change rate of cross-border e-commerce transactions. It rose to 87% at the 13th period and began to stabilize. The netizen change rate reached 22% in the third phase, with a clear short-term promotion effect and a long-term stability of about 1.6%.

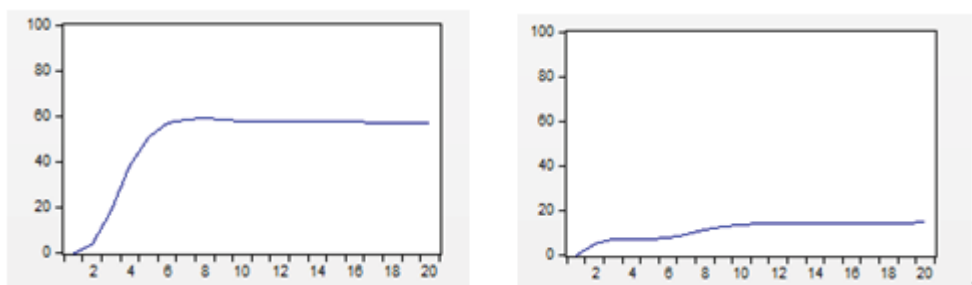


Figure 3. The variance decomposition of LnT

3.7.2 Variance contribution of variables to the cross-border e-commerce transaction volume

The variance contribution to the cross-border e-commerce transaction volume is shown in Fig 4. In the first phase, the cross-border e-commerce was affected by its own impact about 99%, and from the second phase, the foreign trade and the number of netizens's effect began apparent. the impact of foreign trade was basically stable and finally maintained at 11% after the fourth phase reached 13.8%, While the number of Internet users reached the maximum in the eighth period, then slowly declined.

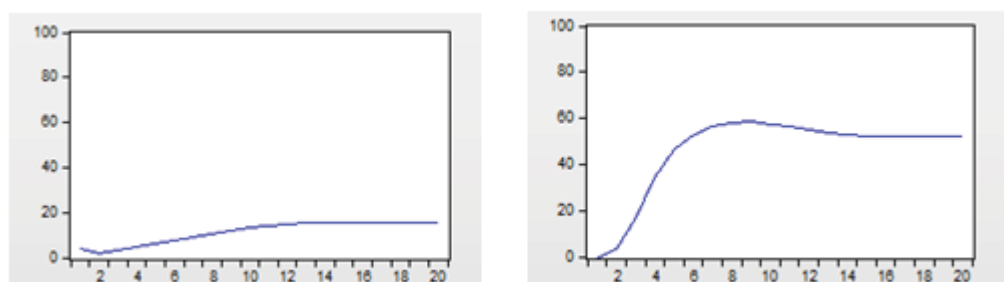


Figure 4. The variance decomposition of LnE

The results of the variance show that there is an interaction between the cross-border e-commerce and foreign trade, the economic development affect of cross-border e-commerce to foreign trade fully demonstrated in one year and the effect is long-lasting. However, the development impact of foreign trade to the cross-border e-commerce is relatively small, less than 15%. After all, we are still a trade surplus nation, the number of Internet users has little effect on foreign trade economy. It has a limited impact. However, Internet users still have an impact on the development of cross-border electricity suppliers in the long run. So talent base is an important driving force for industrial development.

4. CONCLUSIONS

This essay discusses the dynamic relationship between foreign trade and cross-border electronic commerce in Fujian province. The utilized foreign trade data are from Fujian province having the advantages of foreign trade and cross-border electronic commerce, by VEC model, we can find there is the interactive relationship between the total amount of foreign trade and cross-border electronic commerce transactions, the number of Internet users, moreover, we also can find the influence of cross-border electronic commerce on foreign trade is stronger, foreign trade on cross-border electronic commerce relatively weaker.

The results of empirical research showed the relationship between foreign trade and cross-border electronic commerce has the following characteristics: first, the government's great attention, more dividend policies making, especially the establishment of the free-trade area provide conditions for the development of cross-border electronic commerce, and promote its development greatly, and in the short term (1-3 years) can achieve certain profit, which will become an important driving force to foreign trade, if the policy is persistent, this force will last long time; however the advantages of traditional foreign trade does not mean it has the priority in new emerging industries, foreign trade transformation is a long-term job, this need to play a more active role of the government, adjusting industrial policy and industrial structure to promote foreign trade transformation and upgrading based on the various time and conditions. We should take market demand as the guidance, constantly open new fields, improve the proportion and requirements of business and information service industry, draw up relevant policies, set up unique core foreign trade industries and enhance the influence in international trade. Furthermore, with the national policy support, Fujian province establishes experimental free-trade area, the government should strongly recognize the great advantage of cross-border electronic commerce, using the advantage of free-trade area, to accelerate the exchange of information between enterprises, to guide transformation and upgrading of cross-border electronic commerce and foreign trade industry.

The influence of net users are not expected to be obvious, showing that the basic facilities, our Internet penetration rate has reached a certain degree, the online transaction has become a common way of trading, the policy for future guidance should not be using any longer, but should turn to the in-depth application and understand the enterprise need to enhance the international competitiveness.

The rapid development of cross-border electronic commerce in Fujian Province will not only introduce the new business model into the traditional foreign trade industry, but apply the advantages of low cost and cross international exchange to foreign trade, it has become an important driving force of foreign trade of Fujian province; yet the effect of number of net user on foreign trade growth is limited, this lower degree told us there still exist some problems in the actual cross-border electronic commerce in Fujian province.

As everyone knows, cross-border electronic commerce is based on electronic commerce transaction mode^[12], its development can never break away the coordination of the customs, logistics, after sale service and tax etc., the supply chain is not enough, only the government solve the problems of import and export and logistics, can that further motivate the enterprises' initiative, accelerate the upgrading and promote the economic development.

REFERENCES

- [1] B. Lin, C. Liu(2016). Why is electricity consumption inconsistent with economic growth in China? *Energy Policy* 88 ,310–316.
- [2] E Libin, & Huang Yongsheng. (2014). A new way of international trade: the latest research on cross-border e-commerce. *Journal of Dongbei University of Finance and Economics* (2), 22-31
- [3] Chang Cheng. (2015). Research on the interaction between cross-border e-commerce and trade growth. Ms D Thesis. Shenyang University of Technology.

- [4] Fengying, Yu (2015). Strategy analysis of developing cross border electronic commerce in traditional foreign trade. Journal of Changsha Aeronautical Vocational & Technical College.
- [5] Ji Xuejun. (2016). The impact of B2C cross-border e-commerce platform on China's foreign trade -- take Alibaba global express as an example. Ms D Thesis. Capital University of Economics and Business.
- [6] Yu Wenjing. (2016). The study of the impact of cross-border e-commerce on China's international trade and its countermeasures. Ms D Thesis. Shandong Normal University.
- [7] Li Fengyi. (2016). Analysis of the influence of foreign direct investment on foreign trade in China's cross-border e-commerce. Ms D Thesis. Capital University of Economics and Business.
- [8] Feng Zuo(2016). Evaluation of the new normal credit and the development of cross-border electricity supplier. Journal of Wuhan Commercial College, 30 (2), 64-66.
- [9] S. Johansen(1988), Statistical analysis of cointegration vectors, J. Econom. Dynam. Control 12,231–254.
- [10] Bureau of statistics of Fujian (2017). Statistics of Fujian. <http://www.stats-fj.gov.cn/xxgk/nds/>
- [11] S. Johansen(1991), Estimation and hypothesis testing of cointegration, Econometrica. 1551–1580.
- [12] M. Sun, P.-P. Zhang, T.-H. Shan, C.-C. Fang, X.-F. Wang, L.-X. Tian(2012), Research on the evolution model of an energy supply–demand network, Physica A 391,4506–4516.

Research on Influence Factors of Information Technology Enterprise's Operating Performance

Jiangping Wan^{1}, Jiawen Huang², Zhong Wang³*

^{1,2,3}School of Business Administration, South China University of Technology, Guangzhou, China

Abstract: This paper uses grounded theory to interview a listed information technology company in Guangzhou, and establishes the influence factors model of information technology enterprise's operating performance. The model concludes 17 key factors, including national policy, regional environment, market situation, strategy, execution, corporate values and leadership. The purpose of research is to help listed information technology enterprises improve their operating performance.

Keywords: information technology enterprise, operating performance, grounded theory, influence factors model

1. INTRODUCTION

China's ministry of industry and information technology (MIIT) released "The Development Plan of Software and Information Technology Services (2016-2020)" on December 18th 2016. MIIT proposed that the revenue of the industry will exceed 8 trillion yuan with an average annual growth rate of over 13% by 2020, and IT service revenue will account for 55% of business revenue. The revenue from information security products will reach 200 billion yuan with an average annual increase of over 20%. Moreover, the software exports will be more than \$68 billion and the software practitioners will reach 9 million. The development of information technology enterprise grows stronger and becomes more important to the social economy and people's daily life. Therefore, it is worth studying the influence factors of information technology enterprises' operating performance and the sustainable development. Just as Berliner, C and Brimson, J said, performance evaluation is a key factor in ensuring successful implementation of enterprise strategy ^[1].

This paper is organized as follows: section 2 is literature review, section 3 is research design, section 4 is open coding, section 5 is axial coding, section 6 is selective coding, and section 7 is conclusions.

2. LITERATURE REVIEW

According to the data released by MIIT, there were 40,900 nationwide software and information technology service enterprises with a total revenue of 4.32 trillion yuan in software business and a year-on-year growth of 16.6% in 2015. Software business revenue accounted for 28% of the electronic information industry with an increase of 1.4% over the last year. The income from information technology services was 2.21 trillion yuan with a year-on-year growth of 18.4% and an increase of 1.7% over the last year in the software and information technology service industries. The revenue from operation services, including online software operation services, platform operation services and infrastructure operation services, increased by 18.3% over the same period of the previous year. The revenue from e-commerce platform technical services, including online trading platform services, online transaction support services, increased by 25.1% year on year. The revenue from integrated circuit design increased by 13.3% year on year. The revenue from other information technology services, including information technology consulting and design services, system integration, operation and maintenance services and data services, increased by 17.8% year on year ^[2].

Anandhi S. Bharadwaj developed the concept of IT as an organizational capability and empirically examines the association between IT capability and firm performance ^[3]. Wang et al. chose total assets,

* Corresponding author. Email: csjpw@scut.edu.cn (Jiangping Wan), jiawen.huang@hotmail.com (Jiawen Huang)

operating costs, human resources cost and R&D investment as input indicators, net profit after tax, turnover ratio, the proportion of core business income and total assets growth rate as output indicators when he studied on the evaluation of the operating efficiency of the software and Internet industry GEM listed companies ^[4]. Xiong et al. took R&D expenses, R&D personnel, advertising investment, human resource costs and management costs as input indicators and sales revenue and brand equity as output indicators when she researched on the operating efficiency of Chinese high-tech start-ups ^[5]. Wan et al. proposed the business ecosystem of software industry, and constructed the evaluation model of the health of the commercial ecosystem and applied it to the health analysis of the commercial ecosystem in Chinese software industry ^[6]. Wan et al. also analyzed three ICT enterprises in Guangzhou, 21 risk factors of ICT commercialization were identified with grounded theory and the three-level ICT commercialization risk factor model was established in the views of technical management, project management and dynamic capability ^[7]. Wan et al. selected two investors and three entrepreneurs to research on the risk factors of entrepreneurship in Internet industry with the grounded theory and established the model of risk factors that affected entrepreneurship in Internet industry with three dimensions ^[8].

3. RESEARCH DESIGN

Grounded theory is a research method developed by Anselm Strauss and Barney Glaser of Columbia University ^[9]. Grounded theory requires researchers to start from the original data, abandon all the theoretical hypotheses, and based on the collected data, find out the concept of reacting social phenomena through induction, and then establish the relevant theories by developing the connection between these concepts.

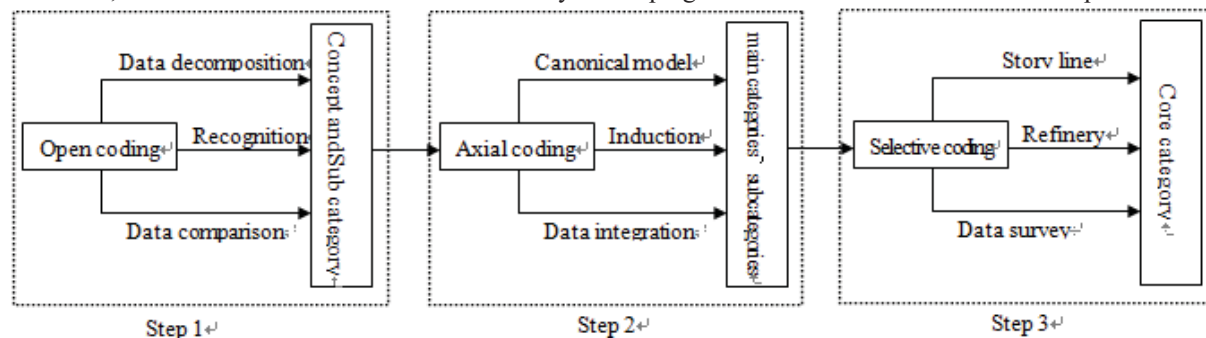


Figure 1. Coding analysis steps of grounded theory

Considering the diversity and uncertainty of the sources of information technology business performance impact factors, this paper adopts a more operational procedural grounded theory. Procedural grounded theory is mainly divided into four steps as follows: (1) A research problem arises. The problem of grounded theory research is usually the action and process about the phenomenon. After the definition of the research problem, we chose the research object. (2) Collect data. We collected data by means of observing, in-depth interviews and document analysis. (3) Open coding, axial coding and selective coding the documents and analyze the codes (Figure 1). (4) Build theories. We compared the concepts, categories, subcategories and main categories formed in the coding process, established their mutual relations, constructed and judged the theory.

We chose one of the 69 information technology listed companies as our interviewee, hereinafter referred to as J Company. J Company, located in Guangzhou, focuses on the development and application of smart technologies and products and is the leading provider of smart city products and solutions in China. The main businesses of the company include smart security, intelligent rail transit, value-added communications, integrated services (including network and cloud computing products and services, and IT integrated services), of which the smart security and intelligent rail transit business are the key businesses.

Participants in the interview included the vice president and CTO, and deputy general manager of company J, who were representative, scientific and professional. We designed a semi-structured interview questionnaire

(Appendix) for interviewees. The questionnaire focused on the key factors that led to the success or failure of the IT business. And the listed questions were open, which gave full play to the enthusiasm of the interviewees and facilitated further understanding of deeper information. We recorded the key information from the interview recording and converted the recording into words on the same day to ensure that the interview information was true and accurate.

4. OPEN CODING

Open coding refers to the decomposition of the original data, continuous comparison between events and between events and concepts through word-by-word analysis, and the formation of concepts and categories [9]-[10]. The procedures of open coding include: (1) Follow the rules of native coding, extract the subject words from the original sentences of interview records, and conceptualize them. (2) Mine categories and name them from multiple sources, such as literature, interview records or the results of discussions with experts. (3) Find the nature and dimensions of categories. In the open coding process, it is necessary to maintain an open mind, not to build any preconceived and predetermined theoretical framework.

After the interview, we carried out structured disorganization of the interview records, discussed with experts in relevant fields with the literature and interview records. Next, we repeatedly compared the records, concepts and categories, deleted low occurrence initial concepts and merged some initial concepts in order to refine initial concepts. Then, we got 32 initial concepts and finally obtained 17 categories after refining. Table 1 lists the open codes of interview records. 2-3 corresponding sentences were selected for each category.

Table 1. Open coding analysis

The records of company J	Initial concept	Category
China vigorously develops smart city and strongly supports smart security, intelligent transportation and other fields.	Industrial policy	National policy support
The state encourages the promotion of the public infrastructure PPP model and gradually introduces social capital to meet the funding needs for construction and financing of rail transit construction. It is foreseeable that PPP mode will be a strategic opportunity in the next three years.	Access to social capital	
Our company is a key software enterprise within the national planning layout and its income tax is at a preferential tax rate of 10%.	Tax incentive	
The general situation of intellectual property protection is better than before, and there has been a great improvement in national legal protection and market respect of intellectual property.	Intellectual property protection	Intellectual property protection
Technology development and industry development become more closely. We provide a series of solutions, including demand research, design and development, operation and maintenance, etc., which is not one or two copied pirated software.	Technology barrier	
The economic situation in the past two years is not very good. Some enterprises are struggling with their development.	Economic situation	The level of economic development
Our development of intelligent security and rail transit is a strategic industry vigorously developed by the country in recent years. With the progress of urbanization, the development of the industry is promising and the development situation is relatively good.	Prospect	Industry development cycle
Guangzhou is the center of traditional trade, and the attention and service to science and technology enterprises are still in the progress.	Government service	Local government service
The Guangzhou Municipal Science and Technology Bureau had a meeting a few days ago to find out what difficulties the enterprise encountered in the current development, and the government tried to help solve those.	Problem solution	
Guangzhou has many college resources, we have a lot of research collaboration with Sun Yat their-sen University.	Research collaboration	The level of science and technology
Guangzhou has well-developed trade, but it needs to learn from other cities such as Shenzhen so as to enhance the overall scientific and technological level.	Technological prowess	
The BT project and the PPP project of our company need huge capital. Some developed finance cities such as Guangzhou can help us raise more funds. By now, the company has set up an industrial fund.	Developed financial industry	Financing environment

The records of company J	Initial concept	Category
We are a listed company that can raise funds by issuing stocks, bonds, etc. There are many financing channels.	Financing channels	
With the development of smart city, the customer demand are increasingly differentiated, refined and integrated, which requires the enterprise to have both depth and height. It is necessary to have a deep insight into the personalized intelligent service needs of different industries such as security, transportation, finance, justice and education in smart cities. On the other hand, it can provide solutions for high-quality housing construction and help customers to solve problems and innovation management and service with various intelligent technologies.	Customer demand	Market insight
Based on our own technological advantages and experience, we focus our development on smart security and rail transit, and we will further promote the development of these two businesses in the future.	Core business	Market positioning
According to the company's strategic adjustment, we stopped IT products supply chain management services business.	Strategic adjustment	
Our company has invested a lot of resources to carry out research on advanced technology related to big data, cloud computing and other industries in order to promote the industrialization of frontier technology, In 2015, the company added 3 new patents, 28 software copyrights, 3 software product registration certificates and 41 product inspection reports. The Video Intelligence Analysis System won the first prize of Guangzhou Science and Technology Progress Award. And the Comprehensive Monitoring of Large-scale Rail Transit System Software Platform passed the identification, and the experts concluded that the results had reached the international advanced level.	Technology research and development	Technical innovation
Our company developed the automatic fare collection system and completed the research of face recognition gates, cloud gates, two-dimensional code payment and a new generation of PSD control system, and applied them to the newly opened Guangzhou Metro Line 6 and Line 7, Guangfo Line.	Technical transformation	
Smart city, smart security and intelligent rail transit require high level of technology and product quality, which can bring high profits.	Target customer	Value proposition
The automatic fare collection system developed by the company supports UnionPay flash and cloud flash payment, making passengers travel more convenient, and the operation of metro companies is more efficient. In the future, mobile payment and face payment will also be vigorously promoted.	Convenience and efficiency	
Our company is mainly responsible for research and development, Our suppliers, including H3C and Hikvision, are responsible for production, operation and maintenance.	The division of value chain	Value network
We signed a cooperative agreement with the Information School of Sun Yat their-sen University to promote the research of vehicle image intelligence analysis technology and products. In April 2015, the company cooperated with Chongqing Research Institute of Chinese Academy of Sciences to invest in Cloudwalk company, and jointly promoted the industrialization of face recognition and machine vision technologies in security, finance and other fields.	R&D collaboration	
The company's customers and suppliers are all over the China. We have been working with our partners for a long time, mutual trust, smooth communication and stable suppliers.	Good relationship	
While maintaining close cooperation with upstream manufacturers, our company strengthens the construction of service network for end users and downstream application integrators, continuously strengthens the optimization of upstream and downstream operation systems and business processes, and improves operational efficiency and project delivery capabilities.	Industry chain collaboration	
Our company's organizational structure and personnel are constantly changing, and constantly improve the business structure in order to meet the needs of business development.	Organizational structure	Organizational change
In 2014, the company began to carry out the equity incentive plan and the employee stock ownership plan.	Equity incentive	Talent policy
Our company is flexible in performance assessment, we set salary by job, skill and performance.	Salary policy	
The company's organizational climate is harmonious, and we respect highly qualified, sincere, open and dedicated high-quality talents. We encourage everyone to trust and understand each other, keep on learning and keen innovation..	Organizational climate	Corporate culture
Our company pays attention to knowledge and often carries out employee training, new staff training, professional ability training and management personnel reserve training, so that employees continue to grow.	Employee training	

The records of company J	Initial concept	Category
The concept of company value is "The City Will Become Better", which means that technology and intelligence makes the city better. And this value concept is also our vision to actively participate in the construction of smart city.	Value concept	Corporate values
The company set up a 3×3 development strategy, which means 2013-2015 years is integration and development period, 2016-2018 is leaping development period, and 2019-2021 is company's innovation and development period. The three-year specific goals are further detailed and detailed. We all have the same goal and make concerted efforts for development.	Development goal	
Our boss, a courageous and far-sighted person, plans and promotes the company's decision-making on major issues, making the company develop better and better.	Leadership	Leadership

5. AXIAL CODING

Axial coding is to classify categories in open coding by cluster analysis, establish the association between correlation and logical order among different categories, and construct a relation network around "main axis" of category to form main categories and subcategories. We can use the model of "antecedent conditions - theoretical phenomena - context - mediated condition - action / interaction strategy - action consequence" to connect categories to the main categories ^[6].

We repeatedly compared the concepts and categories in Table 1, considered the relationships among the categories, and ultimately obtained 2 main categories and 7 subcategories (Table 2).

Although the axial codes in Table 2 cannot present a complete influence factors theoretical model of information technology enterprise, we draw the following conclusions:

(1) The influence factors of information technology enterprise' operating performance are divided into external and internal factors.

(2) The external factors are divided into the national policy, market situation and regional environment, which are the factors that are difficult for information technology enterprises to control, but those will indirectly have a significant impact on the operation and development of information technology enterprise. The national policy includes industrial policy, access to social capital and tax incentive. The market situation includes the level of economic development and industry development cycle. Regional environment includes local government service, financing environment and the level of science and technology.

(3) The internal factors are divided into strategy, execution, corporate values and leadership. The four interrelated factors form the internal operation system of information technology enterprise, those have a direct impact on operating performance. The strategy includes market insight, market positioning, technical innovation and value proposition. The execution includes value network, organizational change, personnel policy and corporate culture.

6. SELECTIVE CODING

Selective coding is to aggregate all categories from the open coding and axial coding to the core category, verify their relationship, and complete the category with incomplete conceptualization ^[6]. The core category is a dominant concept that summarizes all categories and occupies a central place related with all categories.

The steps of selective coding include: (1) Define the story line. (2) Describe the main categories, subcategories and concepts of the codes. (3) Improve the incomplete concepts and categories. (4) Select the core categories. (5) Establish a theoretical relationship between the core categories and other categories ^[7].

The theoretical model of the influence factors of information technology enterprise' operating performance was illustrated in Figure 2 through the analysis of 17 categories, 7 subcategories, 2 main categories and the original records.

Table 2. Axial coding analysis

The intension of category	Category	Subcategory	Main category	
National policy orientation, encouragement of social capital and tax incentive	Industrial policy	National policy	External factor	
The intellectual property protection of the country and enterprise	Intellectual property protection			
The level of national economic development, the cycle of economic development	The level of economic development	Market situation		
Industry development prospects and maturity	Industry development cycle			
Local government's attention and service awareness	Local government services	Regional environment		
Local science and technology and research ability	The level of science and technology			
The development of local financial industry and financing channels	Financing environment			
Understand customer demand and solve customer's pain points	Market insight	Strategy	Internal factor	
Business market positioning, business focus and development strategy	Market positioning			
Scientific and technological innovation, technology research and development, technical transformation	Technical innovation			
Target customer orientation and customer value provision	Value proposition			
Upstream and downstream industry chain collaboration, research collaboration and value network collaboration	Value network	Execution		
Organizational structure and business architecture	Organizational change			
Talent recruitment, talent incentive and salary policy	Talent policy			
Organizational climate and employee training	Corporate culture			
Value concept and development goal	Corporate values	Corporate values		
Leadership vision planning and leader charm	Leadership	Leadership		

As illustrated in Figure 2, we can find that in the following: (1) The operating performance of information technology enterprise is the core category. Around this core category, the main categories, subcategories and concepts are organically integrated. And the operating performance of information technology enterprise is affected both by external factors and internal factors. (2) The story line around this core category is that the external factors of national policy, market situation and regional environment have an external impact on information technology enterprise and the internal factors of strategy, execution, corporate values and leadership have a direct impact on information technology enterprise. (3) The strategy, execution, corporate values and leadership interact with each other and affect the operating performance of information technology enterprises together with the external environment. (4) If information technology enterprises want to improve their performance and maintain sustained and healthy development, they should analyze their own current situation, find their advantages and disadvantages and make continuous improvements from the above aspects. In our understanding, we also should consider e-service and new economic situation in following: Chuang et al. developed an e-service capability, and an innovation strategy that emphasizes service innovation orientation to

examine information-value offering^[11], and Yang et al. proposed the performance evaluation system design of enterprise environment under the new economic situation^[12].

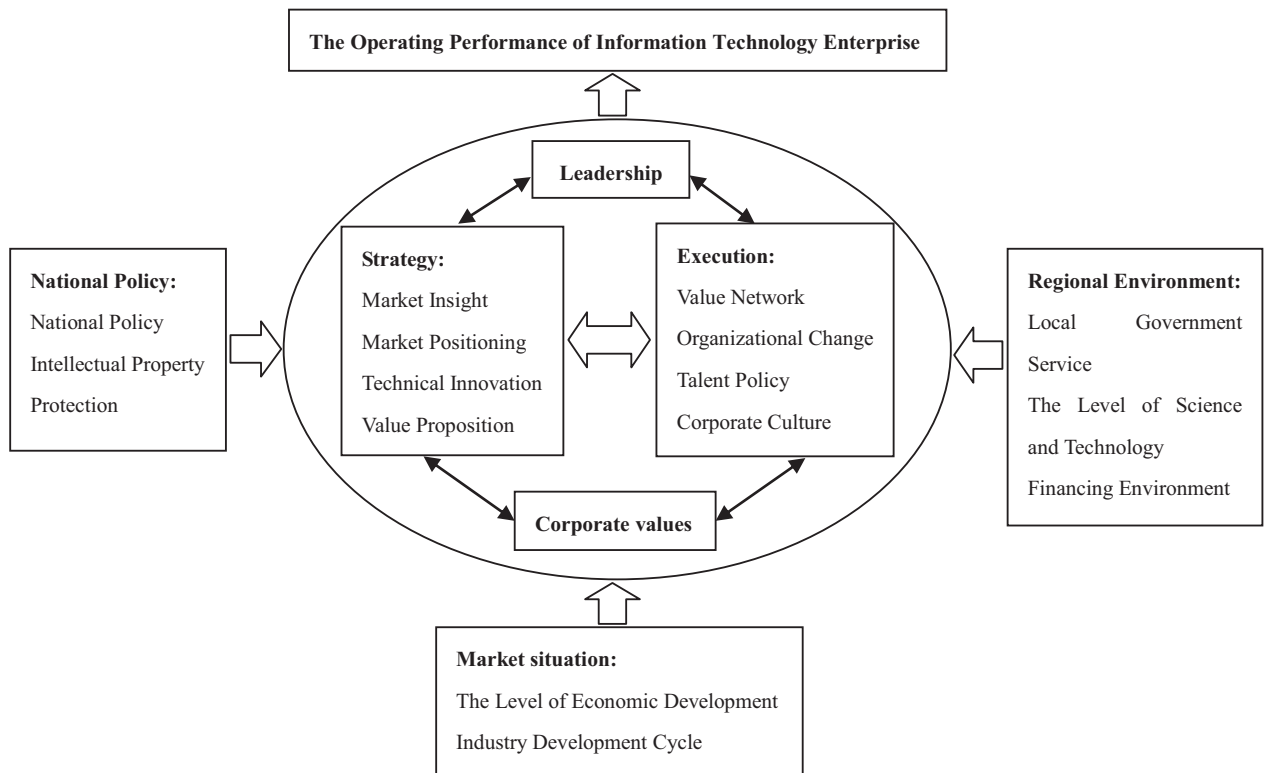


Figure 2. The influence factors model of information technology enterprise's operating performance

7. CONCLUSIONS

This paper is a partial results of the research on the operating performance and influencing factors of the Chinese listed information technology enterprises. Our research first applied DEA method to statically evaluate the operating performance of 69 Chinese listed information technology enterprises in 2015. The results illustrated that the overall operating performance of information technology enterprises was ordinary. Thus, it was necessary to improve their management, the efficiency of resource utilization and scale efficiency. Then, we applied the Malmquist index method to dynamically evaluate the performance changes of these enterprises in five years from 2011 to 2015. The results illustrated that the change of pure technical efficiency and scale efficiency played an important role for business development in 2011-2014 with a gradually reducing contribution. The technological change had a gradually reduced hindrance to business development. However, in 2015, the change of pure technical efficiency and scale efficiency had a hindrance to the business development. The technological change played a role in promoting the business development. Due to length limitation, the above research will be published in another paper. This paper uses grounded theory to interview a listed information technology company in Guangzhou, and establishes the influence factors model of information technology enterprise's operating performance, which concludes 17 key factors, including national policy, regional environment, market situation, strategy, execution, corporate values and leadership. The purpose of research is to help listed information technology enterprises improve operating performance in China. It is notability that May McCreddie et al. illustrated the potential for a new form of grounded theory methodology (GTM) drawing on discursive approaches (DGTM)^[13], in our understanding. DGTM will improve the interactive between researchers and practitioners to get deep insights.

REFERENCES

- [1] Berliner, C, Brimson, J. (1988). Cost Management for Today's Advanced Manufacturing: The CAM-I Conceptual Design. Boston: Harvard Business School Press, 70:13-23.
- [2] Chinese industrial information. (2016). A survey of the development of China's information technology service market in 2016. <http://www.chyxx.com/industry/201607/430325.html>,
- [3] Anandhi S. Bharadwaj. (2000). A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation, *MIS Quarterly*, 24(1):169-196
- [4] Wang Chongyuan, Zhang Chuifan. (2016). Software and Internet industry GEM growth enterprise operation efficiency evaluation-based on DEA method. *Industrial Economy Review*, 3(6):732-737 (in Chinese).
- [5] Xiong Chan, Mai Yiyuan, He Xiaobin, Xiao Renqiao. (2014). A study on operational efficiency of hi-tech startups in China based on DEA methods. *Journal of Management Science*, 28(2):26-37.
- [6] Wan Jiangping, Zhu Shiqing, Hua Lishui. (2009). Some Thoughts on Business Ecosystem in China Software Industry. *Science and Technology Management Research*, 29(4):130-134 (in Chinese).
- [7] Wan Jiangping, Liang Lianyu, Wan Dan. (2013). Research on risk factors of ICT commercialization with grounded theory. *Journal of Computers (Finland)*, 8(9): 2356-2365
- [8] Wan Jiangping, Jiang Qiaowen, Xie Leqi. (2017). Research on Risk Factors of Entrepreneurship in Internet Industry with the Grounded Theory. *WHICEB2017 Wuhan, China*, 99-107
- [9] Corbin, J.M, Strauss, A. (1990). Grounded Theory Research: Procedures, canons, and Evaluative Criteria. *Qualitative Sociology*, 13(1): 3-21.
- [10] Chen Xiangming. (1999). Grounded theory of ideas and methods. *Educational Research and Experiment*, 17(4): 58-63 (in Chinese).
- [11] Shu-Hui Chuang, Hong-Nan Lin. (2017). Performance implications of information-value offering in e-service systems: Examining the resource-based perspective and innovation strategy. *Journal of Strategic Information Systems*, 26(1): 22–38
- [12] Guojun Yang, Xiaohu Zhou. (2017). Performance Evaluation System Design of Enterprise Environment Based on New Situation Theory and Big Data Technology, *Revista de la Facultad de Ingeniería U.C.V.*, 32(5): 641-649.
- [13] May McCreaddie, Sheila Payne. (2010). Evolving Grounded Theory Methodology Towards a discursive approach, *International Journal of Nursing Studies*, 47(6): 781-793.

APPENDIX: INTERVIEW QUESTIONNAIRE

1. Please give a brief introduction to the basic situation and main business of the company.
2. Please introduce the main reasons for the company's previous transformations?
3. What do you think about the development status of China's information technology business?
4. What are the external factors that affect the operation and development of Chinese information technology enterprises?
5. What are the internal factors that affect the operation and development of Chinese information technology enterprises?
7. Which of the above key factors in the company there are advantages? What are their disadvantages?
8. Please give a brief introduction to the future direction of the company and your opinion on the future development direction of Chinese information technology enterprises.

An Analysis with Evolutionary Game of the Resource Sharing in Supply Chain Under Cloud Platform

Ling-wu Zeng¹, Yi Hu¹, Xing-jian Zhou^{1,2}*

¹School of Management, Wuhan Textile University, China

² Research center of enterprise decision support, Research base for Humanities and social sciences in Hubei Province, China

Abstract: Based on the sharing mode of supply chain resources in the environment of cloud service, this research constructed the evolutionary game model of supply chain resource-sharing to reveal the behaviors between two types of enterprise, the equilibrium in model and local stability are analyzed under the state of uniform mixed and non-uniform mixed populations. By using the method of system dynamics, the evolutionary game model is built, and a contrastive analysis of evolutionary results affected by diverse parametric variations is performed. The results of the research shows that the evolutionary trends of the game are significantly influenced by the initial sharing proportion in enterprise group, the cost and benefit of upgrading equipment, and the risk of technological loss. To facilitate the information interaction and resource sharing between enterprises, continuous improvement needed to be done in line with the above aspects.

Keywords: cloud platform, sharing of supply chain resource, evolutionary game, system dynamics model

1. INTRODUCTION

Nowadays, the development of information technology has transformed the competition between enterprises into the competition of global supply chain^[1], and the overall competitiveness of the supply chain can be achieved through the coordinated development between the members. One of the conditions for synergetic development lies in the circulation and sharing of resources. At the same time, with the advancement of information technologies such as Internet of Things and cloud computing, the cloud service platform makes the seamless connection of information and resources among supply chain enterprises accessible^[2]. All of this can build a virtual enterprise alliance, and jointly improve product quality and enhance the overall competitiveness of the supply chain. However, the information among members of the supply chain is not open and opaque, besides, the conditions that companies capitalize on supply chain member's core competencies and protect their own knowledge and technology are widespread, which leads to the low willingness of sharing the supply chain resources, so it is hard to form an effective community of supply chains. Therefore, it is of theoretical and practical significance to study how to maximize the benefits of supply chain members under the premise of ensuring the overall interests of supply chain cooperative organizations.

2. LITURATURE REVIEW

Up to now, the research on supply chain resource sharing mainly includes the following two aspects. First, the impact of resource sharing on the performance of supply chain operations, the representative paper are, Frank Chen^[3](2000) explored the degree of variation of the variance $\text{var}(q^k)$ of orders placed in the k th stage of the supply chain relative to the variance of market demand $\text{var}(D)$ under the pattern of demand determination and uncertainty, indicating that resource sharing can significantly increase the performance of supply chain operations. However, most of the literatures focuses on the distribution of benefits between resource sharing

* Corresponding author. Email: wuliwutu@163.com (Xing-jian Zhou)

parties, Fu Heng^[4](2018) argues that wholesale price of the manufacturer and retailer's pricing strategy are the main factors that determine the distribution of benefits between the supply chain, they constructed a single manufacturer-retailer supply chain system to discuss the each pricing strategy for manufacturer and retailer to find out the best distribution point of interest under the uncertainty market demand, and verify the existence of utility equilibrium in the optimal decision. Based on the retailer's risk appetite, Wang Cong and Yang Deli^[5](2017) studied the conditions that tradition retailer sharing demand information with manufacturers, and discussed the impact of information resource sharing on their business decisions and profits of both retailers and manufacturers. And numerical simulations showed that resource sharing can maximize the benefits of both parties only when the demand is uncertain. Lu Jizhou, Feng Gengzhong^[6](2017) Aiming at the problem of bullwhip effect in the supply chain inventory, they established a cooperative game model covering the individual utility function and the overall collaborative optimization of the supplier and the retailer, and transforms the equilibrium of the resource allocation in the game community into the optimal decision balance, to ensure that the sharing of resources can maximize the interests of both parties.

However, most of the existing literature focuses on the decision of resource sharing through the distribution of benefits. But there is a lack of discussion on the factors that limit the resource sharing between supply chains, and the influence of different situations on the final decision behavior is also lack of argument. In addition, the resource sharing among supply chain is a dynamic process, the dynamic changes of sharing factors will ultimately affects the choice of a balanced strategy for both supply chains. Therefore, this study adopts the evolutionary game method to analyze the evolution trend of the dynamic cooperation process of resource sharing among supply chain members in the cloud service environment, and measure the impact of various factors on the equilibrium strategy in different situations by using system dynamic simulation model.

3. THE FRAMEWORK OF EVOLUTIONARY GAME MODEL

3.1 Problem description

Enterprises can access various resources that are lacked in the enterprise through the resource sharing platform, but they needed to pay a certain access costs, and bear the risk of the leakage of core technologies, but they may also obtain excess returns. This research draw lessons from the game method that Qi Ershi used to analysis the resource sharing between manufacturers, we assumes that the main body of sharing supply chain resources are two types of "group" enterprises, one type of enterprises mainly share hard manufacturing resources(denoted as enterprises group A), and the other type is based on sharing soft service resources(recorded as enterprises group B). The strategies among both of them are $S\{\text{shared, not shared}\}$, and both sides of the game are "bounded rationally". The evolutionary game model is constructed by the strategies and benefits of both sides to study the evolution of the result of both parties under the condition of homogeneous and heterogeneous mixed groups.

3.2 Model assumption

(1) If both group A and group B does not share enterprise's resources, both parties produce and operate independently, the income earned by group A is P_a and that of group B is P_b .

(2) If group A share enterprise resources while group B do not, group A needed to invest a lot of fixed costs to purchase or upgrade existing material-sensing devices, denoted as C_a . At this point, although there is no platform for the cooperation information between supply chain enterprises, but the production and information level of group A has been improved, the proceeds will also change accordingly, here the degree of enterprise informatization is R_a , from the extra income brought by the improvement of information technology was

(3) If group B share enterprise resources while group A do not, group B are mainly rely on back-end productive service businesses in the supply chain and have a higher level of informationization, so the cost of investment in upgrading devices of group B is less than group A, marked as C_b . However, group B have to bear the risk of the leakage of core technologies. Therefore, the loss risk factor for group B is recorded as R_b , and the losses due to technological loss is recorded as L_b . The income of group A is still P_a .

(5) χ is the proportion of individuals who choose to share enterprise resources in group A. $1-\chi$ is the proportion of individuals who choose not in the group A; y is the proportion of individuals choosing to share the enterprise resources in group B, and $1-y$ is the proportion of individuals who choose not.

3.3.1 Homogeneous mixed evolution game model

Table 1. Evolutionary game payment matrix of supply chain enterprises.

$$F(y) = y(I_b^s - I_b) = y(1-y)(I_b^s - I_b^n) = y(1-y)[\chi(P_b^s - C_b - R_b^s L_b) + (1-\chi)(P_b - C_b - R_b L_b) - P_b] \quad (8)$$

Copying the dynamic steady state means that the proportion of game parties adopting both strategies remains constant. So when $F(x)=0$ and $F(y)=0$, we can get the five equilibrium points of the evolutionary game process, respectively are, O (0,0), A (0,1), B (1,0), C (1,1), D (χ_D , y_D), among them, $X_D = \frac{c_b + R_b L_b}{P^s_b - P_b + (R_b - R^s_b)L_b}$, $y_D = \frac{C_a - R_a I_a}{P^s_a - P_a + (R^s_a - R_a)I_a}$. And when $\chi=0$, $\chi=1$ or $y = \frac{C_a - R_a I_a}{P^s_a - P_a + (R^s_a - R_a)I_a}$, the probability of choosing a "share" strategy for group A is stable, while $y=0$, $y=1$ or $\chi = \frac{c_b + R_b L_b}{P^s_b - P_b + (R_b - R^s_b)L_b}$,

the probability of choosing a "share" strategy for group B is stable. Thus, the stability of evolutionary game equilibrium can be obtained by Friedman's Jacobian matrix stability decision method^[9]. Deriving $F(\chi)$ and $F(y)$ from partial derivatives of χ and y respectively, we get the Jacobian matrix of the system as follows:

$J =$

$$\begin{bmatrix} (1-2\chi)[y(P^s_a - C_a + R^s_a I_a) + (1-y)(P_a - C_a + R_a I_a) - P_a] & \chi(1-\chi)[(P^s_a - C_a + R^s_a I_a) - (P_a - C_a + R_a I_a)] \\ y(1-y)[(P^s_b - C_b - R^s_b L_b) - (P_b - C_b + R_b L_b)] & (1-2y)[\chi(P^s_b - C_b - R^s_b L_b) + (1-\chi)(P_b - C_b + R_b L_b) - P_b] \end{bmatrix}$$

consequently, we can get the stability analysis of five equilibrium points on the basis of the Jacobi ratio of the matrix stability analysis showed in Table 2. Because χ and y represent the probability of a strategy for both sides of the game to choose, so $0 \leq \chi \leq 1$, $0 \leq y \leq 1$, then the constraints conditions are available: $C_b + R_b L_b \geq 0$, $P^s_b - P_b + (R_b - R^s_b)L_b \geq 0$, $C_a - R_a I_a \geq 0$, $P^s_a - P_a + (R^s_a - R_a)I_a \geq 0$.

Table2. Balance point and local stability.

Balance point	DetJ	Symbol	TrJ	Symbol	Result
O(0,0)	$(C_a - R_a I_a)(C_b + R_b I_b)$	+	$-C_a + R_a I_a - C_b - R_b I_b$	-	ESS
A(0,1)	$-(P^s_a - P_a - C_a + R^s_a I_a)(P^s_b - P_b - C_b + R^s_b I_b)$	—	$P^s_a - P_a - C_a + R^s_a I_a$	+	Not stable
B(1,0)	$-(C_a - R_a I_a)(C_b + R_b I_b)$	—	$P^s_b - P_b - C_b + R^s_b I_b$	+	Not stable
C(1,1)	$(P^s_a - P_a - C_a + R^s_a I_a)(P^s_b - P_b - C_b + R^s_b I_b)$	+	$-P^s_b + P_b + C_b + R^s_b I_b$	—	ESS
$D(\frac{c_b + R_b L_b}{P^s_b - P_b + (R_b - R^s_b)L_b}, \frac{(C_a - R_a I_a)P^s_a - P_a - C_a + R^s_a I_a}{P^s_a - P_a + (R^s_a - R_a)I_a})$	$\frac{(C_b + R_b L_b)P^s_b - P_b - C_b - R^s_b I_b}{P^s_b - P_b + (R_b - R^s_b)L_b}$	+	0	/	Saddle point

Through the stability analysis we can see that there are 2 out of 5 equilibrium points are evolutionary stability strategy (ESS), respectively, they are O (0,0) point and the C (1,1) point which meet the conditions $\text{Det}J > 0$ and $\text{Tr}J < 0$. So the corresponding strategy is {share, share} and {not share, not share}. A (0,1) and B (1,0) are unbalanced point, D ($\frac{c_b + R_b L_b}{P^s_b - P_b + (R_b - R^s_b)L_b}, \frac{C_a - R_a I_a}{P^s_a - P_a + (R^s_a - R_a)I_a}$) are Saddle point. Analysis shows

that the dynamics game eventually converges to individuals in both types of group select share enterprise resources O (0,0) or choose not to share enterprise resources C (1,1) those two equilibrium points, and which direction does saddle point move is decided by the system initial state, the income that companies share enterprise resources, the investment costs access to cloud service platform, the improvement of information technology and technology losses and other factors that affect the common role. Two groups of enterprises can adjust and control their own implementation of enterprise resource sharing costs, information gains number, the risk of technological loss coefficient and other parameters to make the dynamic game process moves towards the ideal equilibrium point, promoting enterprise to share resources under cloud service environment.

3.3.2 Heterogeneous mixed evolution game model

In the actual situation, the number of a population is limited, and the rate of contact between individuals are different^[10]. So according to Taylor^[11](2006), QuanJi^[12](2013) research on non-uniform contact rate, our study record the contact rate of group A and B choose shared strategy between individuals as R_{11} , the contact rate

between individuals who choose shared strategy in group A and choose not share in group B is R_{12} , the contact rate between individuals who choose shared strategy in group B and who do not in group A recorded as R_{21} , the contact rate between individuals who choose not share in both group A and B recorded as R_{22} .

For the convenience of the model, set $\chi_a = P_a^s - C_a + R_a I_a$, $y_a = P_a - C_a + R_a I_a$, $W_a = Z_a = P_a$, $\chi_b = P_b^s - C_b - R_b^s L_b$, $y_b = Z_b = P_b$, $W_b = P_b - C_b - R_b L_b$.

In this way, the game revenue matrix for supply chain resource sharing of two groups of enterprises in the cloud service environment can be expressed as follows:

$$\begin{bmatrix} (\chi_a, \chi_b) & (y_a, y_b) \\ (W_a, W_b) & (Z_a, Z_b) \end{bmatrix}$$

From the perspective of group A, in the condition of non-uniform contact rate, the income of group A choose to share and not share the enterprise resources are:

$$I_a^s = \frac{x_a R_{11}x + y_a R_{12}(1-x)}{R_{11}x + R_{12}(1-x)}, I_a^n = \frac{W_a R_{21}x + Z_a R_{22}(1-x)}{R_{21}x + R_{22}(1-x)} \quad (10)$$

So according to the copying equation $F(x) = \chi(I_a^s - I_a^n) = \chi(1-\chi)(I_a^s - I_a^n)$, we can conclude that:

$$I_a^s - I_a^n = \frac{(\varphi + \eta - \delta)\chi^2 + (\delta - 2\eta)\chi + \eta}{[R_{11}\chi + R_{12}(1-\chi)][R_{21}\chi + R_{22}(1-\chi)]} \quad (11)$$

Among them,, $\delta = (x_a - Z_a)R_{11}R_{12} + R_{12}R_{21}(y_a - W_a)$, $\varphi = (x_a - W_a)R_{11}R_{21}$, $\eta = (y_a - Z_a)R_{12}R_{22}$. Set as $h(\chi) = (\varphi + \eta - \delta)\chi^2 + (\delta - 2\eta)\chi + \eta$, $T(\chi) = \chi(1-\chi)(I_a^s - I_a^n)$.

And then let $h(\chi) = 0$, when $\delta^2 \geq 4\varphi\eta$, we can conclude that:

$$X_1 = \frac{2\eta - \delta + \sqrt{\delta^2 - 4\varphi\eta}}{2(\varphi + \eta - \delta)}, \quad X_2 = \frac{2\eta - \delta - \sqrt{\delta^2 - 4\varphi\eta}}{2(\varphi + \eta - \delta)} \quad (12)$$

Divided into three cases, we discussed the evolutionary stability on the value of the proceeds of group A:

(1) When $(\chi_a - y_a)(W_a - Z_a)$ and $(\chi_a - W_a)(y_a - Z_a)$ are both positive number,

$$\frac{R_{11}R_{12}}{R_{21}R_{22}} \geq \left(\frac{\sqrt{(\chi_a - y_a)(W_a - Z_a)} + \sqrt{(\chi_a - W_a)(y_a - Z_a)}}{\chi_a - Z_a} \right)^2 \quad \text{or} \quad \frac{R_{11}R_{22}}{R_{12}R_{21}} \leq \left(\frac{\sqrt{(\chi_a - y_a)(W_a - Z_a)} + \sqrt{(\chi_a - W_a)(y_a - Z_a)}}{\chi_a - Z_a} \right)^2, \delta^2 \geq 4\varphi\eta.$$

In this situation, if $\chi_a \exists W_a$, $y_a \exists Z_a$, the $\chi=0$ is unstable while $\chi=1$ is stable, at this time, $\varphi, \eta \geq 0$. But if φ and $\eta \geq \frac{\delta}{2}$, so $\chi_1' \chi_2$, at this point, χ_1 is the stable balanced point while χ_2 is the unstable point. In other word, it means when $\chi_a - W_a = P_a^s - C_a + R_a^s I_a - P_a \geq 0$, $y_a - Z_a = R_a I_a - C_a \geq 0$, and $\varphi, \eta \geq \frac{\delta}{2}$, $0 \rightarrow \chi_1 \leftarrow \chi_2 \rightarrow 1$. On the other hand, if $\chi_a' W_a$, $y_a' Z_a$, $\chi=0$ is stable while $\chi=1$ is unstable, at this time, $\varphi, \eta' 0$. But if φ and $\eta' \frac{\delta}{2}$, so $\chi_1 \exists \chi_2$, at this point, χ_1 is the stable balanced point while χ_2 is the unstable point. In other word, it means when $\chi_a - W_a = P_a^s - C_a + R_a^s I_a - P_a' 0$, $y_a - Z_a = R_a I_a - C_a' 0$, and $\varphi, \eta' \frac{\delta}{2}$, $0 \leftarrow \chi_2 \rightarrow \chi_1 \leftarrow 1$.

(2) When $(\chi_a - y_a)(W_a - Z_a)$ and $(\chi_a - W_a)(y_a - Z_a)$ are both positive number,

$$\left(\frac{\sqrt{(\chi_a - y_a)(W_a - Z_a)} - \sqrt{(\chi_a - W_a)(y_a - Z_a)}}{\chi_a - Z_a} \right)^2 < \frac{R_{11}R_{22}}{R_{12}R_{21}} < \left(\frac{\sqrt{(\chi_a - y_a)(W_a - Z_a)} + \sqrt{(\chi_a - W_a)(y_a - Z_a)}}{\chi_a - Z_a} \right)^2, \delta^2 < 4\varphi\eta.$$

Under this circumstance, there are only two balanced point in evolutionary game system, respectively are: $\chi=0$ and $\chi=1$.

So if $\chi_a \exists W_a$, $y_a \exists Z_a$. the $\chi=0$ is unstable while $\chi=1$ is stable. In other word, it means when $P_a^s - C_a + R_a^s I_a - P_a \geq 0$ and $R_a I_a - C_a \geq 0$, $0 \rightarrow 1$. But if $\chi_a' W_a$, $y_a' Z_a$, the $\chi=0$ is stable while $\chi=1$ is unstable. In other word, it means when $P_a^s - C_a + R_a^s I_a - P_a' 0$ and $R_a I_a - C_a' 0$, $0 \leftarrow 1$.

(3) When $(\chi_a - y_a)(W_a - Z_a)$ and $(\chi_a - W_a)(y_a - Z_a)$ are negative number, or $(\chi_a - y_a)(W_a - Z_a)$ is positive number, $(\chi_a - W_a)(y_a - Z_a)$ is negative number, $\delta^2 \geq 4\varphi\eta$ is permanently established under this circumstance.

So if $\chi_a \exists W_a$, $y_a' Z_a$, the $\chi=0$ and $\chi=1$ both are stable. At this point, $\varphi \geq 0$ and $\eta' 0$, χ_2 is the stable balanced point. it means when $P_a^s - C_a + R_a^s I_a - P_a \geq 0$ and $R_a I_a - C_a' 0$, $0 \leftarrow \chi_2 \rightarrow 1$. But if $\chi_a' W_a$, $y_a \exists Z_a$, the $\chi=0$ and $\chi=1$ both are stable. At this point, $\varphi' 0$ and $\eta \geq 0$, χ_1 is the stable balanced point, it means when $P_a^s - C_a + R_a^s I_a - P_a' 0$ and

$R_a I_a - C_a \geq 0, 0 \rightarrow \chi_1 \leftarrow 1$.

Here, since the group B only differ in their strategic returns from the group A, so they will not be discussed again. In the case, for non-uniform contact rate, the evolutionary result is related to the contact rate only when $(\chi_a - y_a)(W_a - Z_a)$ and $(\chi_a - W_a)(y_a - Z_a)$ are both positive. In other cases, the contact rate only change the position of χ_1 or χ_2 . Because in the income matrix, $W_a = Z_a = P_a$, so $(\chi_a - y_a)(W_a - Z_a) = 0$, indicating that the contact rate in the system only changes the position of the saddle point without changing the direction of the evolutionary result. This also means that when group A chooses the strategy of not share the resources, increasing or decreasing the contact rate between group A and B only affect the process of evolution but won't affect the evolutionary result.

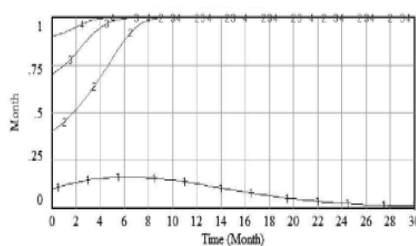
4. SD MODEL OF THE EVOLUTIONARY GAME THEORY AND ITS SIMULATION ANALYSIS

4.1 Simulation analysis of the SD model

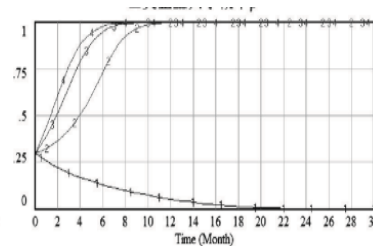
Use Vensim PLE 6.3 to establish the evolutionary game model of sharing enterprise resource under cloud service environment. According to the above analysis, the simulation process in this paper fixed most of the initial value under the premise, from the initial sharing probability, upgrading equipment costs and profit, and technology loss risk factor those three aspects, and with the adjustment of the parameter size, observe the impact of its changes on the evolutionary results.

(1) The influence of the change of initial sharing probability on evolutionary results

Assuming that the INITIAL TIME=0, FINAL TIME=30, TIME STEP=0.5, $P_{as}=2.2$, $P_a=1.2$, $C_a=0.4$, $I_a=0.6$, $r_{as}=0.8$, $r_a=0.4$, $P_{bs}=2$, $P_b=1$, $C_b=0.2$, $L_b=0.6$, $r_{bs}=0.1$, $r_b=0.2$. With the given initial probability $\beta=0.3$ of choosing the shared strategy in group B, we divided the initial probability of choosing the shared strategy in group A into 0.1, 0.4, 0.7 and 0.9, and then observe the changes in the probability of choosing shared strategies between the two types of enterprises. The results can be seen in Fig.1 that when the initial probability of group B is fixed, the trend of evolution of the probability choosing the shared strategy of group A is more significantly affected by their initial probabilities. When the initial sharing probability α is higher, the evolution rate of the probability value converging to 1 is also relatively fast. And also when the initial sharing probability α of group A is higher, the sharing probability of group B will converges to 1; on the contrary, the sharing probability of group B converges to 0, Shown in Fig.2.(among them, $\alpha=10\%$ represent —1—1—1—1—; $\alpha=40\%$ represent —2—2—2—2—; $\alpha=70\%$ is —3—3—3—3—; $\alpha=90\%$ is —4—4—4—4—, and the following is equivalent)



The sharing probability α of group A



The sharing probability α of group B:

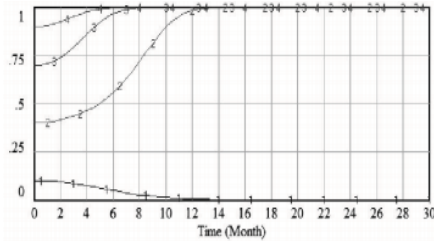
Fig. 1 The influence of the changes of the initial value α on the sharing probability of group A.

Fig. 2 The influence of the changes of the initial value α on the sharing probability of group B.

(2) The impact of the cost and profit of changing the equipment on the evolutionary result

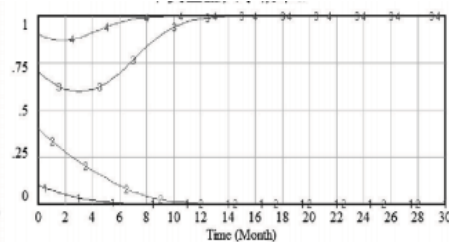
In the other parameters remains the same conditions, we adjust the upgrade equipment cost C_a from the initial 0.4 to 0.6 and 0.9, then get the evolution result shown in Figure3,4. Compared with the Figure 1, when the cost of upgrading equipment in group A gradually increases, the probability of choosing the sharing strategy converge to 1 decreases. In particular, when $C_a = 0.9$, group A with the initial probability 0.4 will eventually converge to 0, which means that group A will choose to abandon the strategy of sharing enterprise resources on

the premise of higher equipment cost. Besides, because the profits is mainly brought by the information improvement factor, so we set improvement factor of information sharing r_{as} of group A rise from the initial 0.8 to 0.9, and the information improvement coefficient r_a rise from the initial 0.4 to 0.6, as shown in Figure 5,6. At this point, the rate of shared probability in group A convergence to 1 is accelerated compared with that of Figure 4, the more prominent change is that the initial shared probability $\alpha=0.1$ converges to 1 in Fig.6. From this we can see that enterprises promote information construction will not only enhance the efficiency of business operation, but also serve as an important guarantee for the cooperation among supply chains.



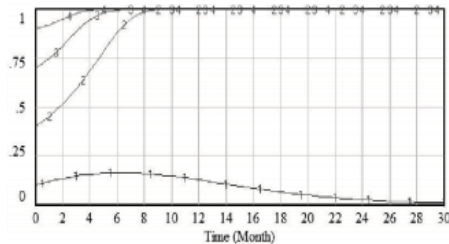
The sharing probability α of group A

Fig.3 The evolution results of the sharing probability in group A when $C_a=0.6$.



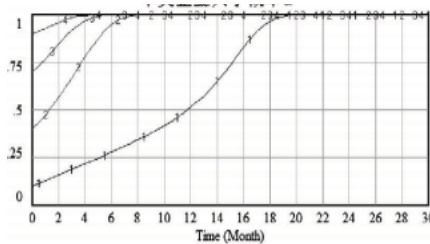
The sharing probability α of group A

Fig.4 The evolution results of the sharing probability in group A when $C_a=0.9$.



The sharing probability α of group A

Fig.5 The evolution results of the shared probability in group A when $r_{as}=0.8$, $r_a=0.4$.

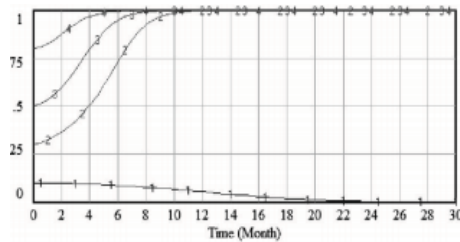


The sharing probability α of group A

Fig.6 The evolution results of the shared probability in group A when $r_{as}=0.9$, $r_a=0.6$.

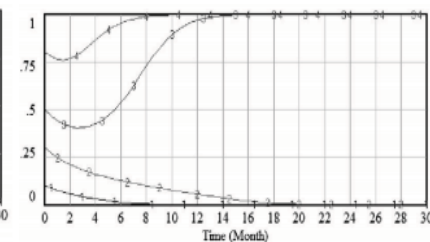
(3) Effect of risk factors of technological loss on evolutionary results

we adjust the technology loss risk factor from $r_{bs}=0.1$ to $r_{bs}=0.4$, then observe the evolutionary result of the shared probability in group B. Comparing Figure 7 with Figure 8, it is not difficult to find that the increase of technology loss risk greatly reduces the rate of initial value $\beta=0.5$ and $\beta=0.8$ converges to 1, and initial value $\beta=0.3$ finally converges to zero. This shows that the probability of choosing a sharing strategy for group B will decrease as the risk of technology loss increases. When the coefficient increases to a certain extent, the enterprise will undoubtedly choose a strategy of not sharing enterprise resources. ($\beta=0.1$ is —1—1—1—, $\beta=0.3$ is —2—2—2—, $\beta=0.5$ is —3—3—3—, $\beta=0.8$ is —4—4—4—)



The sharing probability β of group B

Fig.7 The evolution results of the shared probability in group B when $r_{bs}=0.1$.



The sharing probability β of group B

Fig.8 The evolution results of the shared probability in group B when $r_{bs}=0.4$.

5. CONCLUSION

From the perspective of evolutionary game, this study constructed a strategy selection model of sharing enterprise resources, and uses the method of system dynamics to simulate the model dynamically. The impact of different factors on the evolutionary results is compared and analyzed. Through the discussion, supply chain members can adjust the contact rate between enterprises to make the evolutionary results more favored. Besides, the guiding effect brought by the construction of resource sharing platform and business alliances will make other businesses produce follow-up behavior, which help form the industry rules. And the ration between cost and benefit access to cloud platform is the main factor that enterprise consider when sharing resources, the technology loss risk factor also takes a part. But our research also has some disadvantages, as a resource sharing third-party intermediary, the model in our research didn't include the factors involving cloud service platform provider, besides, the coordinated development of multi-cooperation is also widespread and it makes the evolutionary game model more complex. All of this are the directions of our next research.

ACKNOWLEDGEMENT

This research has been supported by the technical innovation special project of Hubei Science and Technology Program (soft science research) (2017ADC103), the general project of Research base for Humanities and Social Sciences in Hubei Province - Research Center of Enterprise Decision Support (DSS20170302), the general program of China Logistics Society and China Federation of Logistics & Purchasing (2018CSLKT3-068), and the research project of graduate teaching reform for Wuhan Textile University (2017023).

REFERENCES

- [1] Qi Ershi. (2011).Industrial Engineering and Manufacturing Information[M]. Beijin: Machinery Industry Press, 85.
- [2] Tong S.(2008).Contracting and information sharing under supply chain competition. *Management Science*,45(4):701-715
- [3] Chen F, Drezner Z. (2000).Quantifying the bullwhip effect in a single supply chain:The impact of forecasting, lead time and information. *Management Science*, 46(3): 436-443.
- [4] Heng F, Ming J. (2018).The effect of scale economy on supply chain information transaction decision. *Journal of System Science*, 26(2): 111-116. In Chinese.
- [5] Deli Y, Chong W. (2017).A study of information sharing in dual-channel supply chain with risk preference retailer. *Industrial Engineering and Management*, 22(2): 83-88. In Chinese.
- [6] Jizhou L, Gengzhong F, et al. (2017).Factors affecting bullwhip effect of inventory under information sharing. *Journal of Management Science in China*, 20(3): 137-147. In Chinese.
- [7] Peter M M. (2007).A brief history of system dynamics in continental Europe. *System Dynamics Review*, 23(2): 215- 218.
- [8] Galor O, Weil D N. (2000).Population, technology, and growth: from Malthusian stagnation to the demographic transition and beyond. *American economic review*, 17(3): 806-828.
- [9] Friedman D. (1991).Evolutionary games in economics. *Econometrica: Journal of the Econometric Society*,21(3): 637-666
- [10] Jia LuK, Li WenL. (2008). Study of evolutionary stability of a copy equation under non-uniform contact rate. *Journal of Lanzhou University (Natural Science Edition)*, 44(4): 93- 98. In Chinese.
- [11] Taylor C, Nowak M A. (2006).Evolutionary game dynamics with non-uniform interaction rates. *Theoretical Population Biology*, (69): 243- 252.
- [12] Quan J, Wang X J. (2013).Some analytical properties of the model for stochastic evolutionary games in finite populations with non-uniform interaction rate. *Communications in Theoretical Physics*, 60(1): 37-47. In Chinese.

The Analysis on Multimodal Transport Mode of Cross-border E-commerce with 'the Belt and Road' Strategy Based on Personalized Recommendation

Mingli Zhang¹, Yanling Fan², Man Chen³

¹College of Literature, Law, and Economics, Wuhan University of Science and Technology,
Hubei Small and Medium-sized Enterprise Research Center, Wuhan, 430065, China

^{2,3}College of Literature, Law and Economics, Wuhan University of Science and Technology, Wuhan,
430065, China

Abstract: With the further advance of 'the Belt and Road' strategy, China's cross-border E-commerce has obtained powerful policy support and wide world market. But from the view of users' coverage and total import and export of the trade along 'the Belt and Road', China's cross-border E-commerce still has great potential for development, while the high transportation cost is the main resistance in business. Therefore, based on the theory of customer personalized recommendation, combining with the successful cases of personalized services recommendation system from Jingdong and eBay, this article puts forward the multimodal transport service mode of China's cross-border logistics enterprises so as to customize the optimized logistics service system for e-commerce and achieve a win-win situation for customers and enterprises.

Keywords: the Belt and Road; personalized recommendation; cross-border transport; multimodal transport

1. INTRODUCTION

In March 2015, 'The Vision and Proposed Actions Outlined on Jointly Building Silk Road Economic Belt and 21st-Century Maritime Silk Road'^[1] has been issued by National Development and Reform Commissions. In recent years, with the support of national policies, a wave of 'cross-border e-commerce' has been raising in China. According to the report 'eWTP Assisted The Belt And Road Construction -- The Practice of the Alibaba Economy'^[2] issued by Ali Research, Aliexpress has the only cross-border B2C retail platform that cover all countries and regions along the Belt and Road and its users throughout the world located in more than 220 countries and regions. In which, only 45% users are from the Belt and Road countries. According to the National annual statistical bulletin issued by National Bureau of Statistics of the People's Republic of China, China's total import and export to countries along the Belt and Road amounted to 62,205 yuan in 2015, which accounted for 25.3% of the total annual import and export of goods. And in 2016, China's total import and export to countries along the Belt and Road amounted to 62,517 yuan, that accounted for 25.7% of the total annual import and export of goods. The above data shows that China's cross-border E-commerce still has great potential for development in the countries and regions along the Belt and Road. At present, there are problems of high cost, slow efficiency and single transportation form existing in China's cross-border logistics which restrict the development of cross-border E-commerce to a certain extent. Based on the theory of customer personalized recommendation, this paper develops optimized multimodal transport mode for each transaction in order to promote the rapid development of the cross-border E-commerce logistics among 'the Belt and Road'.

2. THE SITUATION OF ROAD TRANSPORT IN COUNTRIES ALONG 'THE BELT AND ROAD'

2.1 The situation on development of highway transportation

Since 2000, the highway freight volume between China and other countries along the Belt and Road have been increased rapidly year by year. From table 2.1, we can see that China's highway freight volume in 2016 is 10 times than that in 2000. In the past 10 years, the highway freight volume of Kazakhstan, Pakistan and Turkey also presents a faster and sustained growth trend. It shows that highway transportation is the dominant way of import and export trade among the transportation of the national and regional along the Belt and Road.

2.2 The situation on development of railway transportation

The railway freight volume of China and Kazakhstan presented a sharp decline. According to table 2.2, we can see that from 2012 to 2016, while the railway freight volume of Iran increased by 1.2 times in the four years. But compared with the railway freight volume from other countries, Iran has problem of low cardinality. And the railway freight volume of Israel has changed little, and the quantity is small. This illustrates that the status of railway transport in countries and regions along the Belt and Road is increasingly weak.

2.3 The situation on development of waterway transport

The waterway transport becomes the main mode of transportation of international trade with its low cost and huge capacity. On the basis of table 2.3, we can see that China waterway transport development is relatively mature, and the container terminal throughput of China, Malaysia, Thailand and India has been increasing year by year from 2012 to 2016 with large cardinal number and its growth is around 1.2 times. This shows that waterway transport has maintained a stable dominant position among the countries along 'the Belt and Road'.

Table 2.1. The highway freight volume between China and other countries along the Belt and Road

Country /Region	Highway freight volume (Million Tons - Kilometer)			
	2000 (year)	2005 (year)	2012 (year)	2016 (year)
China	612940	869320	5137470	6121100
Kazakhstan	/	47120	121070	/
Pakistan	90270	129250	177950	/
Turkey	152210	166830	203070	/

Source: WID database of the World Bank

Table 2.2. The railway freight volume between China and other countries along the Belt and Road

Country /Region	railway freight volume (Million Tons - Kilometer)				
	2012 (year)	2013 (year)	2014 (year)	2015 (year)	2016 (year)
China	2518310	2473477	2308669	1980061	1920285
Kazakhstan	235846	231248	216524	189759	188159
Iran	22604	22400	24461	25014	27243
Israel	1099	1099.233	1099	1155	/

Source: WID database of the World Bank

Table 2.3. The container terminal throughput between China and other countries along 'the Belt and Road'

Country /Region	container terminal throughput (TEU: twenty-foot-equipment unit)				
	2012 (year)	2013 (year)	2014 (year)	2015 (year)	2016 (year)
China	163,372,100	175,805,101	186,852,801	194,755,502	199,565,501
Malaysia	20,588,244	20,910,265	22,367,904	24,012,700	24,570,000
Thailand	7,323,880.75	7,546,522.75	8,119,271	835,945.5	8,239,362.75
India	9,576,716	9,685,160	11,319,000	11,883,003	12,083,010

Source: WID database of the World Bank

Table 2.4. The volume of air cargo between China and other countries along the Belt and Road

Country /Region	volume of air cargo (Million Tons - Kilometer)				
	2012 (year)	2013 (year)	2014 (year)	2015 (year)	2016 (year)
China	15,568.753	16,053.733	17,822.581	19,805.63	21,304.585
Russia	4,132.144	4,249.269	4,413.559	4,761.047	5,863.197
Turkey	1,933.678	2,296.039	2,630.33	2,882.162	3,493.93

Source: WID database of the World Bank

2.4 The situation on development of air transport

In accordance with table 2.4, we see that in air transportation, the volume of air cargo in China, Russia and Turkey has shown a growing trend, but the volume of goods transported is small because of high cost of air transportation. The air transport of the countries along 'the Belt and Road' is at an initial phase, but is still keep on developing slowly. This shows that there is a large development space in air transport.

3. THE PROBLEMS AND DIFFICULTIES OF CHINA'S CROSS-BORDER E-COMMERCE LOGISTICS UNDER 'THE BELT AND ROAD' STRATEG

3.1 High cost of logistics transportation

Retailing is generally the main marketing model of the cross-border e-commerce platform. The users only want to pay limited amount cost in the logistics transportation when they buy small amount of goods with low price. But the link of cross-border logistics is very long, including domestic cargo transportation and customs clearance, international cargo transportation, foreign cargo customs clearance, transportation and so on^[3], which can lead to high logistics costs. The costs of cross-border logistics mainly include transport costs, customs duties and VAT (value added tax) and overseas logistics cost, etc. Even if the user selects the slowest postal international package with the most favorable cost, the logistics cost is still much higher than that of general logistics. The cost of international express will be much higher. Once the user has additional requirements for the timeliness or safety, the logistics cost will be increased correspondingly.

3.2 Simplification of transport form

At present, logistics of China's cross-border e-commerce is immature with single transport mode. The modes of transports are ineffective and wasteful in logistics infrastructure and energy without effective connection. In the past, China's cross-border logistics enterprises did not adopt the multimodal transport mode and different kind of transportation mode has individual information system^[4]. Therefore, most of China's logistics enterprises did not have the ability to use multiple transportation modes flexibly.

3.3 Unsmooth of logistics information

Under normal circumstances, the users hope to know about exactly what they have purchased, where they are and prepare for the reception of the goods after shopping on the cross-border e-commerce platform. With the unlikeliness between cross-border logistics and domestic logistics, cross-border e-commerce logistics is mainly divided into two parts of domestic logistics enterprises and overseas ones. It is difficult for logistics enterprises and overseas logistics enterprises to communicate and check transport logistics information because of the inconformity of information degree between domestic part and overseas part and obstacles in language communication. These factors will weaken users' consuming experience in cross-border e-commerce greatly and cut down the sales of e-commerce platforms to a certain extent^[5].

3.4 Unbalance of freight development

According to the official data issued by National Development and Reform Commissions, the volume and turnover of railway freight traffic began to decrease year by year from 2012. In 2014, railway freight volume accounted for 8.67%, highway freight volume accounted for 77.29% and railway turnover accounted for 21.19%, while highway turnover accounted for 46.97%, so railway lost the status of big artery. In 2015, the turnover of railway goods dropped by 14%. From January to May in 2016, the freight turnover of the railway dropped by 10.3%, and the decline trend of railway freight still continues. On April

19, 2017, the document of 'The Development Planning of Railway Container Multimodal Transport in 13th Five-Year'^[6] issued by National Development and Reform Commissions, Ministry of Transport of People's Republic of China and China Railway Corporation. According to 'the planning', it presents that the proportion of railway freight in the United States and the European Union is about 40%, while the proportion of that in China is only 7.6%. This data shows that the development of railway transportation in China still falls behind. The railway transportation has the advantages of large freight volume, low freight rate and high safety. Therefore, the China's cross-border logistics enterprises should enhance the optimal combination of railway transportation and other modes of transportation.

Compared with the developed countries, China's cross-border logistics has a lot of problems, such as single form of logistics development, high cost, small coverage and low specialization level and so on. However, in strategic planning of 'the Belt and Road', transportation industry is in the basic position. For the implementation of the Belt and Road strategy, logistics is the key to solve the existing problems of cross-border trade. There is a logistics line from Guangzhou to Harbin, which is operated by a logistics company in Liaoning. In which, the length of a single road transport line is 3,390 kilometers, and the freight rate for each standard container is 7,800 yuan. After using the multimodal transport of water, road and railway, the length of the line is shortened to 3,160 kilometers, and the freight rate for each standard container is 4,900 yuan, which can reduce the transportation cost by 37.2%^[7].

On April 21, 2016, the first train of 'joint transport of highway and railway' mode was set up at Duoluokou station in Wuhan to the terminal station Enshi in Hubei province. According to table 3.1, we can see that if a single road transport is adopted, the logistics enterprises need to spend 4,600 yuan. After the joint transport, the freight rate will be 2,635 yuan, and the transportation costs have been reduced by 42.7%^[8]. Thus it can be seen that the flexible use of the arrangement and combination of various modes of transportation will greatly reduce the logistics cost. For the long distance and multiple batches in cross-border logistics, multimodal transport will bring more economic benefits to cross-border logistics enterprises and will also promote the rapid development of cross-border e-commerce.

Table 3.1. The cost comparison of two transportation modes

Starting place	Single transport cost (yuan)	Multimodal transport costs (yuan)	Cost saving
Guangzhou- Harbin	7,800	4,900	37.2%
Wuhan - Enshi	4,600	2,635	42.7%

4. A SUCCESSFUL CASE ANALYSIS OF CUSTOMER PERSONALIZED RECOMMENDATION SYSTEM FOR CROSS-BORDER E-COMMERCE ENTERPRISE

4.1 The theoretical analysis of customer personalized recommendation system

In 1990s, the research on personalized recommendation has just begun. Personalized recommendation is the system which can infer users' interested information, service or merchandise through learning and mining users' historical purchase behavior and comments, so as to generate recommendation list which can be recommended to users. Personalized recommendation system can help people find all kinds of resources under the network, so it can save time and cost through recommending information to users initiatively when people are looking for target information.

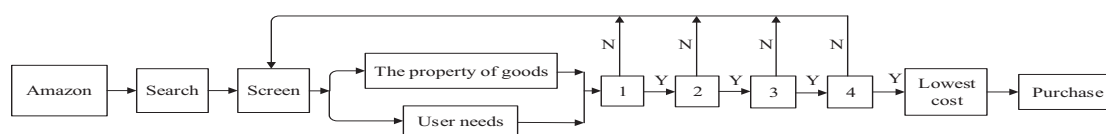
The core content of the personalized recommendation system is the algorithm. In China, many cross border e-commerce enterprises who do well in personalized recommendation adopt a classical personalized recommendation algorithm: Recommendation algorithm based on utility^[9] and recommendation algorithm based on association rule. Recommendation algorithm based on utility refers to the enterprise analyzes the

matching of good's attributes and user's needs by qualitative or quantitative analysis method, so as to meet the user's subjective and objective demands from the macro perspective, and to maximize the effect of recommendation service. Recommendation algorithm based on association rule^[10] refers to the process of an enterprise recommends products to user through analyzing the user's purchase history information to conclude a rule condition and use the rule condition to predict the user's purchase target good. This algorithm is a very basic method which is widely used in all kind of e-commerce enterprises.

4.2 Case analysis of customer personalized recommendation system for e-commerce enterprise

Since 2001, Amazon set 'customer-centric' as its service goal, and started the exploration of customer personalized service. According to table 4.1, we can see when linking the Amazon shopping interface, we will input the goal merchandise in the 'search' column, and in the upper right corner of the interface will appear a 'screening' option, here we can find all kinds of option related to product's material, price, size and comment, when we making decisions, we will take those options, our needs and payment ability into account. This is the recommended algorithm based on utility, it can quickly match the commodity's attribute with user's need, so as to effectively help users select their target goods with lowest cost. It is a win-win method, which not only satisfy the user's needs, but also improve the sale volume in the unit of time of e-commerce platform.

In 2013, with the advent of the big data era in domestic, Jingdong developed very fast in this year. Supplying multi-screen and multi-type products is the result of Jingdong's personalized recommendation system, which is the recommendation algorithm based on association rule. According to table 4.2, we can see the main method is that the e-commerce enterprise predict the customer's target product and recommend it to the customer based on his or her behavior records in the Jingdong and the other related platforms, such as analysis of customers browsing, shopping cart, attentions, searching, purchasing, and comments in the Jingdong shopping platform, and commodity information browsing on Tencent QQ and WeChat. This recommendation is a process in which the e-commerce platform recommend product to meet consumer demand from a macro perspective, which can make personalized recommendation more accurate, and help customer quickly select a product, and improve the loyalty and experience of consumer, so as to improve the quality and efficiency of user's shopping decisions and shorten the user's shopping path.



Notes: Y represents satisfaction, N represents not satisfaction

Figure4.2. Recommendation algorithm based on association rule

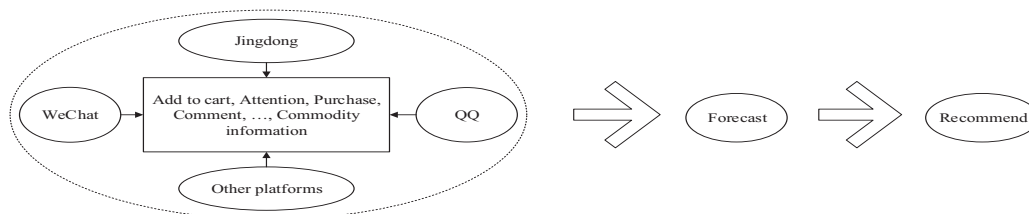


Figure4.1. Utility-based recommendation algorithm

5. THE CONSTRUCTION OF MULTIMODAL TRANSPORTATION OF CHINA'S CROSS-BORDER LOGISTICS BASED ON CUSTOMERS PERSONALIZED RECOMMENDATION UNDER 'THE BELT AND ROAD' STRATEGY

5.1 The personalized multimodal transport mode of recommendation algorithm based on utility

In April 21, 2017, Ali Research and DT Financial Research Institute issued the report named 'EWTP assisted The Belt and Road construction -- the practice of the Alibaba economy'. According to the report, the top five countries' export volume of China's cross-border e-commerce businesses are Russia, Ukraine, Israel, Belarus and Poland. Most popular products are Mobile phones and spare parts, fashion jewelry, women's clothes, clothing and accessories, and nail products. The top five countries' import volume of China's cross-border e-commerce businesses are Thailand, Singapore, Malaysia, Israel and Czech. China has a high demand for latex pillows and latex mattresses of Thailand, whose sales volume reach nearly 280,000 in Tmall international last year; in 2016, Israel's total commodity turnover has an increase of 371% in the Tmall international, and China has the most demand for personal care products from Israel.

According to the report of Ali, the commodity's attributes of cross-border electronic business for export and import is different, each product has a different volume, timeliness and value with each user also has different consume need. Therefore, China's logistics enterprises can set different options of commodity attributes with the help of cross-border e-commerce platform by predicting the intermodal transport method preferred by customers so as to recommend best multimodal transport model as followed.

5.1.1 Highway-railway transport: suitable for customer demand of multiple batches and high timeliness

Highway-railway transport is a more efficient delivery method, which can combine the advantages of highway and railway effectively on the one hand: combining the safety, punctuality, cheapness of railway transportation and the flexibility, convenience and efficiency of highway transportation, and overcome their disadvantages on the other hand: overcoming the slow speed, less network, low efficiency of railway transportation and high cost of highway transportation. The users not only can get the benefits of combined railway and highway transport, but also can appoint the location, shipping time and delivery time of goods according to their own needs. Therefore, when the users need the logistics service to transport good with small amount or multi batches or high timeliness, they can choose the highway-railway transport.

5.1.2 Sea-railway transport: suitable for customer demand of high-volume, long distance and low cost.

Railway transport is suitable for mass goods transportation, with high safety factor, less influencing from natural factors, punctuality and low freight. The sea transport has the advantages of larger volume, long distance, flexibility and low freight. Sea transportation and railway transportation both have the advantages of low cost, large volume, connection convenience and low freight. Sea-railway combined transport not only can control costs but also can solve the shortcomings of the limited coverage of single transport model. This transport method can fully connect ocean and land, and it can not only help users to purchase bulk goods from different regions at a low price, but also to achieve goods with high quality and integrity through the convenience of sea-railway transport. When customers want to deliver goods with large volume, long distance and low freight, they can choose the sea-rail transport.

5.1.3 Land-air transport: suitable for customer demand of high value, high timeliness, small size

If air transport is used only, it will limit the goods' amount and the cost is higher. If using domestic charter flights, the cost will be even more expensive. Therefore, in the case of large quantity of goods, we often choose the method of land to airport, and to international flight. Because the highway transportation is more flexible in the distribution, so land-air transport is usually being chosen. Land-air transport is a

multimodal transport model combined with land transport and air transport. There are 3 forms: a joint of railway, highway and air transport, a joint of railway and air transport and a joint of highway and air transport. The multimodal transport under the air and land combination model integrates the advantages of the quickness and safety of air transportation with the extension and flexibility of the highway, and realizes the seamless connection between them. When customers want to deliver goods with high value, high efficiency, small volume, they can choose the land-air transport.

5.1.4 Sea-air transport: suitable for customer demand of low cost, high efficiency and long distance.

Sea-air transport is a mode of transportation combined with sea transportation and air transportation. They all aim at providing fast and reliable transport services at low freight rates. Compared with single sea transport, the biggest advantage of Sea-air combined transport is saving time. While compared with single air transport, the biggest advantage of Sea-air combined transport is saving cost. This mode combined the low cost of sea transportation with the high speed of air transportation. In general, most part of this transportation is borne by sea and the final delivery section is borne by air so as to give full play to the unique advantages of both sides. When customers want to deliver goods with low cost, high value and long distance, they can choose the Sea-air transport.

5.2 Personalized multimodal transport model of recommendation algorithm based on association rule

There are six main types of countries along the ‘Belt and Road’, including China, Mongolia in east Asia, 18 countries in West Asia, 8 countries in South Asia, 5 countries in Central Asia, 7 countries in the Commonwealth of Independent States and 16 countries in Central and Eastern Europe. Every country and region along the route has different geographical features and advantages. Therefore, cross-border logistics enterprises can carry out the personalized recommendation of the user's commodity transportation mode from a macro perspective according to the overall geographical features of each region by using the recommendation algorithm based on association rule as followed.

Table 5.1. Personalized multimodal transport model of recommendation algorithm based on utility

Intermodal approaches	Product attributes	Customer needs	Advantages
Highway-railway transport	small amount, multiple batches, high timeliness	consumer goods	flexibility
Sea-railway transport	high-volume, long distance, low cost	bulk goods	low price
Land-air transport	high value, high timeliness, small size	High-end goods	efficiency
Sea-air transport	Low cost, high efficiency, long distance	fresh or high value-added products	Cheap and efficiency

Table 5.2. Personalized multimodal transport model of recommendation algorithm based on association rule

Intermodal approaches	Area	Topographical features	Climatic conditions
Sea-air transport	Southeast Asia, South Asia	Peninsula, Archipelago; Straits of Malacca	Tropical rain forest, tropical monsoon
Sea-rail transport	Central Asia, West Asia	More plains, hills; ‘sea traffic arteries’	Temperate continental
Land-air transport	China, Mongolia, the Commonwealth of Independent States and Central and Eastern Europe	Multiplayer, plateau	Mainly temperate continental

5.2.1 Sea-air transport: suitable for the freight to other Southeast Asia and South Asia countries

Southeast Asia has many peninsulas and islands, and the South Asia's topography is high in the north and south and low in the middle. Therefore, it is very inconvenience to developing land transport in this area. While most of the climate types in Southeast Asia and South Asia are belong to the tropical rain forest and tropical monsoon climate with abundant water. Southeast Asia is located at the crossroads of Asia and Oceania, the Indian Ocean and the Pacific Ocean. It is an important hub of world maritime transport and air transport. The Malacca Strait between Malay Peninsula and Sumatra Island is a natural waterway connecting the Pacific Ocean and the Indian Ocean. It is also an important waterway connecting the ports of Europe, the Indian Ocean and the Pacific's west bank.

In addition to the advantage of natural ports and air, economic factors are the main reason for the countries and areas in Southeast Asia and South Asia along 'the Belt and Road' choosing sea-air transport. In November 3, 2017, Ali announced the launch of the first digital free trade zone in Malaysia, which will create a complete logistics ecosystem, reduce costs and increase trade volume. Once implemented, all packages are expected to be delivered to ASEAN countries within 72 hours via the well-established air and sea connections at Kuala Lumpur International Airport and Klang Port, Malaysia. For the users in this region, China's cross-border logistics enterprises should recommend the mode of sea air transportation.

5.2.2 Sea-rail transport: suitable for the freight to other countries of Central Asia and West Asia

Central Asia is located in the Asian continent, its terrain is mainly plains and hills and its climate is temperate continental climate-based. And Central Asia is the area that passes by the ancient Silk Road in China and the Second Eurasian Continental Bridge in modern times, so it has a good natural and economic basis for rail transport. And West Asia is an important area connecting the Atlantic, Indian Ocean, Asia and Europe, with its location among Mediterranean, the Red Sea, the Black Sea, the Arabian Sea and the Caspian Sea. It has two superior maritime transport routes: one is the Persian Gulf - the Strait of Hormuz - the Mandeb Strait - the Suez Canal - the Strait of Gibraltar route, another is the Black Sea - the Turkish Channel - the Mediterranean route. Therefore, Sea-rail transport should be taken for China and other countries in Central Asia, West Asia along 'the Belt and Road'. For the users in this region, China's cross-border logistics enterprises should recommend the mode of Sea-rail transport.

5.2.3 Air-rail transport: suitable for the freight between China and the countries of the Commonwealth of Independent States, Central and Eastern Europe along 'the Belt and Road'

The Commonwealth of Independent States (CIS) stands for the Commonwealth of Independent States after the dissolution of the former Soviet Union. It is located in the northern part of Asia and the eastern part of Europe. Central and Eastern Europe and North Asia are characterized by temperate continental climate, with multi-plains and plateaus, flat terrain, and are suitable for the development of land-air transport. Transport in the region has been dominated by rail transport, with a radial Moscow-centered railway network in the west, the Trans-Siberian Railway in the east and a second Eurasian continental bridge running through the region. In addition, in recent years, Hainan Airlines Group strove to build an air channel connecting China with the countries along 'the Belt and Road'. At present, there are 65 international routes along 'the Belt and Road'^[11]. The region has an excellent railway infrastructure and an opportunity for rapid development of air transportation. Therefore, air-rail transport should be taken for China and other countries in CIS and Eastern European along 'the Belt and Road'. For the users in this region, China's cross-border logistics enterprises should recommend the mode of air-rail transport.

6. CONCLUSION

For the cross-border e-commerce logistics enterprises, they should firmly grasp the strategic opportunity of 'the Belt and Road' and give full play to their own advantages. And they should formulate a plan for the transport of goods that meet the customer's own actual needs, combined with the advantages of multimodal transport, to play a variety of transport advantages, promote the development of the rail transit actively, co-ordinate the integration of sea-railway transport scientifically, expand the market of land, sea and land transport continuously, and to strive to achieve the optimal combination of sea and air transport. It is significant to contribute its own strength to the smooth the realization on China's Belt and Road and create a beautiful win-win situation for cross-border e-commerce logistics enterprises and customers.

ACKNOWLEDGEMENT

This research is supported by Hubei Provincial Department of Education (Project Grant No.:16D014); supported by Library and Information Committee for Academic libraries of Hubei Province (Project Grant No.: 2017-YB-04); supported by Library and Information Committee for Academic libraries of Hubei Province (Project Grant No.: 2017-YB-03).

REFERENCES

- [1] National Development and Reform Commission. (2015). The vision and proposed actions outlined on jointly building Silk Road Economic Belt and 21st-Century Maritime Silk Road.
- [2] Ali Research. (2017). eWTP assisted The Belt and Road construction -- the practice of the Alibaba economy.
- [3] Xujie Yang. The logistics difficulties and countermeasures of China's cross-border E-commerce enterprises[J]. China Journal of Commerce, 2016(8):132-133.
- [4] Rundong Liao. The difficulties and countermeasures of the retail export of small and medium-sized foreign trade cross-border enterprises[J]. Enterprise Economy, 2017(11):62-67.
- [5] Xiaheng Zhang. The difficulties and Countermeasures of China's cross-border e-commerce logistics[J]. Contemporary Economic Management, 2015(5):51-54.
- [6] National Development and Reform Commission, Ministry of transportation and transportation, China Railway General Corporation. (2017). The development planning of railway container multimodal transport in '13th Five-Year'.
- [7] People's Daily. (2017). Objective: the development of multimodal transport flow cost reduction. This is the key.
- [8] Hubei Daily. (2016). The first trip of 'road and railway transportation' freight train in our province.
- [9] Taiju Liu. (2013). The Web personalized service model based on utility. M Thesis. Beijing: Beijing University of Posts and Telecommunications
- [10] Jing Wang. (2012). The Design and implementation of personalized recommendation system of book sales website based on association rules. M Thesis. Chengdu: University of Electronic Science and Technology of China
- [11] Guangzhi Ling, Huiyu Wang. (2017). The survive of a mid-Europe airport due to 'The Belt and Road'.

Scheduling Limited Resources in Engineering Projects

Samer Ben Issaa¹, Yiliu Tu^{1}*

¹Schulich School of Engineering, Department of Mechanical and Manufacturing Engineering,
University of Calgary, Canada

Abstract: Because of high customizations in the one-of-a-kind production companies (OKP companies), these companies need to find a way for reducing the production cost, shortening the production lead-time, and maintaining the quality of the productions as in the mass production system MP.s. Currently, production scheduling in OKP system follows the traditional mass production system, which focus on time and inventory, and it is inapplicable. Actually, OKP system works based on customer requirements, where each order can be representing as multi-project based. In this paper, One-of-a-kind production OKP has been referred as a project-based production and as a flexible resource- constrained project scheduling problems (FRCPPs); because in practice, some of project activities cannot be pre-determined due to its high customizations and great uncertainties. A new model has been proposed based on these assumptions to create production schedules for OKP system, which focuses on time and resources as in project management system PM. s, and deals with the problem which have three categories of project activities A, B, and C. The per-findings indicated that the model enhances the applicability of resulting schedules, emulates what a project manager in practice does (i.e. adding or removing resources from tasks to have the project completed in time), increase the number of feasible solutions, and reduces the project duration.

Keywords: one-of-a-kind production system OKP. s, Flexible Resource-constrained project scheduling problem FRCPP, Mass production system MP. s, Project management system PM.s.

1. INTRODUCTION

The classic version of the resource constrained project scheduling problem (RCPSP) normally consists of one mode. An Activity duration and resource quantity required must be predetermined for each project activities, and normally they are non-preemptive. The typical aim of the RCPSP is to create a schedule that minimizes the total project duration, subject into two constraints: 1) precedence constraint, which means an activity cannot be started until the preceding activities in the project network are finished (finish-start relationship with zero-time lag); and 2) resource availability constrain, which means resources are limited during the execution of an activity period. The time constraints are caused by the relationships between each activity and the durations of the activities. Resource constraints rise from the relation between the availability of a resource type at a given time and resource required of all activities being executed at that time. Meanwhile, each activity in Multi-mode Resource-Constrained Project Scheduling Problems (MRCPP) can be accomplished in one out of several modes, which means an activity in RCPSP has been extended by allowing several alternatives (modes) in which an activity can be performed. Each mode reflects a feasible way to combine an alternative duration with different levels of resource requirement to accomplish the activity (i.e. project activities can be pre-planned into one or more running modes). Moreover, resources assignment over the activity duration in (RCPSP) and (MRCPP) are normally constant. In the RCPSP & MRCPP, there are two different activity categories have

been considered: 1) Category A, which referred to non-preemptive activities with constant resources profile (CRP) as depicted in Figure (1A): Project activities are displayed as rectangular shape, where X-axis represents the duration of the activity (5), and Y-axis represents the resource required per time unit (2). Category B, which referred to preemptive activities with constant resource profile (CRP): Activities can be executed by several different ways as depicted in (Figure 1B).

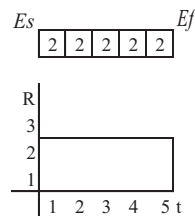


Figure 1A. Non-preemptive activity executes by constant resource profile (CRP-RCPS).

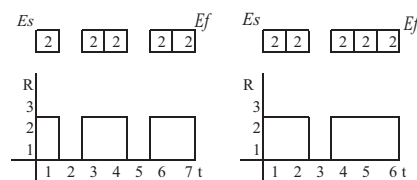


Figure 1B. Preemptive activity executes by constant resource profile (CRP-MRCPS).

The RCPS problems have been well documented, and they have been gained high attention. Meanwhile, there are several models have been proposed to find the best solution for MRCSPs ^{[1]-[2]-[3]-[4]}. MRCSP is NP-hard problem, and if there are at least two non-renewable resources, the problem of finding a feasible solution is already NP-complete ^[20]. Genetic algorithm presented for the preemptive and non-preemptive resource constrained project scheduling problem^[7], a model formulation and solution method for RCPS established where assumed that the project structure should be provided in advance^[8], and lastly, Branch and Bound model (B & B) modified to solve MRCSPs by allowing activities to be splitting^[9], and a new criterion called Minimum Moment Of Resource Required around X-Y axes (MMORR) has been used to find the best feasible solution as indicated in (Figure 2). However, none of the previous models can be used for solving large-sized projects in reasonable computation time (the number of project activities which have been used in these models are less than 100 activities ^[10]), and they used only two categories of activity.

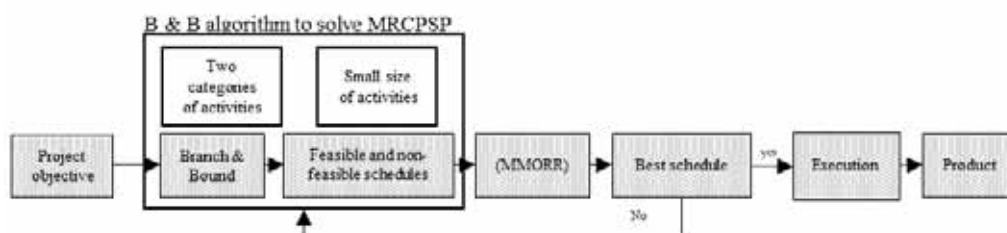


Figure 2. The B & B model used to solve MRCSPs.

The MRCPS problem has been extended, where the activities can be classified into three categories as follows: 1) Category A; Non-Preemptive activities with a constant resource profile (CRP); 2) Category B; Preemptive activities with a constant resource profile (CRP); and 3) Category C; Preemptive or non-preemptive activities with Flexible Resource Profile (FRP), which means variable units of resource (Y-axis) may work on an activity in each time unit due to expected or unexpected reason, and they can freely split along time duration (X-axis) during the execution as are depicted in (Figure 3a, 3b, 3c and 3d) as examples. In such a case, the total amount of resource required to complete an activity over its duration must at least satisfy its resource required. This type of project activities is in general characterized and defined by a work contents^[11], e.g. person-day or man-hour...etc. The number of resource units allocated to execute an activity of this category usually may vary over execution-time. Consider e.g. the example in (Figure 3) which corresponds to a work content of 10 resource - time units; this task can be processed within 10-time units by one resource unit (1x10) each day as illustrated in (Figure 3a), or can be processed by several other ways such as (1x3, 2x2, and 3x1), (3x2, 1x2, and 2x1), or (2x1, 3x2, and 1x2) as illustrated in (Figure 3b, 3c, and 3d) respectively.

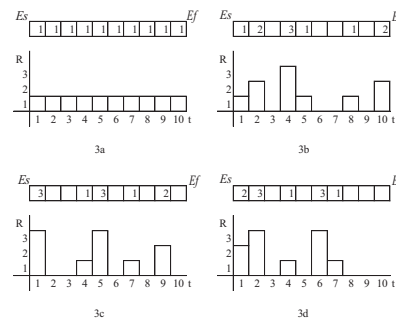


Figure 3. FRP-Activity executes by using a flexible resource rate (FRCPS).

Defining the work-content per activity and resource type is quite similar to the MRCPS but adds greater flexibility during the scheduling process. In the latter category, neither resources usage nor activity duration are known. The problem called MRCPS problem with flexible resources (FRCPS), which is easy to define, but so difficult to solve. The FRCPS deals with a set of N activities which must be scheduled by somehow to minimizing project duration, subject to: 1) Precedence constraint; 2) Resources available during the project execution are limited; 3) Splitting activities is not allowed; 4) Minimum block length constraint, which means that there must be at least a number of a consecutive period having a constant resource usage; 5) Maximum and minimum resources profile confining the resource allocation per time to be within a specific range; 6) Synchronization for resource allocation to each activity^[12].

The resources required by an activity are classified into three types: 1) A principle resource is the resource which used as basis for other resources to process the activity; 2) A dependent resource is the resource which its use dependent on its principle resource; And 3) An independent resource is a resource whose allocated quantity is independent from any other resources.

They are sorted as follows; 1) Renewable resources which are available in limited quantities during each unit time; and 2) Non-renewable resources which are limited for the entire project.

A hybrid metaheuristics model proposed for FRCPSPs where for each activity the given resource requirement is allocated in a variable number of period during the activity execution ^[6], and the resource constrained project scheduling problem addressed with flexible resource profile where each activity may start, end or change its resource allocation at any point of time ^[12]. However, in the practice, Project Manager deals with the following project scheduling problem: Some of project activities, that are pre-planned and specified as non-preemptive activity with (CRP) as category A, some other project activities are pre-planned and specified as preemptive activities with (CRP) as category B, and the rest of project activities are assumed and specified as work content and have to be processed by the work-content resource (FRP) as category C. Thus, the new model deals with the problem which have three categories of project activities A, B, and C, assumes that each activity in category C could be splitting during the execution. In this paper, we will refer the One-of-a-kind production (OKP) as a Project-based production, and as FRCPS because some of the project activities cannot be predetermined due to its high customization and great uncertainties. The per-findings indicate that the model enhances the applicability of resulting schedules, emulates what a project manager in practice does (i.e. adding or removing resources from tasks to have the project completed in time), increase the number of feasible solutions, and reduces the project duration.

The remainder of the paper is organized as follows. Section 2 provides the literature review. While section 3 describes the problem statements. Section 4 illustrates the methodology. Section 5 addresses numerical example, and section 6 conclusion.

2. LITERATURE REVIEW

The first studying for the flexible Resource-Constrained Project Scheduling Problem (FRCPSP) was in 2006 in the context of real-world application in pharmaceutical research and development projects and deals with the lead optimization phase of pharmaceutical research where a number of leads (molecules as a basis for potential drugs) are proposed by different department in order to optimize their biochemical ^[13]. The project scheduling problem with work-content constraints formulated as a mixed-integer linear program for small instances to assess an existing method ^[16]. A procedure has been proposed to find all feasible work profile of each project activity instead of a fixed duration and resource requirement, the total work content is given, and using genetic algorithm to schedule activities ^[17]. A priority-rule scheduling method proposed to develop a feasible serial schedule-generation scheme and to determine a feasible resource-usage to each period. The usage of the work-content resource can be varied between a lower and an upper bound and must be integer ^[11]. Mixed-integer linear program MIP-based heuristic proposed to schedule the activities sequentially. The heuristic starts by ordering the activities in a precedence-feasible list, the activities are then scheduled one at a time. Each time a prescribed number of activities has been scheduled, a subset of activities is rescheduled. The inefficient resource allocation can be reduced by considering multiple activities simultaneously in the rescheduling phase ^[18]. Mixed-integer linear programming formulation presented to solve the FRCPSP and to show that the problem can be solved efficiently for small and medium sized instances using a commercial solver ^[19]. A hybrid metaheuristic HM to solve the FRCPSP has been proposed, the best schedules are improved in a variable neighborhood search by transferring resource quantities between selected activities. The HM consists of a

genetic algorithm GA combined with a variable neighborhood search ^[6]. Four discrete-time model formulation for the FRCPSP presented and investigate the model efficiency in terms of problem size, solution quality, and runtime ^[15]. A mixed-integer linear programming model proposed to solve FRCPSP, using the continuous-time system to synchronize activities and resources ^[14].

In short, to the limit of our knowledge, none of all the papers mentioned above have assumed that there are three categories of activity A, B, and C (Preemptive, Non-Preemptive, and flexible resource profile activities) to solve the multi-resource constrained project scheduling problems as is in OKP activities.

3. PROBLEM STATEMENT

In this paper, a project consists of $V=\{1, \dots, n\}$ activities as well as dummy source activity 0 and dummy sink activity $n+1$ to be scheduled. The activities are categorized to be $(A, B, \text{ and } C)$, where category A is defined as activities that must be executed by assigning constant resource profile (CRP), and activities splitting is not allowed, category B is defined as activities that must be executed by assigning constant resource profile, and activities splitting is allowed, and category C is defined as activities that can be executed by assigning flexible resource profile and activities splitting is allowed. For category A and B ; durations and resources required are known in advance, while they are unknown in category C . The work-content that essentially represents how much of work has to be applied for executing an activity in category C should be considered. The One-of-a-kind production (OKP); is a production system to produce customized products within a product domain, which means productions are characterized by product designs that change with every order. Usually the quantities are small or just one. The characteristics of OKP.s can be summarized as:

- Great uncertainties in production control.
- High customization.
- Complicated and dynamic production system.
- Creating production schedule and control is extremely difficult.
- Production planning and scheduling is represented as multi-project schedules and sometimes is represented as a mega project scheduling problem, and that is because of the daily large numbers of the customer requirement.
- A one-of-a-kind production (OKP) system is a Flexible Manufacturing System with lot size of one ^[21].

Currently, production scheduling in OKP system focuses on time and inventory. It follows the traditional mass production system. OKP system works based on customer requirements, where each order can be representing as multi-project basis. Meanwhile, customer expects these orders to be delivered in shorter time and at nearly the same cost as mass production with high quality. Projects manager have been tried to apply production schedules of mass production system in OKP system. However, those production schedules are inapplicable because of resource constraints that imposed on the project. Moreover, project management system has been applied mainly for civil constructions because it focuses on time and resource constraints. Thus, OKP system can be considered as project management system because its production-schedules focus on time and resource constraints too. As a result, the objective of this research is how to create the production schedules of

PM.s in OKP.s, reduce the variable cost of customized production system (OKP), shorter the lead-time, while maintaining the same quality level. Therefore, creating a fast method or an algorithm to speed up the setup times of scheduling and rescheduling the production in OKP companies from time to time can be effectively way to achieve these goals (customer satisfaction). The engineering efforts usually take the major part of the variable cost in this type of production system, and one of these efforts is process planning and production scheduling and control ^[21]. (Figure 4) demonstrates the effects of applying the model proposed on customer satisfaction and what are the benefits expected comparing with other production systems (i.e. use the production schedules of PM. s in the OKP.s).

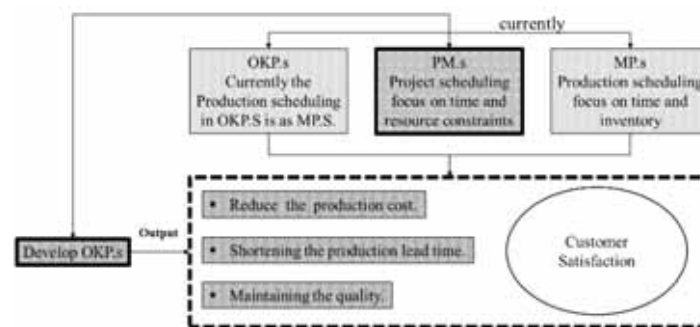


Figure 4. The customer requirements by applying the model proposed

4. METHODOLOGY

A new model has been addressed the OKP projects, which represents as Resource-Constrained Project Scheduling Problem with Flexible resources (FRCPSP) for determining the starting time, resource profile, and activity duration in order to minimize project duration, with allowing activities to be splitting as indicated in (Figure 5). In OKP projects, there are some activities can be pre-determined and executed as are indicated in (Figure 1A & 1B), and some others can be pre-determined as work content and can be executed by different ways as indicated in (Figure 3). Due to high customization and grate uncertainties in OKP, Project manager should find an effective algorithm to schedule these three-different types of activities to obtain a best feasible solution out of several other solutions. To put it another way, the objective of this paper is;

- To generate a guideline or method to find how automatically split activities under deferent constraints.
- To adapt quickly the changes of production scheduling from time to time.
- To deal with a multi-project and multi-size activities in each project.

Furthermore, the previous models have been assumed that all activities in FRCPSPs can be split. which is not true according to the activity categories that we have assumed and mentioned in this paper. From (Figure 5), the FRCPSPs in OKP system can be characterized as Flexible Resource Profile activities, Flexible project structure,

multi-project schedule, and the activities are large-sized (e.g. daily orders are 2000-2500 in some windows and doors production companies which categorized as OKP companies).

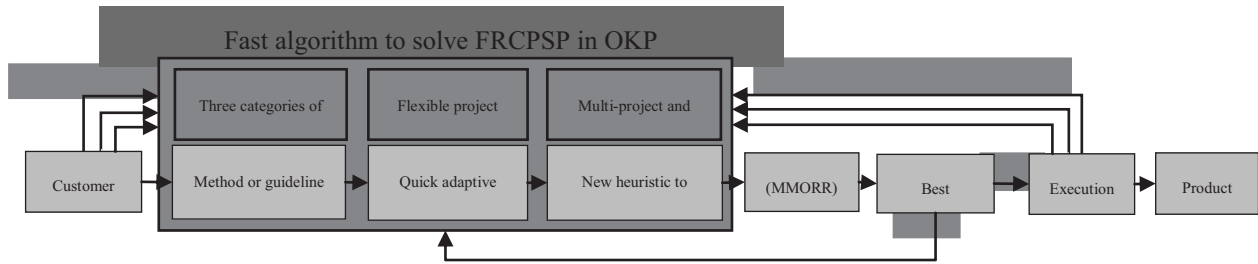


Figure 5. The new model proposed to solve FRCPSPs in OKP.s

Managing OKP activities, as project production based, increases both the number of the feasible solutions for the problem and the efficiency usage of the resources and decreases the project duration. In the model proposed; we are going to attack the RCPSP, MRCPSp and FRCPSp in OKP as follows: if project activities included only the pre-determined activities i.e. category A&B, whether single or multi-mode, the model defines and deals with the problem as RCPSP or MRCPSp. In case the project activities included the non-predetermined activities i.e. Category C, in addition to other categories A&B, the model defines and deals with the problem as FRCPSPs, as depicted in (Figure 6).

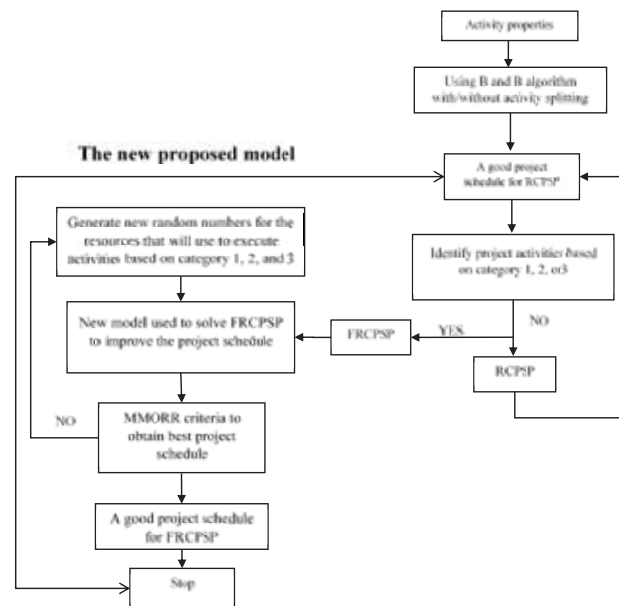


Figure 6. Flowchart for the new proposed model

According to the second case (the problem addressed in this paper): The work required to execute (FRP) activities has been referred as work-content ω , so the highest and the lowest quantity of resource required \bar{r} and \underline{r} should be identified in advance as a domain for resources required. Scheduler can identify the shortest duration \underline{d} to execute the activity based on the highest quantity of resource required \bar{r} as the first step. The Earliest start times (E_s) and the shortest durations \underline{d} for each project activity in category C is determined by dividing work-content on the highest quantity of resource required:

$$\underline{d} = \omega / \bar{R} \quad (1)$$

The longest activity duration can be identifying when the resource required $R=1$, which means $\bar{d}=\omega$.

Next step is: using Gantt chart, ignoring the resource availability, and sorting project activities based on the relationships among them. After that, compare for each unit- time the resource required with the resource available and follow the next steps:

1. In case the summation of the resources required at any unit time less than the resource available, assign the resource available to complete these activities.
2. Assume at any unit-time, there is resource conflict is occurred between the activities in categories A, B, and C:
 - a. Firstly, start by assigning the resource available on the activities in category A. Calculate the resources remaining by using

$$R_r = R_{ava} - R_A \quad (2)$$

- b. Secondly, assign the resource available R_r on activities in category B, and the rest can be assigned to cover a portion of the activity in category C.
 - c. In case the summation of the resource required on activities in category A&B is more than the resource available, try to allocate the resource available on activities in category A&C, taking into the account the domain of the resource required allowed to complete the activity.

$$R_C \neq R_{ava} - R_A \quad (3)$$

It is worth mentioned in this paper that the model proposed consists of an algorithm that generates a huge number of execution modes based on the assumptions for the category C. For example, activity in category C has the work-content $\omega = 3$, and the highest and the lowest values of the resource required $\bar{R}=3$ and $\underline{R}=1$ respectively, and the activity could be executed during the period of 7 days because there are 4 days as slackdays. The number of the feasible solutions have been obtained equals 56 i.e. the activity can be executed by 56 ways as indicated below:

```
000003 000012 000021 000030 000102 000111 000120 000201 000210 000300 001002 001011
001020 001101 001110 001200 002001 002010 002100 003000 010002 010011 010020 010101
010110 010200 011001 011010 011100 012000 020001 020010 020100 021000 030000 100002
100011 100020 100101 100110 100200 101001 101010 101100 102000 110001 110010 110100
111000 120000 200001 200010 200100 201000 210000 300000
```

You can imagine the huge number of the project feasible solutions that could be created based on the number of project activities in category C, where each activity has different factors from others.

5. NUMERICAL EXAMPLE

To better understand the proposed model, let us consider the following example presented in 2003 ^[5], which will be used throughout this paper to illustrate the basic ideas and modeling frameworks.

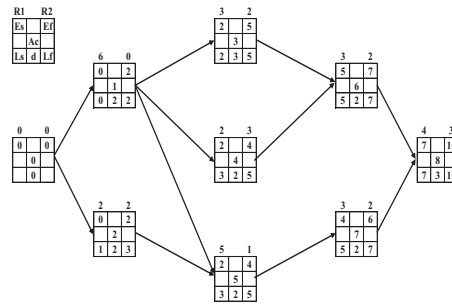


Figure 7a. The project network for the given example

The example of a project network is given to present project activities on the node (AON) as depicted in (Figure 7). The project is constrained by precedence relationships and resource constraints. From the perspective of the planners, the resources are usually allocated without considering competition for the resources of other activities. Therefore, the generated process plans are somewhat unrealistic and cannot be easily executed for a group of activities. Accordingly, the resulting optimal process plans often become non-feasible when they are carried out in practice at the later stage. In this example, the project completion time is 10 days, where resource requirements R_1 & R_2 are 6 & 15 respectively. Figure 8 shows the non-feasible project schedule because of violations exist in the resources.

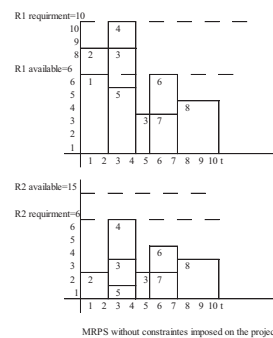


Figure 8. Non-feasible project schedule

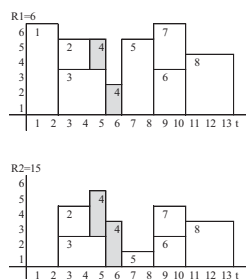


Figure 9a

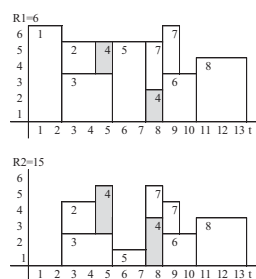


Figure 9b

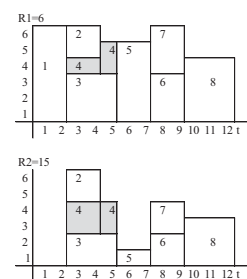


Figure 9c

When all project activities are classified as category A, i.e. project activities are non-preemptive with constant resource profile CRF, then the resource required to complete the activity 4 is 2 units from R1 and 3 units from R2 in 2 consecutive days as depicted in (9a), the project needs 13 days for completion. When some of

project activities are classified as Category A (activities 1, 2, 3, 5, 6, 7, and 8) and some others as category B (activity 4) with constant resource profile CRP, then the resource required to complete activity 2 is 2 units from R1 and 2 units from R2 in 2 separately days as depicted in (9b), the project needs 13 days for completion. While, if the project activities are classified as Category A (1, 2, 6, 8), category B (3, 5, 7), and category C (4) with flexible resource profile FRP, then activity 2 needs 2 units from R1 and 2 units from R2 in 2 consecutive days, and activity 4 needs 1 units from R1 in the first 2 days and 2 units in the third day and needs 2 units from R2 in 3 consecutive days as depicted in (9c), the project duration is reduced to 12 days. In worth mentioned, there is a high number of feasible solutions can be obtained by using the new model proposed to obtain the same results $t=12$ days as is in the third type; however, the best solution can be determined among them by applying the MMORR criterion.

6. CONCLUSIONS

As demonstrated in this paper, it is worth noting that using production schedules of PM.s in the OKP.s and classifying the activities into three categories, seems very affectively way to reduce the project duration, as well as project resource utilization in the OKP.s.

REFERENCE

- [1] Slowinski, R. (1980). Two approaches to problems of resource allocation among project activities--a comparative study. *Journal of the Operational Research Society*, 711-723.
- [2] Talbot, F. B. (1982). Resource-constrained project scheduling with time-resource tradeoffs: The non-preemptive case. *Management science*, 28(10), 1197-1210.
- [3] Patterson, J. H., Slowinski, R., Talbot, F. B., & Weglarz, J. (1989). An algorithm for a general class of precedence and resource constrained scheduling problems. *Advances in project scheduling*, 187, 3-28.
- [4] Sprecher, A., & Drexel, A. (1998). Multi-mode resource-constrained project scheduling by a simple, general and powerful sequencing algorithm. *European Journal of Operational Research*, 107(2), 431-450.
- [5] Alcaraz, J., Maroto, C., & Ruiz, R. (2003). Solving the multi-mode resource-constrained project scheduling problem with genetic algorithms. *Journal of the Operational Research Society*, 54(6), 614-626.
- [6] Tritschler, M., Naber, A., & Kolisch, R. (2017). A hybrid metaheuristic for resource-constrained project scheduling with flexible resource profiles. *European Journal of Operational Research*, 262(1), 262-273.
- [7] Van Peteghem, V., & Vanhoucke, M. (2010). A genetic algorithm for the preemptive and non-preemptive multi-mode resource-constrained project scheduling problem. *European Journal of Operational Research*, 201(2), 409-418.
- [8] Kellenbrink, C., & Helber, S. (2015). Scheduling resource-constrained projects with a flexible project structure. *European Journal of Operational Research*, 246(2), 379-391.
- [9] Issa, S., & Tu, Y. (2017). Integrated multi-resource planning and scheduling in engineering project. *Journal of Project Management*, 2(1), 11-26.
- [10] Van Peteghem, V., & Vanhoucke, M. (2014). An experimental investigation of metaheuristics for the multi-mode resource-constrained project scheduling problem on new dataset instances. *European Journal of Operational Research*, 235(1), 62-72.

- [11] Fündeling, C. U., & Trautmann, N. (2010). A priority-rule method for project scheduling with work-content constraints. *European Journal of Operational Research*, 203(3), 568-574.
- [12] Naber, A. (2017). Resource-constrained project scheduling with flexible resource profiles in continuous time. *Computers & Operations Research*, 84, 33-45.
- [13] Kolisch, R., & Meyer, K. (2006). Selection and scheduling of pharmaceutical research projects. *Perspectives in Modern Project Scheduling*, 321-344.
- [14] Naber, A. (2017). Resource-constrained project scheduling with flexible resource profiles in continuous time. *Computers & Operations Research*, 84, 33-45.
- [15] Naber, A., & Kolisch, R. (2014). MIP models for resource-constrained project scheduling with flexible resource profiles. *European Journal of Operational Research*, 239(2), 335-348.
- [16] Baumann, P., & Trautmann, N. (2013, December). Optimal scheduling of work-content-constrained projects. In *Industrial Engineering and Engineering Management (IEEM), 2013 IEEE International Conference on* (pp. 395-399). IEEE.
- [17] Ranjbar, M., & Kianfar, F. (2010). Resource-constrained project scheduling problem with flexible work profiles: a genetic algorithm approach. *Scientia Iranica. Transaction E, Industrial Engineering*, 17(1), 25.
- [18] Zimmermann, A. (2016, December). An MIP-based heuristic for scheduling projects with work-content constraints. In *Industrial Engineering and Engineering Management (IEEM), 2016 IEEE International Conference on* (pp. 1195-1199). IEEE.
- [19] Correia, I., Lourenço, L. L., & Saldanha-da-Gama, F. (2012). Project scheduling with flexible resources: formulation and inequalities. *OR spectrum*, 34(3), 635-663.
- [20] Kolisch, R., & Drexel, A. (1997). Local search for non-preemptive multi-mode resource-constrained project scheduling. *IIE transactions*, 29(11), 987-999.
- [21] Tu, Y., & Dean, P. (2011). *One-of-a-kind Production*. Springer Science & Business Media.

An Analysis on the Development Model of China's County-level E-commerce

Kegui Liao¹, Yuzhen Zhang², Xuewen Gui^{3}*

¹Marketing Department, Hubei Science and Technology College, Wuhan, 430072, China

²Department of International Economics & Trade, Wuchang University of Science and Technology,
Wuhan, 430223, China

³E-commerce Department, Wuchang University of Science and Technology, Wuhan, 430223, China

Abstract: Based on the case studies of e-commerce activities in six counties of China's eastern, middle and western regions respectively, this paper has probed into the characteristics of e-commerce development of each county, which are then classified into four development models of county-level e-commerce in China, featuring the integration and aggregation of resources endowment and production factors. The paper further analyzes the key factors contributing to the success of county-level e-commerce development, in a bid to provide reference and guidance for other counties in their e-commerce activities.

Keywords: County-level e-commerce, Development model, Key factors, Factors analysis

1. INTRODUCTION

China's county-level regions serve not only as the link between urban and rural areas, but also an important carrier promoting the development of China's economy and new urbanization. According to the report on rural e-commerce development in China (2016-2017), online retail sales of rural areas in the first half of 2017 amounted to 84.93 billion U.S. dollars, up 38.1% over the same period last year. By the end of 2016, there had been 8.32 million online stores in the rural areas and 200 e-commerce industrial parks for all kinds of agricultural products had been built. Nationwide, the number of Taobao Villages had reached 1311 by the end of August 2016. The coastal provinces of Zhejiang, Guangdong and Jiangsu accounted for the biggest number of Taobao Villages. ^[4]

Since China's county-level e-commerce activities are mainly concentrated in the economic zones including Yangtze River Delta (YRD), Pearl River Delta (PRD) and Bohai Economic Rim (BER), six counties in China's eastern, middle and western areas were selected for case study, based on representativeness, diversity and significance, and a comparative analysis was conducted on rural e-commerce model, development link and key factors.

2. A BRIEF INTRODUCTION TO E-COMMERCE DEVELOPMENT IN CHINA'S SIX COUNTIES

2.1 Model 1: Farm produce-driven e-commerce model; Typical cases: Suichang County and Tongyu County

Under this model, local farmers start e-commerce business with the support from local government departments as well as the assistance of local e-commerce association, network marketing platforms and e-business service operators. As a result, an ecosystem of county-level e-commerce is taking shape with the win-win cooperation among three parties, namely the network operators, suppliers and the third-party service providers. ^[6]

Being a typical mountainous county with mountain area accounting for 88% of the land area, Suichang County in China's eastern Zhejiang Province is underdeveloped in terms of industrial economy but abundant in agricultural resources and forestry products. With the support from local government, Suichang E-commerce Association has actively explored a model of selling its agricultural and forestry specialty such as bamboo charcoal, roasted potatoes and chrysanthemum: producers + internet service providers + network distribution platforms (with e-commerce businesses settled in). The core part of this model is the e-commerce service station, a nonprofit service station established by Suichang E-commerce Association. ^[16] The important roles this

e-commerce service station plays in the online sales of agricultural products include the following: first, to lay down a criterion for agricultural products and introduce the criterion into practice; second, to encourage farmers and processing enterprises to process and produce according to this criterion; third, to set up a display hall and make a data packet for online stores; fourth, to unify the procedures of storage, delivery, after-sales service and inventory management; fifth, to promote collaboration among each sector; sixth, to oversee the construction of the project known as 51ganjie (which means “I want to bring my products to market” in Mandarin, and whose full name is Rural E-commerce Service Station, <http://info.51ganjie.com>). (E-commerce structure shown as figure 1)

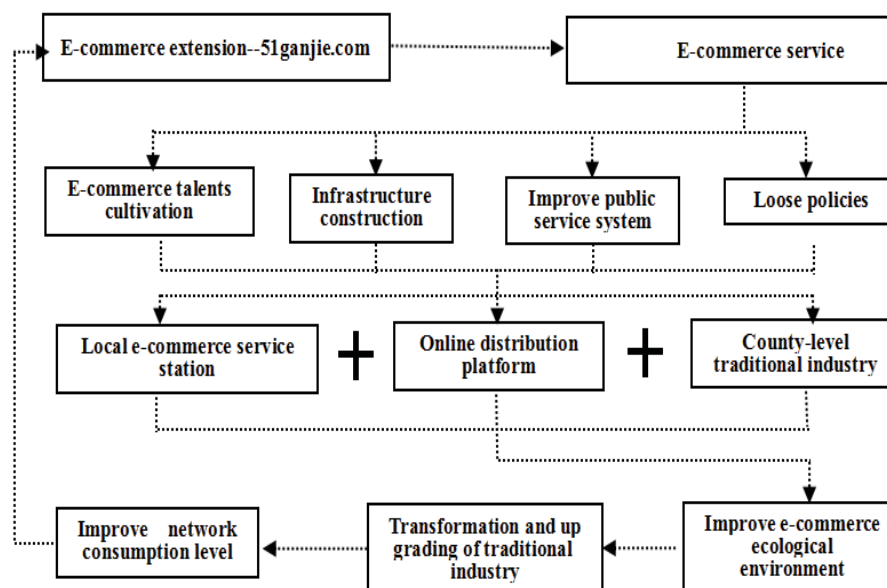


Figure 1(left) . The structure of e-commerce model of Suichang County, Zhejiang Province

The remote Tongyu County of China's north-eastern Jilin Province is one of the national-level impoverished counties, yet it ranks the first nationwide in terms of yields of green beans, hence the name of “homeland for grains and beans”. E-commerce in Tongyu County is driven by top management commitment, with specialized agencies and funds established and preferential policies offered for e-commerce business start-ups. A county-level e-commerce development center, the first of its kind in the province, has been founded. What's more, an e-commerce company named “Yunfeihewu E-commerce Company Limited” has also come into being invested by social funds and backed up by the local government. Therefore, Tongyu County's e-commerce model can be summarized as : Farmers (Producers)+ Yunfeihewu E-commerce Company Limited + Online Merchants.^[5] Backed up and supported by the local government and production base, the agricultural products sold online are directly sourced from their places of origin with uniform brand through innovative marketing.

2.2 Model 2: Interactive development of rural processing industry and e-commerce; Typical cases: Shaji County and Qinghe County

The feature of this model is to build regional processing product brands either based on existing rural processing bases or inspired by the originality of local successful online merchants. B2B and B2C e-commerce are therefore conducted through e-commerce platforms, and a relatively complete comprehensive county-level e-commerce service system is established as a result.^[5]

Shaji Town, Suining County in China's eastern Jiangsu province has a rather weak agricultural basis with a cultivated land area per capita of less than 1 mu (about 0.0666 hectares). However, e-commerce in this town has developed quickly through online sales of assembled furniture. The model therefore is: assembled furniture factories + online merchants.^[1] Shaji Town is taking a bottom-up development path, which advances step by step. Since Mr. Sun Han, one of the so-called Three Musketeers of Online Merchants, first opened an online store in 2006, more and more villagers have followed him to start their own businesses on internet, just like the

cell clone. Almost during the same period, processing facilities, logistics and other infrastructure services have also emerged and as a result an e-commerce industry chain has taken shape gradually, with costs of each link decreased. (E-commerce structure shown as figure 2)

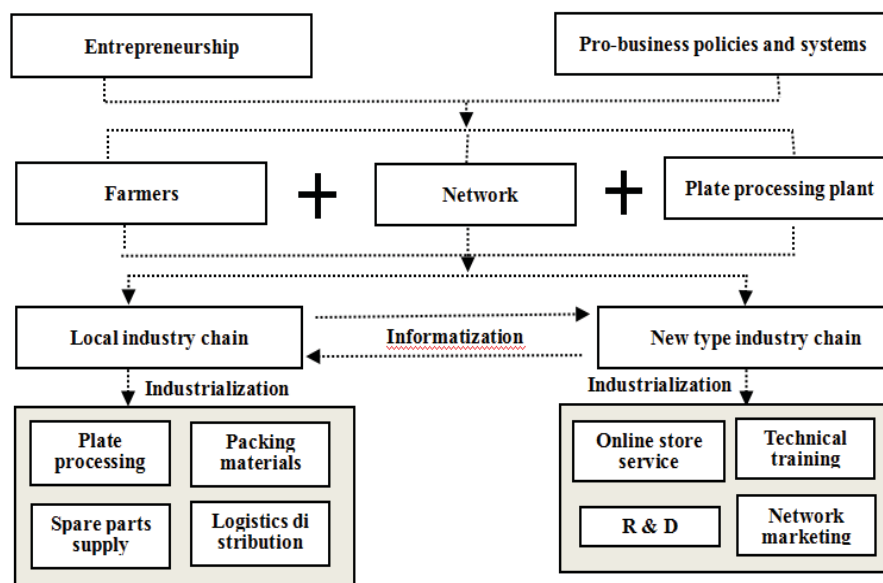


Figure 2 (right) . E-commerce structure of Shaji Town, Suining County, Jiangsu Province

Qinghe County in north China's Hebei Province, being well-known as the capital of cashmere and China's largest processing base for cashmere, has a very strong industrial basis. However, in 2008, Qinghe's export sales for cashmere greatly declined due to the global financial crisis. The county government then began to vigorously develop e-commerce, putting forward the development thought of "to implement interactions from online to offline and to realize the mutual complementation between tangible market and intangible market".^[3] Therefore Qinghe County's model is: cashmere processing factories + online merchants, with the following characteristics: firstly, various measures have been adopted by local government in order to create a good environment for e-commerce business; secondly, strong support from traditional industry; thirdly, leading roles are played by online merchants. By the end of 2017, the county's online sales volume for Qinghe cashmere amounted to 550 million U.S. dollars, with deep processing products accounting for 65% of the whole industry.

2.3 Model 3: Mutual development of transportation and e-commerce; Typical case: Wugong County.

The core of this model is, taking advantage of the geographical location and preferential policies, to attract neighboring agricultural products to the local e-commerce industrial park in order to build a "one-stop two-way integrated e-commerce system", a two-way circulation channel for both industrial products and agricultural products.^[10]

Situated in the west of Guanzhong Plain and as an important passage to the east for Xinjiang Uygur Autonomous Region, Qinghai Province and Gansu Province, Wugong County in northwest of China's Shaanxi Province is an important transportation hub and distributing center. Taking advantage of Wugong's unique geographic location, the county government issued in March 2014 its *Opinions on Encouraging and Supporting Development of Electronic Commerce*, later invested 2 billion RMB in May 2015 to build an e-commerce industrial park covering an area of 650 mu (about 43.33 hectares). According to *White Paper on Agricultural E-commerce (2016)* issued by Ali Research Center, Wugong County ranked the first in the northwestern area and the fifth nationwide among China's top 50 agricultural products e-commerce markets. Therefore, Wugong County's model can be summarized as: distributing center + e-commerce,^[6] with characteristics as follows: firstly, to give full play to the advantages of transportation, warehousing and logistics; secondly, to attract e-business enterprises to settle down in the e-commerce industrial park; lastly, to integrate resources in the northwestern area for sale across the whole country.

2.4 Model 4: Transition to e-commerce: leveraging land-transfer. Typical case: Xiji County

This is a new type eco-friendly leisure & tourism model based on online customized private ecological farms, relying on local ecological resources and tourism resources to expand agricultural functions, promote agricultural transformation and industrial upgrading.

Jixi County of China's central Anhui Province has a representative project called Jutudi, an online personal customized farm project launched by the e-commerce giant Alibaba Group. Farmers participating in this project transfer their land-use rights to Zhejiang Xinhe E-commerce Co., Ltd, an enterprise directly under Zhejiang Supply and Marketing Cooperative, then the e-commerce company entrusts the production and management of land to the local farmer's cooperative. Customers participating in the project can subscribe the right to use the land via the online platform www.juhuasuan.com, a leading Alibaba-owned group-buying e-commerce platform, and get the land output. The local cooperative hires farmers to work in the subscribed land, therefore the farmers can get wages in addition to an annual circulation payment.^[7]

Jixi County was chosen as one of pilot demonstration counties to carry out online farming projects. Because of its dispersed farmland, the sales volume of agricultural products in Jixi County is rather limited due to lack of scale effect. The online farming project allows the crops grown to be ordered in advance by customers via e-commerce platform, and the unattended land to be circulated to the local cooperative for management. It has combined very well both the demand of urban dwellers and rural farmers. On the one hand, the organic products from rural area can get into the urban area, and on the other hand, urban dwellers can feed back the farmers through rural tourism. Since the project was launched, 465 mu (about 31 hectares) of land have been subscribed by customers nationwide, the sales volume amounted to 2.38 million yuan (\$320,000). Farmers participating in the project had an increase in income up to 9.6 times of that before (e-commerce structure is shown as figure 3).

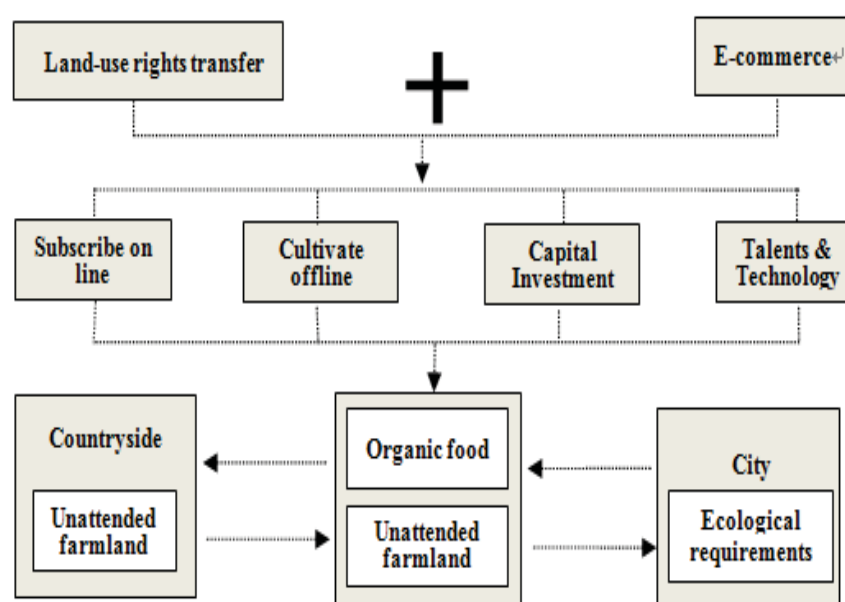


Figure 3 (left). E-commerce structure of Jutudi Project

3. COMPARATIVE ANALYSIS ON KEY FACTORS OF E-COMMERCE DEVELOPMENT IN THE ABOVE SEVEN COUNTIES

Studying the above four development models, we found that five key factors are absolutely necessary which contributed to the initial success of county-level e-commerce in China, namely government support, resources, talents, service and marketing.^[8] (The five key factors and their functions are shown in figure 4). These five key factors were linked together and supported one another, ensuring the smooth development of e-commerce in the counties.

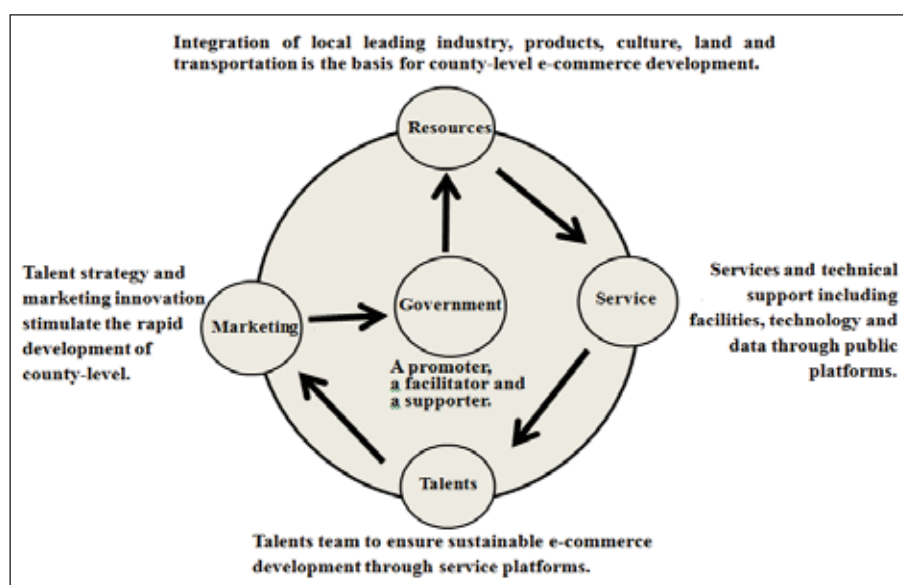


Figure 4 (right) . Key factors for county-level E-commerce development

3.1 Government support

The local government departments have adopted a series of measures to support rural e-commerce, including: 1) setting up management organizations for e-commerce development; 2) formulating preferential policies and regulations on e-commerce; 3) investing in e-commerce business; and 4) updating service concept.

[3]

Table 1. Measures taken in the six counties to support e-commerce

County	Measures taken to support e-commerce
Suichang	Implementation Opinions of Suichang County Party Committee on Accelerating the Development of Rural Electronic Commerce.
	Strategic Plan for Electronic Commerce Development in Suichang County (2014-2020).
	Inclusive Finance: Measures for the Administration of E-commerce Loan: Credit Loans through Dian Shang Tong (E-commerce Access).
	Defining the government role: inductor, supporter and service provider.
Tongyu	Implementing Plan on Comprehensive Demonstration Work of Promoting E-commerce Entering into Rural Areas.
	Setting up government guiding fund (US\$ 2.05 million).
	Establishing incentive mechanism and setting green channels.
Wugong	Opinions on Encouraging and Supporting the Development of Electronic Commerce.
	Setting up a RMB 5 million special fund to support e-commerce development.
	Investing US\$316 million to build an e-commerce industrial park.
Quning	Implementation Opinions of Quning County Party Committee on Accelerating the Development of E-commerce to Achieve Leapfrog Development.
	Providing guarantee of no less than US\$9.48 million per year.
	Providing loan of more than US\$15.80 million per year.
	Investing and building Shaji e-business incubator.
	Defining the government role: neither 'vacancy' nor 'offside'.
Qinghe	Putting forward the development thought "to implement interactions from online to offline and to realize the mutual complementation between tangible market and intangible market".
	Formulating incentive policies to promote industrial transformation and upgrading.
	Setting up quality supervision and inspection institutions for cashmere products.
	Hiring professional design institutions and senior designers.
Jixi	Opinions on Accelerating the Development of Electronic Commerce.
(Jutudi project)	Investing US\$47,000 as a special fund for the repairs and maintenance of tractor-ploughing road and irrigation canals.
	Local government and e-commerce company jointly entrust an agricultural cooperative to supervise and organize production.

(Data source: website of local governments, official website and other publicity materials; Date: by the end of 2017)

From table 1, we can see that the local governments in these six counties have set up official management organizations respectively for e-commerce development, taken a series of measures such as 'policies and regulations + capital input + project implementation', and continuously innovated management and service concept. Compared with other counties, Quning County Government and Qinghe County Government are

playing a significant role in the promotion of industrial upgrading and transformation, offering the so-called ‘postpositional service’.

3.2 Resource endowment

Resource endowment mainly includes the agricultural and forestry specialty resources and other resources such as industry, land, tourism, traffic and entrepreneurship, ^[4] which serve as the basis for e-commerce development in rural county areas.

Table 2. Resource endowment for E-commerce development in the six counties

County	Superior resources (agricultural & forestry specialty, industry, transportation, tourism, etc.)
Suichang	Rich in agricultural & forestry specialty resources: bamboo shoots, bamboo charcoal products, baked sweet potatoes, chrysanthemums, etc. With the influence of ‘A bite of China’, a popular documentary featuring delicious Chinese food, make best use of historical and cultural resources as well as ecological environment resources.
Tongyu	A well-known homeland for coarse cereals and beans.
Wugong	An important transport hub and a distribution center. Well developed in the field of logistics and cold-chain. An obvious advantage in terms of road, railway and infrastructure.
Suining	Abundant Italian poplars resources. Adjacent to Pizhou - one of China’s four plywood processing bases. Popular online products: assembled furniture.
Qinghe	China’s largest deep processing base distribution center for cashmere products.
Jixi (Jutudi project)	Good ecological environment and transferable land-use rights. Was named one of China’s top 100 counties in terms of travel and tourism competitiveness (2013).

From table 2, we can see that among the six counties selected, five counties have fully tapped their own endowment of natural resources and taken it as the basis for developing e-commerce, namely, the local agricultural and forestry specialty resources, industries with local advantages, land resources and tourism resources, and convenient transport location, etc. However, Shaji Town of Quning County is taking a totally different way without relying on its traditional industry at all. Being innovative and by selling its neighboring county’s plates, Shaji Town has blazed a new trail of developing its economy through informatization.

3.3 The comprehensive service network

The comprehensive service network mainly consists of e-commerce association, supporting infrastructure and related service facilities including public service stations and logistics system, information service and technical support. ^[6] At present, the e-commerce industrial parks which have been built or are being built in these counties, can perform general functions and provide services including brand promotion, logistics services, personnel training, technical support and quality & safety management.

Table 3. Comprehensive service network in the above six counties

County	Comprehensive service network.
Suichang	An incubator and startup base for rural e-business is under planning based on the existing Bamboo Charcoal Industrial Park.
Tongyu	County-level logistics nodes have been set up; The only e-commerce service center in Jilin Province has been set up.
Wugong	A logistics park with an e-business incubator is under construction.
Suining	An e-commerce pioneer park has been set up.
Qinghe	A three-tier logistics system connecting county, town and village has been built. A logistics park is under construction;
Jixi (Jutudi project)	An e-commerce industry cluster district is under construction. Cold-chain logistics support system for agricultural products is not available for the moment.

From table 3, all the six counties have set up e-commerce service centers and can provide one-stop integrated services including e-commerce service (logistics and delivery service, thirdpartnar (TP), e-commerce training, exhibition, financial and human resources, etc.), the government affairs service and life service; five have set up county-level e-commerce associations and public service platform by means of establishing e-commerce industrial park or logistics park. In addition to Wugong County in the west and Tongyu County in the northeast of China, the other 4 counties can provide good information services and technical support, which reflects a deep integration of information technology and industrialization.

However, due to lack of cold-chain logistics support system for agricultural products in Jixi County where the Jutudi project is implemented, part of leaf vegetables cannot be delivered to the customers within the preservation time limit.

3.4 Talent team

The e-commerce industrial chain is a systematic project involving goods, technique, logistics and customer service, therefore it needs not only technical and business professionals, but also inter-disciplinary talents such as strategic management personnel. The talent team construction mainly includes the establishment of online and offline talent team training system so as to cultivate e-commerce professionals. ^[5]

Table 4. E-commerce Talent Team in the Above Six Counties

County	E-commerce talent team
Suichang	A talent team of about 500 e-commerce professionals.
	Suichang County Agricultural E-commerce Institute.
	Suichang Ganjie Vocational Training School.
	Suichang Ganjie Vocational Training School.
	More than 6000 e-commerce practitioners.
	Holding training classes aiming at improving e-commerce skills for local enterprises.
Tongyu	E-commerce training classes at elementary and intermediate level were held with annual trainees of over 1000.
	E-commerce training base set up by the local human resources department.
	Online public training platform: Tongyu Pavillion.
	Rural Taobao Farmer e-businessmen Training Center.
Wugong	Implementation Plan of Tongyu County for E-commerce Talent Cultivation and Training.
	Training staff of more than 20,000.
	Shaanxi E-commerce Training Base.
	Taobao University, Shaanxi Branch.
Suining	Special lectures are given and practical training classes are held.
	More than 2800 local residents have received training on e-business by the end of 2017.
	Number of Taobao Towns: 5, number of Taobao Villages: 40.
	Number of online stores: over 2000, number of TP service providers: over 40.
Qinghe	Program set up on training 10,000 rural e-commerce practitioners in the Year 2014.
	More than 12,000 local residents have received training.
	Was named national e-commerce demonstration base.
	Number of Taobao Towns: 1, number of Taobao Villages: 8.
Jixi (Jutudi project)	Over 60,000 e-business practitioners.
	More than 30,000 local residents attended Taobao training classes.
	Partner of the project: Alibaba Group
	Contractors: comprehensive e-commerce operator and service provider.
	<i>Several Opinions on Accelerating the Development of E-commerce</i> issued by the local government.
	Over 200 local residents received training in 2016.
	The project is entitled to take on e-business personnel from its parent company Alibaba Group's talent pool.

Table 4 shows that among the six counties selected, three counties (Suichang, Suining and Qinghe), have a rather good basis for e-commerce development where localized e-commerce talents are increasing and Taobao Villages and Towns are emerging. Jointly launched by e-commerce giant Alibaba Group Holding Ltd and Zhejiang Xinghe E-commerce Co., Ltd, Jutudi project in Zhejiang's Jixi County has a rich talent pool while the other two counties are stepping up efforts to train e-commerce personnel. At present in China's county-level areas, most of the online merchants are doing jobs requiring only basic skills, due to lack of such talents as R&D staff, business professionals, and strategic management personnel.

3.5 Marketing model

The marketing model mainly includes the selection of e-commerce model, brand strategy, product standardization and marketing innovation strategy, etc. ^[1]

Table 5. E-commerce marketing model of the six counties

County	Innovative Marketing Strategies
Suichang	County government departments create and abide by a complete set of strict and scientific quality management system and professional delivery and inventory management criterion as well as optimized work-flow in terms of storage and delivery. Ganjie Project: The aim of the project aims is to integrate e-commerce, local life and rural entrepreneurship by providing one-stop comprehensive service for rural residents in multiple cities and counties.
Tongyu	Government support: Government authorities encourage the local residents to engage in e-commerce Base support: a project has been planned and started to set up a direct production base for cooperative e-commerce enterprises, covering 10,000 mu of land to cultivate various kinds of agricultural products Unified brands: 2 brand names "Sanqianhe " and "Dayounian " have been registered. Marketing innovation: An e-commerce company ("Yunfeihewu E-commerce Co. Ltd) was established, aiming to sell online all the agricultural products (including millet, green beans, oats and beans) it has collected from local farmers, production bases, cooperatives or processing enterprises, etc.
Wugong	Taking advantages of its convenient geographic location and logistic facilities, an e-commerce park was set up in order to integrate resources from northwest of China for online sales by encouraging e-commerce enterprises from other places.
Suining	The county has explored a path characterized by industrial division of labor and socialized cooperation. Based on the successful experiences of local online merchants, the furniture industry has made a so called "three-level" achievement, namely from " ready-to-assemble furniture ", then "wood and steel furniture ", and to "customized furniture ". This has stimulated the development of related industries including manufacture, logistics distribution and delivery, raw materials processing, and spare parts supply, thus a new industry eco-system facing a larger market is taking shape.
Qinghe	Industrial basis: well-established traditional industry and specialized market. Demonstration effect: The local fine examples' demonstration expedited the transformation of local enterprises from traditional business model to platform-based e-commerce business model. Marketing advantages: favorable prices and strong competitiveness.
Jixi (Jutudi Project)	Under the Jutudi Project, land-use rights are transferred through online platform, valid for one year. Customers subscribing the right to use the land can either take care of the land by themselves or let local farmers work for them free of charge. All the agricultural yields within the subscribed year will be sent to the customers twice a month. In addition, customers are usually offered some preferential treatment such as free tickets to the local scenic spot and free accommodation during their farm stay.

Generally speaking, county-level e-commerce model falls into three categories, namely B2B, B2C and C2C. It can be seen from table 5 that in the selected counties, B2C e-commerce is the dominant model while B2B and C2C are only supplementary. All the six counties are implementing the brand strategy; however, due to lack of well-known brands, it is difficult to form strong brand competitiveness. National product standards are generally adopted to guarantee the quality of products sold online. In terms of marketing strategy, Suichang County plays a vanguard role by carrying out the so-called "Ganjie project" under the guidelines of its new concept for developing rural e-commerce; Meanwhile, Suining County has explored a path characterized by industrial division of labor and socialized cooperation based on its success in selling furniture online.^[11] For Qinghe County, besides selling its cashmere products on internet platforms such as Taobao.com and Tmall.com (both owned by internet giant Alibaba), it has also carried out other kinds of online marketing activities at the same time, including establishing a trade center ("Xinbaifeng Cashmere Electronic Trading Center") , a B2C website (<http://www.qingheyr.com>), and a O2O website "Baironghui" to sell its cashmere products; In order to sell its coarse cereals and mixed beans, Tongyu County has innovated its marketing strategy including carrying out uniform brand strategy which has got support from the local government as well as its industrial base;^[13] The marketing strategy conducted Wugong County is a combination of government marketing plus hot sale items strategy.^[14] As for Jixi County where the Jutudi project is carried out, it has explored a new model of taking advantages of land resources through online customized farm project with crowd funding.^[16] On the whole, the marketing innovation in these selected counties is critical to the vitality of rural e-commerce development, though further improvement still needed.

4. REVELATIONS FROM CHINA'S COUNTY-LEVEL E-COMMERCE DEVELOPMENT

One county's e-commerce development mode and its future direction depend to a large extent on whether the county Party Committee and county government have a good understanding of e-commerce and accurate positioning of the government role, and how they will push forward e-commerce development in rural areas.^[9]

Local farm produce, industrial basis and entrepreneurship are important factors contributing to the internal comparative advantage whereas high quality supporting service facilities and preferential measures play important roles in attracting capital and expertise from outside.^[15] In addition, in order to change people's outdated marketing concepts, implement the brand strategy and innovate the existing marketing mode, it is necessary to introduce and cultivate talents at all levels in the rural counties, so as to expand and extend the industrial chain, and eventually play a leading role in the integration of primary industry, secondary industry and

tertiary industry.

Firstly, introduce brain-trust to have a deep study of the key issues regarding the county-level e-commerce development, including the strengths, weaknesses, objectives and supporting policies etc, and to have an overall plan and top-design of the county's e-commerce development.^[10] The government services and market efficiency should be combined effectively under the principle of "government investment, enterprise operation, public welfare oriented, and market as a supplement". Taking e-commerce development as its top management commitment, local government shall take targeted measures to support e-commerce development in terms of policies, fund and projects.

Secondly, choose geographical indication products, specialty products, and special resources products or manufactured products with strong industrial base. In case of lack of enough resources, all the manpower and physical resources of the whole county shall be gathered together in order to build hit products. Please note that one county's e-commerce is not necessarily confined to the resources within the county. If the county enjoys a favorable geographic location with convenient transportation facilities, for example, it then can be built into a comprehensive regional e-commerce center gathering all kinds of e-commerce factors; in this way its e-commerce may develop even faster with a better development prospect.

Thirdly, enhance the construction of infrastructure facilities, e-commerce industrial park, logistics industrial park and storage center, etc so as to provide comprehensive one-stop service including e-commerce service, administrative service and daily life service, etc.^[14] We shall build a comprehensive e-commerce service system which not only integrates the functions of subject cultivation, hatch support, platform construction and marketing promotion, but also undertakes transformation of various resources and the demand from "government, internet enterprises, and suppliers".

Furthermore, attach great importance to talented personnel cultivation and take it as priority in e-commerce development^[16]. Our general idea is to train our own personnel while introducing talents from other places. That is, to train technical talents and introduce professional personnel at the same time, including R&D staff, business professionals and strategic and project management personnel. On the one hand, to introduce leading figures from other places, on the other hand, to encourage local leading figures to play exemplary roles in promoting e-commerce development; And lastly, training effects should be taken into account when it comes to talents team.

Finally, encourage innovation in marketing mode, especially innovation in the rural e-commerce, which plays a pioneering role in the county-level e-commerce development. We shall deeply develop the products and carry out brand strategy.^[11] Online and offline business shall be combined in order to realize overall marketing. More creative marketing theme activities shall be planned in order to maintain market exposure. It's necessary to set up an e-commerce association as well as a quality control mechanism which is responsible by multi-parties including government departments, local farmers, cooperatives, e-commerce associations and platforms.

5. CONCLUSIONS

County-level e-commerce development is a systematic project, which means administrative measures or individual enterprises alone have limited roles, instead, it must rely on the effective interaction among service providers, e-commerce merchants and enterprises, traditional industries and local government, so as to build a new type of e-commerce ecology and boost e-commerce development.

REFERENCES

- [1] Peter F·Drucker .(2009).The effective Executive. XU Shixiang(translator).Beijing: Machinery Industry Press, 155 (in Chinese).
- [2] Michael A. Crew, Timothy J. Brennan. (2015). Postal and Delivery Innovation in the Digital Economy [M]. Springer International Publishing.323-325.
- [3] Hubei E-Commerce Research Center of Huazhong Normal University.(2016). The 13th Five-Year Plan for Electronic Commerce Development in Hubei Province (2016-2020). Hubei E-Commerce paper No.24-156.Department of Commerce of Hubei Province.Wuhan: Hubei People's Press (in Chinese).

- [4] Alibaba Group Research Centre.(2016).Research Report on China County-level E-commerce Development .http://www.askci.com/news/hlw/20160606/16155425768_2.shtml (in Chinese).
- [5] YU Han , ZHANG Chang.(2016).A Research into the Model of Rural E-commerce Development in Hebei Province: Commercial Times,(12):172-174 (in Chinese).
- [6] HONG Tao.(2016).On the Development of China's Agricultural Products E-commerce in 2015 and Outlook in 2016:.Commercial Times, (11):63-66 (in Chinese).
- [7] WANG Fengliang.(2016).To Expand the Development Path of Rural E-commerce under the Stage of New Norma:Economic Review, (7):35-36 (in Chinese).
- [8] CP Kang , XD Wang.(2013).Present Status Quo of Shaji E-commerce and the Studying of 'Shaji Model 2.0':Journal of Xuzhou Institute of Technology,28(3):14-17 (in Chinese).
- [9] Borsenberger C.. (2015).The Concentration Phenomenon in E-commerce. Postal and Delivery Innovation in the Digital Economy. Springer International Publishing. 31-41.
- [10] HUANG Ruo.(2013). E - commerce In My Eyes. Beijing: Electronic Industry Press, 135 (in Chinese).
- [11] Yan Q., Wu S., Wang L., Wu P., Chen H., Wei G. . (2016). E-wom from e-commerce websites and social media: which will consumers adopt?. Electronic Commerce Research & Applications, 17(C): 62-73.
- [12] Yan Q., Wang, L., Chen W., Cho J. .(2016). Study on the influencing factors of unplanned consumption in a large online promotion activity. Electronic Commerce Research,16(4): 453-477.
- [13] Leong C., Pan S. L., Newell S., Cui L. .(2016). The emergence of self-organizing e-commerce ecosystems in remote villages of china: a tale of digital empowerment for rural development. Mis Quarterly,40(2):475-484.
- [14] Zhu S., Chen J. .(2016). E-commerce use in urbanising china: the role of normative social influence. Behaviour & Information Technology,35(5): 357-367.
- [15] Uzoka F. M. E., Seleka, G. G. (2006). B2C e-commerce development in Africa: case study of Botswana.ACM Conference on Electronic Commerce, 38(7) :290-295.
- [16] Turvey C. G., Xiong X. .(2017). Financial inclusion, financial education, and e-commerce in rural china. Agribusiness,33(2):28-35.

Understanding WeChat User's Intention to Use Various Functions: from Social Cognitive Perspective

Shenhao Zhou¹, Xiaoli Yu², Chuan Luo^{3}*

¹²³School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, 611130, China

Abstract: Based upon social cognitive theory, this study explores the effect of personal and environment factors on Wechat user's continuous intention to use various functions. Online survey is used to collect data from the WeChat users. The results confirms that some personal factors (relationship benefit and performance benefit) have a positive effect on intention to use, while image does not have significant effect. Besides, three social environmental factors, the popularity of WeChat, subjective norm and company guarantee, all have significant impacts. Furthermore, we find that environmental factors' effects are stronger than personal factors. Finally, we propose our theoretical and practical implications according to the findings of this study.

Keywords: Social cognitive theory, WeChat, Various functions, Continuous usage intention

1. INTRODUCTION

The Internet technology has witnessed unprecedented boom in recent years. According to the 40th China Internet Development Statistics Report, Up to June 2017, the number of Internet users in China has reached 751 million. The proportion of mobile Internet users achieved a further rise from a high base. Meanwhile, the social network services featured by virtual environment, personalized customization, interactive participation and other new features are expanding rapidly in China and around the world. At present, there are various kinds of Social networking services (SNS) in China, such as QQ, WeChat. These SNS APPs now take various kinds of functions, instead of just an instant messenger.

We found a lot of previous studies focus on the continuous usage intention of WeChat as an instant messenger. For example, Chunmei Gan's studies have shown that pleasure plays the most important role in influencing the use intention of WeChat^[1]. Che-Hui Lien et al. found that WeChat can improve the quality of service so as to improve customer satisfaction, and thus increase the user's will to use^[2]. However, these studies just treat WeChat as an instant messenger, while ignoring the WeChat's various other functions, such as the moments (Friend Circle) function, wallet function and so on. Thus, this study will try to explore why Wechat's users are willing to utilize various new functions of this SNS APP.

The rest parts of this paper organize as follows. First we introduce the theoretical background of this study. We then propose our research model and hypotheses. Then we introduce the methodology of this study. Next we will describe the data analyses and the statistic results. Finally, we will summarize our findings and discuss the theoretical contribution and practical implications of this paper.

2. LITERATURE REVIEW

Through a comprehensive literature review, we found that among the study of SNS' use intention, many previous studies focus on the users' benefits^{[3][4][5]}. The analysis of users' benefits includes interpersonal income, perceived advantage, perceived usefulness, knowledge sharing and other factors. In addition, the significant impact of the social environment on users' use intention has also been confirmed by previous studies^{[6][7][8]}.

* Corresponding author. Email: luochuan@swufe.edu.cn(Chuan Luo)

In previous research, scholars only focus on the usage of these SNSs as communication tools. However, there is little literature to explore the user's intention to use various other functions of these SNSs. Thus this study try to fill in this research gap, we attempt to investigate why the users are willing to use various other functions of WeChat, from the social cognitive perspective.

Table 1. Summary of literature on instant messaging products

Article Title and research target	Findings
Uses and Gratifications of Twitter: An Examination of User Motives and Satisfaction of Twitter Use (2009) ^[3] [Twitter]	The motivation for information acquisition has a positive impact on the use intention of Twitter users.
The role of personality traits in motivating users' continuance intention towards Facebook: Gender differences (2017) ^[4] [Facebook]	The user's personal characteristics, expectations and satisfaction have a combined impact on the their intention to use Facebook.
Exploring the intention to continue using social networking sites: The case of Facebook (2015) ^[5] [SNS]	Perceived ease of use is a major factor influencing user's intention to use SNS. Awareness of fun is also influential.
Continuance Usage Intention of WeChat By Users In Malaysia (2016) ^[6] [WeChat]	Promoting social integration, enhancing friendship, bringing entertainment etc. can promote user's using intention.
Examining Users' Intention to Continue Using Wechat Based on the Expectation-Confirmation Model, Social Presence and Flow Experience (2015) ^[7] [WeChat]	Externality and group behavior have a significant impact on the continuous use intention of WeChat users.
How social influence affects we-intention to use instant messaging: The moderating effect of usage experience (2011) ^[8] [SNS APP]	Group norms and social identity have a significant impact on the user's intention to use SNS APPs.

3. SOCIAL COGNITIVE THEORY

Social cognitive theory originated from the field of psychology, it is proposed by American psychologist Bandura firstly, which believes that the individual factors, the environment factors, and people's behavior are mutually influential and mutually determined^[9].

Jin et al.^[10] used social cognitive theory to study the virtual community's information sharing behavior from different perspectives. Chiang et al.^[11] used social cognitive theory to study the user stickiness of video sharing sites. Mohamed et al.^[12] studied the privacy of users in social networks and the privacy measures they took. In short, social cognitive theory has been widely used by scholars to explain the using behavior of software and network, such as computer using and training, the impact of online social networks on online learning, the formation of digital division, etc.

As social cognitive theory has been widely involved in many aspects of management. This study will utilize social cognitive theory as the theoretical foundation. We consider that the individual factors, the environment factors will affect WeChat's intention to use various new functions embedded in this APP.

4. RESEARCH MODEL AND HYPOTHESES

There are many personal factors that can affect users' intention to use SNS APPs, such as benefits, perceived ease of use, personal preferences, and so on. We consider that among these factors, personal benefits is one of the main factor that affects people's intention to use WeChat. Furthermore, we consider there are three kinds of benefits people can obtain when utilizing WeChat, they are image benefit, relationship benefit and performance benefit.

Porter et al. argue that improvements in self-image can effectively reduce social anxiousness and provide support for social activities^[13]. Nadkarni et al. found that self-presentation is one of the main motivations for people to use Facebook^[14]. For WeChat users, they can chat through private chat, group chat and other chat functions to express their personal views. They can also share photos and publish videos in the circle of friends to show their state of life. We consider they utilize various kinds of WeChat functions to improve their personal image and to get approval from their friends. Therefore, we predict that the image benefit will increase the WeChat user's continuous intention to use various Wechat functions.

H1. Image benefit positively affects the WeChat users' continuous usage intention.

Besides, we consider that the relationship benefit will also affect WeChat users' continuous use intention.

Chao-Min Chiu et al. found that individuals' relationship in the virtual community would promote people's knowledge-sharing behavior^[15]. Tung-Ching Lin et al. found that to strengthen the association between individuals could improve the intention of people to share knowledge^[16]. For WeChat users, we consider that they can utilize various functions to obtain relationship benefit. Such as, they can use the chat function to communicate with others and keep in touch. They can also try "WeChat Swing", "People Nearby" and other functions make new friends, expand the circle of friends. WeChat can meet people's needs of making new friends, maintaining the existing interpersonal relationships, and expanding the circle of interpersonal networks. We speculate that relationship benefit will have a positive impact on users' continuous intention to use various WeChat functions.

H2. Relationship benefit positively affects the WeChat users' continuous usage intention.

We consider that performance benefit is another kind of benefit, which can affect WeChat users' continuous use intention of various functions. Nripendra P. Rana et al. found that the Indian people would prefer to use the online public grievance redressal system(OPGRS) if the system could improve their work efficiency and quality^[17]. While examining the acceptance of tablet PC, Jieun Yu et al. found that the performance benefits perceived during the use of the product are one of the factors that affect the user's acceptance intention^[18]. Now the WeChat is no longer a SNS APP, it has become all aspects of life and work assistant. Using WeChat, people can talk about work, send and receive documents, or use online wallet and other functions. These functions will bring convenience for the user's work and life, which is conducive to improving the efficiency of work and convenience of life. We attribute these benefits as 'Performance benefit' and predict performance benefit perceived by users will positively affect their continuous intention to use various WeChat functions.

H3. Performance benefit positively affects the WeChat users' continuous usage intention.

In addition to individual factors, social cognitive theory believe that the environment factors will also have an impact on people's behavior. This theory consider that environment and behavior are interdependent and mutually determined.

In our research, we consider there are various kinds of environmental factors that could affect wechat users' usage intention, such as the surrounding people and the Wechat's company. Therefore, in our model, we analyses the environment factors from two aspects: interpersonal environment and business environment. Firstly, the usage of WeChat APP is affected by user's interpersonal environment. The surrounding people always affect the individual, with varying degrees of herd mentality. The usage of WeChat around the people will affect the personal usage intention. Secondly, Usage of WeChat is associated with the business environment provided by WeChat's company since it provides system guarantee and technical support for WeChat.

As for the interpersonal environment, we consider that the popularity of WeChat will be an antecedent of the users' continuous usage intention. Most consumer behavior models recognize the importance of social impact on consumers' intentions, they believe that consumers incline to use products that are widely welcomed in their friend circles^[19]. Studies have confirmed that the popularity of the product directly affects people's buying behavior of luxury goods^[20]. Following this line of research, we predict that the popularity of WeChat, deriving from the users' interpersonal environments, can affect their usage intention of various WeChat functions. Furthermore, we consider that WeChat users' friends may also encourage, support people to use this SNS APP since it can facilitate their communications, this social pressure is often named as subjective norm in previous study^{[21][22]}. Thus, we hypothesize that subjective norm will also significant affect WeChat users' intention to use various functions.

H4. Popularity positively affects the WeChat users' continuous usage intention.

H5. Subjective norm positively affects the WeChat users' continuous usage intention.

Previous studies have explored the mechanism of company' institutional assurance^[23]. It is the signal of product quality, and consumers tend to observe this signal to judge product quality^[24]. It also can reduce the

user's perceived risk thus to encourage users to try the product^[25]. In general, institutional assurance have a regulatory role in perceived quality and perceived risk^[26]. Therefore, in this paper, we consider institutional assurance is the WeChat user's business environment factor, which may affect their intention to use various WeChat functions. When the company's institutional assurance can protect users' privacy, rights and interests, they have more reason to use WeChat continuously.

H6. Company guarantee positively affects the WeChat users' continuous usage intention.

Figure 1 shows the research model of this study.

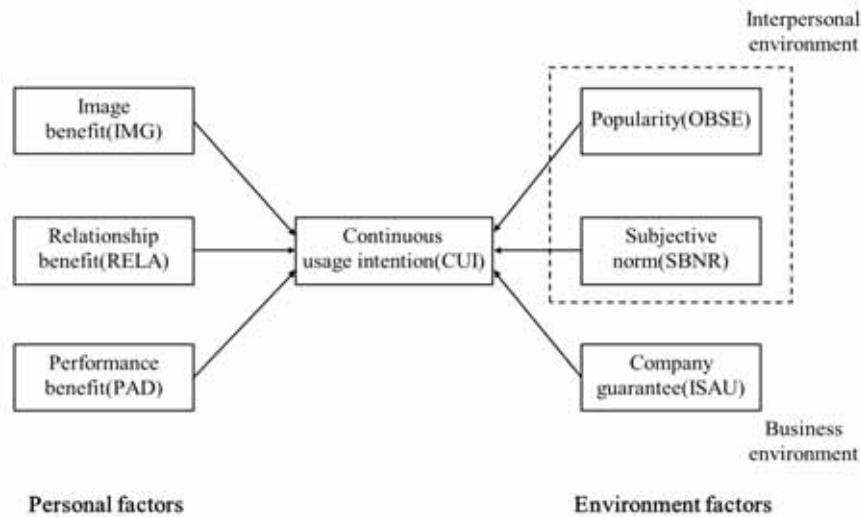


Figure 1. Research model

5. METHODOLOGY

Online survey is used to collect data from the WeChat users in China. We believe the field data from the users can improve the reality of the findings of this study. The questionnaire items for each construct in our research model are adapted from previous studies^{[4][7][8]} with slight amendment to fit our research context. We contact with the WeChat users and invite them to fill in the questionnaire, each user who agreed to join in would get 5 RMB reward. This data collection work lasted 3 weeks, finally we get 532 samples, they are from more than 70 cities in mainland China. Table 2 shows the demographic information of the subjects.

Table 2. Demographic information

Cities (of 100)	74	74.0%
Provinces (of 34)	24	70.6%
Gender		
Male	243	45.7%
Female	289	54.3%
Age range		
<20	73	13.7%
21-25	259	48.7%
26-30	81	15.2%
31-35	43	8.1%
36-40	22	4.1%
40+	54	10.2%
Education		
Junior high school or less	48	9.0%
High school / technical school	71	13.3%
Junior college	54	10.2%
Bachelor's Degree	244	45.9%
Master's Degree and above	115	21.6%
Net age		
<1	7	1.3%
2-4	78	14.7%
5-7	155	29.1%
8-10	140	26.3%
10+	152	28.6%

6. DATA ANALYSIS AND RESULTS

Confirmatory factor analysis (CFA) is utilized to test the measurement model. Fornell and Larcker^[27] suggest that Cronbach's alpha, composite reliability, average variance extracted (AVE), and item loadings can be utilized to assess the convergent validity. As shown in table 3, we confirmed that the values of Cronbach's alpha are larger than 0.8, the composite reliability values are larger than 0.9, the AVE values are larger than 0.7, and the loadings of all the construct items are larger than 0.8. These results confirm the high convergent validity of our data. We further tested the discriminant validity of the constructs in our research model. Table 4 shows, the AVE square roots of all the constructs are much higher than the cross-correlations; it approves the high discriminant validity.

Table 3. Psychometric properties of measurements

Construct	AVE	Composite Reliability	Cronbach's Alpha	Item	Item Loading
Image benefit (IMG)	0.906	0.966	0.948	Item1	0.954
				Item2	0.952
				Item3	0.948
Relationship benefit(RELA)	0.846	0.943	0.909	Item1	0.894
				Item2	0.936
				Item3	0.929
Performance benefit(PAD)	0.732	0.916	0.878	Item1	0.809
				Item2	0.885
				Item3	0.868
				Item4	0.856
Popularity (OBSE)	0.875	0.955	0.929	Item1	0.901
				Item2	0.955
				Item3	0.950
Subjective norm(SBNR)	0.803	0.924	0.877	Item1	0.881
				Item2	0.921
				Item3	0.886
Company guarantee(ISAU)	0.882	0.968	0.955	Item1	0.950
				Item2	0.947
				Item3	0.957
				Item4	0.903
Continuous usage intention(CUI)	0.866	0.963	0.948	Item1	0.941
				Item2	0.948
				Item3	0.911
				Item4	0.921

Table 4. Psychometric properties of measurements

	IMG	RELA	PAD	OBSE	SBNR	ISAU	CUI
IMG	0.952						
RELA	0.673	0.920					
PAD	0.567	0.636	0.855				
OBSE	0.170	0.429	0.376	0.936			
SBNR	0.559	0.626	0.570	0.480	0.896		
ISAU	0.658	0.608	0.568	0.302	0.573	0.939	
CUI	0.516	0.634	0.573	0.518	0.666	0.583	0.930

Smart PLS 2.0 was utilized to test the hypotheses in our research model. In general, the 6 independent variables explained 58.4% of the variance of intention to use. These results approved the validity of our research model. We find that relationship benefit and performance benefit can significantly affect users' intention to use, with $\beta=0.175$, $P<0.01$; $\beta=0.108$, $P<0.01$ respectively, this confirms our H2 and H3. However, image benefit has no significant effect on intention to use, thus H1 is not supported. Besides, we find that all environment factors can affect users' intention to use, with $\beta=0.213$, $P<0.01$; $\beta=0.276$, $P<0.01$; $\beta=0.169$, $P<0.01$ respectively, thus H4, H5, H6 are all supported. All the analyses results are summarized in table 5.

Table 5. Hypotheses test

Path	β	T	Hypotheses
IMG→CUI	0.036	0.685	H1(not supported)
RELA→CUI	0.175	2.867	H2(supported)
PAD→CUI	0.108	2.423	H3(supported)
OBSE→CUI	0.213	5.780	H4(supported)
SBNR→CUI	0.276	5.283	H5(supported)
ISAU→CUI	0.169	3.561	H6(supported)

Dependent variable: CUI, $R^2=0.584$

7. DISCUSSION AND CONCLUSION

7.1 Discussion

Based upon social cognitive theory, this study predicts that personal and environment factors can affect WeChat users' intention to use various functions of this APP. We find that relationship benefit and performance benefit have positive effects on intention to use, while image do not have significant effect. We consider the ineffectiveness of image benefit is due to the following reason: WeChat's initial market-position is a concise social software between acquaintances. User's WeChat friends are mostly relatives and friends around, their personal image do not need to show through this SNS APP. WeChat users have weak demand on their image show. Whereas, WeChat users need to communicate with friends from time to time, so as to maintain their interpersonal relationships. In addition, WeChat can improve users' work efficiency and facilitate their quality of daily life. So the performance benefit has a significant role in promoting user's continuous use intention.

Besides, We analyze the social environmental factors' impact on user's willingness to use WeChat, such as the popularity of WeChat, subjective norm and company guarantee. These three environmental factors all have significant impacts. We further noticed that environmental factors' effects are stronger than personal factors.

7.2 Theoretical contribution and practical implication

This research has some theoretical contributions. First, we utilize social cognitive theory to test users' intention to use various functions of a SNS APP, we find most of the personal and environmental factors in our research model have significant effects. This confirms the effectiveness of social cognitive theory in this research domain. Second, we further partition the environment into interpersonal environment and business environment. This is an important theoretical innovation of this study. We confirm the significance of both kinds of environment factors through statistical analyses. Third, previous studies often treat these SNS APPs as an instant messenger and ignore their various functions. This study innovatively utilize the intention to use various functions of WeChat as the dependent variable and explore its antecedent factors.

In addition to the theoretical contributions, this study also can provides WeChat-developers many practical implications. First, the results show that the environmental factors have positive effects on the user's intention to use various functions. We suggest that WeChat-developers should continue to create a favorable market environment. Through maintaining WeChat friends-circle and carrying out word-of-mouth marketing, WeChat can occupy an irreplaceable position in the market. In short, maintaining the positive effect of popularity and subjective norms on user's intention is necessary.

Furthermore, we confirm that institutional assurance can effectively promote the user's continuous use intention of WeChat. Current WeChat have abundant functions, such as the WeChat Payment, WeChat Wallet etc., the protection of user's property and privacy is very important. Developers should continue to provide more protection measures, from both institution and technology aspects, such as, signing security agreement, standardizing the non-confidential-payment and monitoring user's regularly check of account's security, so that users can use WeChat in a more secure environment.

In addition, relationship benefit and performance benefit can significantly affect the user's intention. WeChat should strengthen its function of communication, it also need to incorporate more function to enhance users' work efficiency. We suggest that WeChat should strengthen the cooperation between computer version and mobile version, so that it can be more convenient for users to transfer office documents. On the other hand, cooperation with other software, such as shopping and game platforms, cooperation with these platforms can promote the diversification of WeChat's functions and create a full range of powerful services.

ACKNOWLEDGEMENT

This work is supported by the Fundamental Research Funds for the Central Universities (Grant No. JBK1802012, JBK120505), the Fund from Collaborative Innovation Center for the Innovation and Regulation of Internet-based-Finance (Grant No. JBK150407), and Fund from Laboratory for Financial Intelligence and Financial Engineering, Southwestern University of Finance and Economics.

REFERENCES

- [1] Chunmei Gan. (2017). Understanding WeChat users' liking behavior: An empirical study in China. *Computers in Human Behavior*, 68:30-39.
- [2] Young-Duk Seo, Young-Gab Kim, Euijong Lee, Doo-Kwon Baik, (2017). Personalized recommender system based on friendship strength in social network services. *Expert Systems with Applications*, 69:135-148.
- [3] Johnson P R. (2009). Uses and Gratifications of Twitter: An Examination of User Motives and Satisfaction of Twitter Use. *Computers in Human Behavior*, 43(3): 245-258.
- [4] Samar Mouakket, (2017). The role of personality traits in motivating users' continuance intention towards Facebook: Gender differences, *The Journal of High Technology Management Research*, ISSN 1047-8310.
- [5] Chia-Chin Chang, Shiu-Wan Hung, Min-Jhih Cheng, Ching-Yi Wu, (2015). Exploring the intention to continue using social networking sites: The case of Facebook. *Technological Forecasting and Social Change*, 95: 48-56.
- [6] M Syuhaidi, A Bakar, (2016).Continuance Usage Intention of WeChat By Users In Malaysia, *Iafor International Conference on the Social Sciences-dubai*.
- [7] Dai B, Liu Y. (2015). Examining Users' Intention to Continue Using Wechat Based on the Expectation-Confirmation Model,Social Presence and Flow Experience. *Journal of Modern Information*, 23: 32-51.
- [8] Shen A X L, Cheung C M K, Lee M K O, et al. (2011). How social influence affects we-intention to use instant messaging: The moderating effect of usage experience. *Information Systems Frontiers*, 13(2):157-169.
- [9] Bandura, A. *Social Learning theory*. (1977). Englewood Cliffs, N.J.Prentic-Hall.
- [10] Jin J, Li Y, Zhong X,et al. (2015). Why users contribute knowledge to online communities: An empirical study of an online social Q&A community. *Information & Management*, 52(7SI):840-849.
- [11] Chiang HS, Hsiao KL. (2015).YouTube stickiness: the needs,personal, and environmental perspective. *Internet Research*, 25(1): 85-106.
- [12] Mohamed N, Ahmad IH. (2012).Information privacy concerns,antecedents and privacy measure use in social networking sites: Evidence from Malaysia. *Computers in Human Behavior*, 28(6):2366-2375.
- [13] Eliora Porter, Dianne L. Chambless, (2017). Social Anxiety and Social Support in Romantic Relationships, *Behavior Therapy*, 48(3): 335-348.
- [14] A. Nadkarni, S.G. Hofmann. (2012).Why do people use Facebook? .*Personality and Individual Differences*, 52:243-249.
- [15] Chao-Min Chiu, Meng-Hsiang Hsu, Eric T.G Wang, (2006). Understanding knowledge sharing in virtual communities An integration of social capital and social cognitive theories. *Decision Support Systems*. 42:1872-1888.
- [16] Tung-Ching Lin, Chien-Chih Huang. (2010).Withholding effort in knowledge contribution The role of social exchange and social cognitive on project teams. *Information & Management*. 47:188-196.

- [17] Nripendra P. Rana, Yogesh K. Dwivedi. (2015). Citizen's adoption of an e-government system Validating extended social cognitive theory(SCT). *Government Information Quarterly*, 32(2):172-181.
- [18] Jieun Yu, Hwansoo Lee, Imsook Ha, Hangjung Zo, (2017). User acceptance of media tablets: An empirical examination of perceived value, *Telematics and Informatics*, 34(4):206-223.
- [19] P. Shukla, (2010). Status consumption in cross-national context: Socio-psychological. Brand and situational antecedents, *International Marketing Review*, 27:108-129.
- [20] Paurav Shukla, (2011). Impact of interpersonal influences, brand origin and brand image on luxury purchase intentions: Measuring interfunctional interactions and a cross-national comparison, *Journal of World Business*, 46(2):242-252.
- [21] Nurul Aqila Hasbullah, Abdullah Osman, Safizal Abdullah, Shahrul Nizam Salahuddin, Nor Faizzah Ramlee, Hazalina Mat Soha, (2016). The Relationship of Attitude, Subjective Norm and Website Usability on Consumer Intention to Purchase Online: An Evidence of Malaysian Youth, *Procedia Economics and Finance*, 35: 493-502.
- [22] Arun Kumar Kaushik, Amit Kumar Agrawal, Zillur Rahman, (2015). Tourist behaviour towards self-service hotel technology adoption: Trust and subjective norm as key antecedents, *Tourism Management Perspectives*, 16:6278-289.
- [23] Liyin Jin, Yunhui Huang, Yanqun He, (2016). When does a service guarantee work? The roles of the popularity of service guarantees and firm reputation, *Tourism Management*, 57:272-285.
- [24] C.A. Kelly, (1988), An investigation of consumer product warranties as market signals of product reliability, *Journal of the Academy of Marketing Science*, 16(2): 72-78.
- [25] H. Hogreve, D.D. Gremler, (2009), Twenty years of service guarantee research: a synthesis, *Journal of Service Research*, 11(4): 322-343.
- [26] Cedric Hsi-Jui Wu, Hsiao-Chun Liao, Kuang-Peng Hung, Yi-Hsuan Ho, (2012). Service guarantees in the hotel industry: Their effects on consumer risk and service quality perceptions, *International Journal of Hospitality Management*, 31(3):757-763.
- [27] Fornell, C., Larcker, D.F., (1981). Structural equation models with unobservable variables and measurement errors, *Journal of Marketing Research*, 18(1). 39-50.

APPENDIX 1

Variable	Construct	Measure
Image benefit	IB1	The use of WeChat can promote my personal image in the circle of friends.
	IB2	The use of WeChat can help me win the approval of my friends.
	IB3	The use of WeChat allows me to get a higher reputation.
Relationship benefit	RB1	The use of WeChat can help me to strengthen my contact with friends.
	RB2	The use of WeChat help me make a lot of friends.
	RB3	The use of WeChat can expand my friend circle.
Performance benefit	PB1	The use of WeChat improves my quality of life.
	PB2	The use of WeChat improves my efficiency of work.
	PB3	The use of WeChat provides useful help for my life and work.
Popularity	PB4	The use of WeChat improves my efficiency of finishing some task.
	P1	I have seen other people using WeChat.
	P2	In my company(school), many people are using WeChat.
Subjective norm	P3	People who use WeChat can be seen everywhere.
	SN1	Those persons who can impact me think I should use WeChat.
	SN2	People around me think we should use WeChat.
Company guarantee	SN3	The environment I belonged to prompts me to use WeChat.
	SG1	The institution and technology of WeChat's company can ensure my security.
	SG2	The institution and technology of WeChat's company can ensure my personal benefit.
	SG3	The institution and technology of WeChat's company can protect my rights.
Continuous intention	SG4	The institution and technology of WeChat's company can protect my private information.
	CUI1	I am willing to use various functions of WeChat continuously.
	CUI2	In a long time, I will continue to use the various functions of WeChat.
	CUI3	In the future, I will often use the various functions of WeChat.
	CUI4	If I need, I will continue to use the various functions of WeChat.

Online-to-offline (O2O) Platforms, Proprietary Platforms and Firm Performance

Xing Wan^{1}, Xue Feng¹*

¹ School of Business Administration, Nanjing University of Finance and Economics, Nanjing 210023, China

Abstract: This paper studies the impact of O2O (Online-to-offline) platforms on the performance of participating firms. Specifically, the study uses a sample of 3406 cinemas to investigate the moderating effect of O2O platforms, particularly proprietary platforms, on the relationship between cinemas' resources and their performance. Resources are measured in two dimensions. One dimension is physical resources, measured by the number of seats, while the other dimension is intangible resources, measured by the operating duration of a cinema. The findings show that both O2O platforms and proprietary platforms have significantly positive impact on the relationship between the number of seats and firm performance. O2O platforms negatively moderate the relationship between duration and firm performance, while proprietary platforms positively moderate the relationship. Implications of the findings are discussed.

Keywords: O2O platforms, proprietary platforms, contingent resource-based view

1. INTRODUCTION

Online-to-offline (O2O) has risen as a business model of electronic commerce. Third-party data and analysis agency iMedia Research has published reports, which showed that the whole O2O markets in China had reached 665.94 billion yuan in 2016, with a year-on-year increase of 42.7%. The O2O model has been penetrated into many services, such as food, education, and medical industries. O2O is defined as a new generation of e-commerce that combines online searching and booking products and services, with consumption in brick-and-mortar stores^[1].

Digital platforms have become the information and business infrastructure of the digital economy^[2]. O2O platforms have emerged, including both third-party and proprietary O2O platforms. For the third-party platforms, companies could benefit from its IT technology and a large installed base of users due to network effects^[3]. However, the competition on the third-party platforms could be fierce, and the level of customers' loyalty is low. For the proprietary platforms, companies can collect data about consumers. As the concept of big data has been widely accepted, the value of the data could be important. In fact, data can be valuable resources for companies. In addition, proprietary platforms can build higher customers' loyalty.

A few studies have investigated the role of O2O platforms on firm performance^[4]. Literature studies whether the characteristics of digital platforms will have influence on firm performance^[5], or whether some specific third-party platforms will influence firm performance^[1]. However, extant literature concentrates on third-party platforms, and few researches investigate proprietary platforms. There are two reasons. Firstly, in O2O models, the third-party platform has occupied large market share, and thus has attracted a majority of related researches. Secondly, with the big data era coming, corporations in O2O industry need to control their own consumer' data, hence we hold proprietary platforms need to be emphasized. This paper intends to fill the gap between empirical researches. Specifically, this paper will explore the role of O2O platforms in the

* Corresponding author. Email: 675975069@qq.com(Xing Wan) ; 925232066@qq.com(Xue Feng)

relationship between firm resources and performance. O2O platforms here include both third-party platforms and proprietary platforms.

This paper has two major contributions. Firstly, the paper contributes to the understanding of the contingent role of O2O platforms in the relationship between firms' resources and their performance. According to the contingent resource-based view, it's the combination of resources, rather than a single type of resource, renders competitive advantage in a specific context. Platforms can provide strategic complementary resources for participating firms^[6]. However, few empirical papers examine the complementary role of platform resources. This paper intends to fill the gap. Secondly, this paper helps to understand the emerging O2O platforms. So far C2C and B2C platforms have been well studied. However, few papers have investigated O2O platforms, which play an increasingly important role on electronic commerce. The O2O business model has greatly increased firms' efficiency and improved consumers' experience. This paper investigates the impact on platform choice in the context of China's O2O platforms. The research and its results will have practical implications.

2. THEORETICAL FRAMEWORK AND HYPOTHESIS

This paper selects the motion picture industry as the context, focusing on cinemas' and exhibition-chain firm's behaviours. According to statistics from the China Report Network, online ticketing accounted for 83.9% market share of the film industry in China's first-tier cities in 2017, 83.5% market share in the country's second-tier cities, and 81.3% and 77.5% market share in the country's third and fourth-tier cities respectively. It shows that O2O platforms have become the main outlet for ticket sales in China's film industry. As O2O model has become a booming business in China^[7], local service firms (LSFs), such as cinemas, regard it as an opportunity to improve their performance by leveraging internet and information technologies^[8]. The traditional value chain of china's motion picture industry is from producers, distributors, exhibition chains, and cinemas to the final viewers. All cinemas need to join one of the exhibition-chain firms to achieve screening permit. Box-office revenues will be divided between them. However, the emerging of digital platforms has reshaped the structure of china's motion picture industry. Due to network effects, digital platform can gather a large number of users^[3]. When a supplier participates in a digital platform, it can access many demand-side users.

In this paper, we discuss three ways for exhibition-chain firms to build proprietary platforms. The first one is based on website which has booking function; the second one is APP software which will be available for the large number users of mobile phones; the third one is based on wechat official accounts platform which can have access to the users of WeChat. In addition to proprietary platforms, companies can also choose to join a third-party platform. In this paper, we will discuss the influence of companies' platform strategy on its performance. We want to find out the impact of O2O platforms and proprietary platforms on the relationship between cinemas' resources and their operating performance.

2.1 Online-to-offline platforms

The resource-based theory (RBT) has been one of the most important theories in business literature, which states that different firms have different resources, and valuable, rare, immobile and non-substitutable resources will bring firms competitive advantage^[9]–^[10]. However, resources rarely act alone in creating or sustaining competitive advantage^[11]. This requires paying attention to the role of other resources or complementary resources, which are linked to primary resources in creating sustainable business value. Synergies between primary and complementary resources have been expected^[12]. Thus, as an extension of RBT, contingent resource-based theory (CRBT) explains that the value of contingent resources upon the context and the linkages between the primary and complementary resources.

Primary resources have context dependency^[12]. In the motion picture industry, seats and duration are discussed as two most important elements. Digital platforms in china not only have reshaped china's motion picture industry, but also provide important resources for participating firms.

The end users of different digital platforms can be different, and the quantity could be large^[13]. Therefore, resources of companies can have access to a wide range of customers. A digital platform provides participants with both an IT infrastructure and a business infrastructure, including financial, marketing and logistics support, leading to a better performance^{[14]-[15]}. In the context of the motion picture industry, O2O platforms have a large installed base of end users. Cinemas linking with these O2O platforms can take advantage of the user resources. Also, cinemas can leverage business functions embedded in digital platforms to market their movies and advertise their promotions, thus both the speed of access to customers and the efficiency of resources utilization will be promoted, and more box-office revenues will be created.

Digital platforms provide inside-out resources for suppliers, among others, IT infrastructure and cost-efficient IS operations. Firms linking with platforms can make use of platforms' IT infrastructure to improve its operation effectiveness and efficiency. In the context of China's motion picture industry, O2O platforms provide an electronic marketplace, which is capable of reducing transaction costs for both customers and cinemas by IT-enabled aggregation and matching services^{[3]-[16]}. The distance between cinemas' resources and the customers has been greatly shorten, and the circulation speed of the resources could be enhanced. For example, an O2O platform enables customers to search, select seats and realize online payment, all of which, comparing to the traditional offline channel, can greatly reduce operation cost and improve resource utilization efficiency of cinemas. Further, online booking trend can help cinemas adjust movie schedules in such a way that cinemas' resource utilization can be improved and thus cinemas can generate more revenues. In sum, we present the following hypothesis:

H1: The higher service fees a cinema's audience pay, the stronger the relationship between the cinema's resources and its performance.

2.2 Proprietary platform

Proprietary platforms can enhance customers' loyalty. For one respect, due to local network effects and people's consumption habits, costumers attracted by proprietary platforms could be more loyal than those on third-party platforms; for the other respect, the combination of proprietary platforms and offline stores could provide users with personalized value-added services, which will improve consumers satisfaction^[4] and consumers' loyalty could be consolidated. Therefore, the demand of consumers booking from proprietary platforms could be steady and accurate, resources utilization of cinemas will be enhanced, thus the box-office revenue of cinemas can be promoted.

Further, proprietary platforms give companies direct access to consumers' data. The installation of end users, information and business infrastructure differ between digital platforms. Thus, participation in multiple platforms will allow vendors to access large end users^[13], and have a more complete and accurate understanding of customers' needs. Further, upsurge in digitization and the increase in the amount of data have promoted companies to reexamine their fundamental business models and explore opportunities to improve and innovate^[17].

In context of China's motion picture industry, through digital platforms, consumers can buy movie tickets at anywhere any time. On one hand, cinemas have more channels to earn box-office revenues; on the other hand, cinemas and exhibition-chain firms may lose a valuable resource, i.e. consumers' data. As the internet economy booms in china, consumer's data has been a valuable resource, from which companies can know more about their target consumer groups^[18]. Thus, it can be the complementary resources. The three different ways for exhibition-firms to build proprietary platforms will generate different knowledge information, and the data

collected from each channel will give firms different and abundant information about consumers. In addition, with more accurate information about the loyal customers of cinemas, higher resource allocation and integration efficiency could be expected. Moreover, adjusting movies schedules according to the analysis of the data could help cinemas make full use of its idle resources, thus more box-office revenues will be generated. In sum, we present the following hypothesis:

H2: The more proprietary platforms an exhibition chain has established, the stronger the relationship between resources of the chain's affiliated cinema and the cinema's performance.

3. METHODOLOGY

3.1 DATA

This paper treats cinemas as the unit of analysis. We use cross-sectional data in the paper, with the sampling period from January 2017 to November 2017. The sample covers cinemas in both first-line and second-line cities of China. To ensure data's availability, selected sample cinemas all adopt computer-aided ticket system (CTS). Finally, we get a sample of 3406 cinemas. Data are collected from Entgroup database, a professional database about China's motion picture industry.

3.2 Variables

(1) Dependent variable. In this article we use theater's performance as the explanatory variable, which is measured by the cinemas' box-office revenues. In china, box-office revenues are the main source of income for theaters. Therefore, it is reasonable to regard it as the cinemas' performance.

(2) Independent variable. This paper uses seats and duration to measure the resources of a cinema. The independent variable $Seat_i$ represents the capacity of audience a cinema can accommodate, measured by the number of seats in the cinema. $Duration_i$ represents operating duration of a cinema, measured by the number of business days from a cinema's establishment to the sampling period.

(3) Moderating variables. There are two moderating variables, $Selfplat_i$ and $Service_i$. $Selfplat_i$ represents an exhibition-chain firms' behavior of establishing proprietary platforms. In this paper, we discuss three ways for companies to build their proprietary platforms. The first is building a web site which has booking function; the second is developing an APP, and the third one is registering a wechat official accounts platform. All three ways will make both cinemas and exhibition-chain firms to have a direct channel toward consumers. When exhibition-chain firms have only one proprietary platform, the value of $Selfplat_i$ is 1; when exhibition-chain firms have two proprietary platforms, the value of $Selfplat_i$ is 2; when exhibition-chain firms have three proprietary platforms, the value of $Selfplat_i$ is 3.

The moderating variable $Service_i$ represents audience' platform usage, measured by platform-charged fees paid by customers. Both third-party platforms and propriety platforms will charge the service fees. However, through investigation, we find that unlike third-party platforms, propriety platforms charge it in a more flexible way. For example, some propriety platforms may only open to its members. That means consumers who may want to use the propriety platforms will have to grant membership first. Considering the local network effects cinemas have, to some extent, proprietary platform may help cinema managers distinguish the loyal customer and know more about their taste. In conclusion, the variable $Service_i$ is a comprehensive indicator encompassing contribution of both third-party and propriety O2O platforms.

(4) Control variables. This paper includes four control variables, i.e. $Onlineratio_i$, $Screen_i$, $Price_i$ and $Attendance_i$. Control variables capture the characteristics of the cinemas; $Price_i$ reflects the market power of a cinema, measured by the average price of tickets at the cinema in the sample months; $Attendance_i$ reflects the

operation efficiency of a cinema, measured by the ratio of the actual number of viewers and expect number of viewers in the sample months; $Screen_i$ represents the ability of a cinema to exhibit films simultaneously, measured by the total number of screens of a cinema.; $Onlineratio_i$ represents the proportion of online revenue.

4. RESULTS

4.1 Descriptive statistics

Table 1 gives descriptive statistics of variables. For all variables, the standard deviations are smaller than their mean value. Variance inflation factor(VIF) was used to check whether the variables have the potential problems of multicollinearity. Result shows that the mean value of VIF is 1.61 and the maximum value of VIF is 2.99. Consequently, the potential possibility of multicollinearity could be excluded.

Table 1. Descriptive statistics

Variable	Mean	Std. Dev.	1	2	3	4	5	6	7
1. Price	3.5076	0.172	1.000						
2. $Screen_i$	6.637	2.736	0.2144***	1.000					
3. $onlineratio$	0.770	0.216	0.0631***	0.3488***	1.000				
4. $Attentance$	13.700	7.517	0.3432***	0.3112***	0.3491***	1.000			
5. $Duration$	6.897	1.282	0.1034***	-0.1113***	-0.0465***	0.2269***	1.000		
5. $Service$	12.486	1.696	0.3053***	0.4728***	0.5602***	0.6085***	0.3010***	1.000	
6. $Selfplat$	1.465	1.107	0.0695***	0.1480***	0.1475***	0.0556***	0.0120***	0.1246***	1.000
7. $Seat$	6.654	0.634	0.1779***	0.7087***	0.3318***	0.2158***	0.0305***	0.5853***	0.1722***

Notes: *** represents 1% significance, and ** represents 5% significance.

4.2 Regression results

Table 2 gives regression results. Model 1 is a baseline model, including control variables. Model 2 adds the independent variables $Seats_i$ and $Duration_i$ based on the Model 1. Model 3 includes moderators $Service_i$ and $Selfplat_i$. Model 4 and 5 include interaction terms between the independent variable and moderators.

The result of Model 2 reveals that $Seats_i$ and $Duration_i$ have a significant and positive impact on the cinemas' operating performance. The coefficient of the interaction terms $Seats_i * Selfplat_i$ and $Duration_i * Selfplat_i$ are positive and significant, thus supporting H2. It demonstrates that, with all else being equal, cinemas have more propriety platforms tend to have stronger relationships between its resources and operating performance. The coefficient of the interaction term $Seats_i * Service_i$ is positive and significant, however, the coefficient of the interaction term $Duration * Service$ is negative and significant, thus H1 is supported partially. It demonstrates that cinemas with high level service fee tend to have stronger relationships between seats and operating performance.

To the interaction term $Duration * Service$, the regression result shows that, the higher a cinema's service fee, the weaker the relationship between its resources and its performance. Since service fee refers to the contribution of O2O platforms, we may explain it in two ways. First, well-established cinemas enjoy stable, large and loyal consumer groups. Even digital platforms provide convenient IT technology, the traditional ticket channel of these cinemas can be mature and familiar to customers. Second, customers of well-established cinemas may face higher switching cost. Well-established cinemas are more likely to establish mature offline channels, including membership. The mature offline operation will help improve customers' experience and enhance their loyalty. Further, customers of well-established cinemas tend to be older than those of newly-established ones. These old customers may meet difficulty in using digital platforms to book and pay their tickets. Thus, consumers of well-established cinemas have less willingness to use the O2O platform.

Table 2. Regress result

	Model 1	Model 2	Model 3	Model 4	Model 5
<i>Screen</i>	1.147*** (0.035)	0.064*** (0.042)	0.201*** (0.025)	0.195*** (0.025)	0.221*** (0.025)
<i>Onlineratio</i>	2.119*** (0.081)	2.203*** (0.067)	-0.676*** (0.052)	-0.687*** (0.052)	-0.640*** (0.051)
<i>Price_i</i>	1.455*** (0.094)	1.217*** (0.078)	0.809*** (0.044)	0.780*** (0.044)	0.768*** (0.045)
<i>Attendance</i>	0.087*** (0.002)	0.075*** (0.002)	0.028*** (0.001)	0.029*** (0.001)	0.031*** (0.008)
<i>Duration</i>		0.282*** (0.010)	0.088*** (0.006)	0.059*** (0.009)	0.385*** (0.027)
<i>Seat</i>		0.695*** (0.030)	0.225*** (0.019)	0.206*** (0.024)	0.061 (0.056)
<i>Selfplat</i>			0.035*** (0.006)	0.036*** (0.006)	0.037*** (0.006)
<i>Service</i>			0.606*** (0.008)	0.609*** (0.008)	0.588*** (0.032)
<i>Seat*Selfplat</i>				0.039*** (0.010)	
<i>Duration*Selfplat</i>				0.020*** (0.005)	
<i>Seat*Service</i>					0.028*** (0.004)
<i>Duration*Service</i>					-0.026*** (0.002)
Adjusted R ²	0.704	0.798	0.922	0.923	0.926

Notes: Values in parentheses are standard errors. *** represents 1% significance.

5. CONCLUSIONS

This paper examines the impact of O2O platforms on the relationship between firms' resources and their performance in the context of China's motion picture industry. Generally, we find that linking with O2O platforms, including proprietary platforms, positively moderates the relationship between a firm's resources and its performance. The resources can be both physical and intangible. Therefore, cinemas' performance is not only affected by primary resources, but also influenced by platform strategy they choose. Companies must realize that, as internet economy booms, consumers' data has been an increasingly important resource. Third-party platforms will allow companies to benefit from their resources and capabilities. The increased potential customers will enhance cinemas' operating performance. However, building proprietary platforms is still an optional way. Considering the local network effect of cinemas and peoples' consumption habits, customers captured by proprietary platforms will be more loyal. Thus, these costumers deserve more attention from cinemas. Data about these costumers will be more instructive for cinemas.

Formulating a platform strategy is a complex thing. A lot of aspects needed to be considered, e.g. whether to join a third-party platform or whether to build a proprietary platform. Some factors should be taken into account, such as the characteristics of the industries and the companies, reputation and customers' habit. The findings show that proprietary platforms and cinemas can have a good synergy of resources. However, the result is different when O2O platforms are seen as a whole.

There are some limitations of this paper, which give directions for future research. First, we investigate the research question in the context of Chinese motion picture industry. To generalize the results of this paper, it is necessary to examine the relationship between platform choice and firms' performance in the context of other industries and other countries. Second, although the sample of this paper is large, cross-sectional data employed cannot reveal the change of the relationship over time. Future research can employ panel data to study the dynamics of the focal relationship. Third, the regression method we adopt can only test the correlation relationship. To establish causal link, future research can adopt methods like propensity score matching or

difference in difference.

ACKNOWLEDGEMENTS

The authors appreciate all the helpful and insightful comments of anonymous reviewers. The authors owe their thanks for the financial support provided by the National Social Science Foundation of China (Grant No. 17BGL200), the Humanities and Social Science Project of Ministry of Education (Grant No. 15YJA630065), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. KYCX17_1170)

REFERENCES

- [1]Wan, X. and J. Chen, *The relationship between platform choice and supplier's efficiency- evidence from China's online to offline (O2O)e-commerce platforms*. Electronic Markets, 2017.
- [2]Wan, X., et al., *Unraveling Platform Strategies: A Review from an Organizational Ambidexterity Perspective*. Sustainability, 2017. **9**(5): p. 1-18.
- [3]Bhargava, H.K. and V. Choudhary, *Economics of an Information Intermediary with Aggregation Benefits*. Information Systems Research, 2004. **15**(1): p. 22-36.
- [4]Wan and Yang, *Internet platform choice, vertical integration and firm performance*. China Industrial Economics, 2017(7): p. 156-174.
- [5]Bharadwaj, A.S., *A resource-based perspective on information technology capability and firm performance: an empirical investigation*. Mis Quarterly, 2000. **24**(1): p. 169-196.
- [6]Fréry, F., X. Lecocq, and V. Warnier, *Competing With Ordinary Resources*. MIT Sloan management review, 2015. **26**(3): p. 69-77.
- [7]Lau, A., et al., *China's iConsumer 2015: A Growing Appetite for Choice and Change*. McKinsey & Company(<http://www.mckinseychina.com/chinasiconsumer-2015-a-growing-appetite-for-change/>), 2015.
- [8]Baesens, B., et al., *Transformational issues of big data and analytics in networked business*. MIS quarterly, 2016. **40**(4): p. 807-818.
- [9]Barney, J.B., *Firm Resource and Sustained Competitive Advantage*. Journal of Management, 1991. **17**(1): p. 99-120.
- [10]Barney, J.B., D.J. Ketchen, and M. Wright, *The Future of Resource-Based Theory*. Journal of Management Official Journal of the Southern Management Association, 2011. **37**(5): p. 1299-1315.
- [11]Wade, M. and J. Hulland, *Review: The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research*. Mis Quarterly, 2004. **28**(1): p. 107-142.
- [12]Chae, B., et al., *The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective*. Decision Support Systems, 2014. **59**(2): p. 119-126.
- [13]Landsman, V. and S. Stremersch, *Multihoming in Two-Sided Markets: An Empirical Inquiry in the Video Game Console Industry*. Journal of Marketing, 2011. **75**(75): p. 39-54.
- [14]Suarez, F.F., *Dethroning an Established Platform*. Social Science Electronic Publishing, 2013. **53**(4): p. 35-41.
- [15]Davcik, N.S. and P. Sharma, *Marketing resources, performance, and competitive advantage: A review and future research directions*. Journal of Business Research, 2016. **69**(12): p. 5547-5552.
- [16]Malone, T.W., J. Yates, and R.I. Benjamin, *Electronic markets and electronic hierarchies*. Communications of the Acm, 1987. **30**(6): p. 484-497.
- [17]Woerner, S.L. and B.H. Wixom, *Big data: extending the business strategy toolbox*. Journal of Information Technology, 2015. **30**(1): p. 60-62.
- [18]Xie, K., et al., *Value co-creation between firms and customers: The role of big data-based cooperative assets*. Information & Management, 2016. **53**(8): p. 1034-1048.

Research on the Influencing Factors of the Willingness to Pay for Knowledge Consumers in the Knowledge Payment Platform

Xinyan Liu¹, Jun Feng^{2}*

¹ College of Business Administration, Zhongnan University of Economics and Law, China

² College of Business Administration, Zhongnan University of Economics and Law, China

Abstract: In 2016, “knowledge payment” had become the most popular Internet phenomenon in China. As we all know, knowledge content is an intangible information product, which is essentially different from the physical product. We cannot judge its quality directly from the external features. When the quality of knowledge products cannot be directly judged, what are the factors affected by the consumer in the process of making a purchase decision? This article studies the factors that may affect the willingness to pay for consumers in the knowledge pay platform. We use Python to collect “Zhihu live” data and analyze it through the SPSS tool. The results show that online social capital of knowledge sharers has a significant positive impact on consumers’ willingness to pay, but there is no significant effect of offline social capital and knowledge price on consumers’ willingness to pay.

Keywords: social capital, price, willingness to pay, knowledge payment.

1. INTRODUCTION

With the development of mobile payment and the change of the structure of social demand, sharing economy has developed rapidly. In 2015, a knowledge program “Logic Show” was popular in the network. Since then, the knowledge sharing economy has developed in China. According to the 2017 China knowledge payment industry development report, Chinese knowledge payment APP active users have reached 520 million people. However, the rapid growth of user scale makes us curious about the outburst of knowledge payment industry. We learn that consumers’ purchase intention is usually driven by five factors: consumer’s individual characteristics, product internal clues, product external cues, consumption situations and socioeconomic factors. Consumers usually make a purchase decision based on the judgment of the quality of the product and the preference of the individual. In the physical products with obvious characteristics, the internal and external clues are mostly reflected in the quality of the product. And knowledge, as a kind of information product, has nonmaterial and experiential. These unique properties make it impossible for consumers to perceive the quality of knowledge externally, and thus may be plagued by the purchase decision. So what can be used as a signal to judge the quality of knowledge in the knowledge payment platform? How do these factors affect the willingness to pay for knowledge consumers? Therefore, this paper, taking the “Zhihu live” as the research object, tries to discuss what factors affect the willingness to pay for knowledge consumers.

“Zhihu live” is a real time question and answer interactive payment product launched by “Zhihu” (China’s largest online knowledge question and answer community). According to statistics, the amount of knowledge transactions generated by “Zhihu live” has reached nearly 100 million Yuan, and the average number of people who participate in knowledge payment is 400 per day. The sharer can create a live, and the live home page needs to display information about sharing knowledge that will appear in the information flow of the followers. General knowledge payment platform includes two basic information. One is the information of knowledge content, such as title, price, start time, comment and so on, the other is the sharer’s personal information, including the number of fans, the amount of attention, the personal profile, and so on. “Zhihu live” has the same

* Corresponding author. Email: yanzido@163.com(Xinyan Liu), hnndfjsifer@163.com (Jun Feng)

characteristics. Knowledge consumers can not directly judge the quality of knowledge, but only by identifying knowledge sharer's personal information and sharing knowledge related information as a signal to judge the quality. Because knowledge usually adopts the pre-sale mode, the review mechanism will not be generated until the end of the sharing. Therefore, this paper explores the factors that influence the willingness to pay of knowledge consumers before the review. In the previous research on the network community, the user's personal information is divided into online and offline social capital. Therefore, through the research, we find that online social capital of knowledge sharers has a significant positive impact on consumers' willingness to pay, knowledge sharers offline social capital has no significant impact on consumers' willingness to pay.

2. LITERATURE REVIEW

2.1 Social capital

2.1.1 The concept of social capital

The concept of "capital" can be traced back to the research of scholars such as Weber, Marx and Engels in economics. Until the early 20th century, Hanifan first described the social capital in his thesis. He pointed out that credit, friendship and social interaction have formed a certain accumulation of social capital ^[1]. Subsequently, the theoretical research related to "social capital" gradually started in the field of sociology. "Social capital" was initially used in community research and generally used to explore networks of relationships ^[2]. Social capital plays an important role not only for communities and individuals but also for human capital ^[3]. The concept of social capital has long been used to articulate a wide range of social phenomena, but Pierre is the first scholar to define clearly social capital. He defines social capital as the sum of real and potential resources embedded in the networks of relationships owned by individuals or social units ^[4]. Later scholars defined the definition of "social capital" on the basis of social relationship network. For example, Putnam pointed out the relationship between social capital representing individuals - social networks and the reciprocal values that are formed on this basis ^[5]. In the management literature, Nahapiet & Ghoshal defined social capital as "the sum of the actual resources and potential resources embedded in the relationship network owned by individuals or social units" ^[6]. Lin Nan argued that social capital was a resource embedded in social networks that enable actors to access and use these resources ^[7]. Although there were many previous studies on social capital, they did not give a unified definition. According to the object and content of the study, this article defines social capital as the resources owned or used by individuals or organizations that are embedded in social networks. The research shows that there is a necessary relationship between social capital and social network. Therefore, this paper has obtained important theoretical support in the research of knowledge payment platform.

2.1.2 Social capital classification and measurement

Putnam suggested that the most important study of social capital is to clarify the dimensions. He divides social capital into bonding of social capital and bridging of social capital. Bonding Social capital performs more strongly in emotional and personal relationships, while bridging social capital reflects more in weak links in social networks. Based on this classification, Williams divided social capital into four types: online and offline bridging social capital and online and offline bonding social capital, in the study of the distinction between online and offline in human relationships ^[8]. In Twitter's research, Hofer & Aubert pointed out that online social capital is related to the number of users ^[9], and some research shows that users can increase online social capital by using Twitter. From a manifestation of the form, social capital can be divided into structural social capital and cognitive social capital ^[10]. Structural social capital refers to objective social structures such as social organizations and networks. Cognitive social capital is a standard, value, attitude and belief that emerges from ideas and consciousness and contributes to cooperative behavior and mutually beneficial behavior. Zhang Lu, in

his study of the community, divides user social capital into online and offline structural social capital and online and offline cognitive social capital^[11]. From macro and micro perspectives, macro-level research on social capital mainly focuses on regional or national perspectives, such as the study of the impact of social capital stock on the economic growth in the region, micro-social capital focuses on the individual resources contained in social networks, such as the structural characteristics of individuals in the overall social network and their own social status^[12]. In the study of online communities, usually focusing on the micro level, for the measurement of individual social capital, scholars generally use the nominal and positioning method. Nominal method is to measure the resources in the network by understanding the information features of the network members. Positioning method can measure the structural position of social capital and the resources in the network. The core content of the positioning method is to calculate the individual social capital by the professional prestige in the social network^[13]. From the definition of social capital, we can see that resources are the core elements of social capital, and these resources include power, status, wealth, information and so on. In a specific network, members can obtain important social capital, which are presented in the form of network connection, social status or reputation. Therefore, on the basis of previous studies, this paper divides the social capital of the sharers in the community into four categories, as shown in Table 1:

Table 1. Classification of social capital

Classification	Structural social capital	Cognitive social capital
Online social capital	Number of fans	Number of praise & thanks
Offline social capital	Institutional rankings	Professional prestige

The study found that social capital has an important influence on user knowledge sharing behavior. Users with more fans can get more likes and higher rankings^[14]; and in reality, users with higher social status are more willing to answer other people's questions and ask very few questions. In the study of the medical online community, doctors' offline prestige and social status will bring more virtual gifts and patients' online voting^[15]. Some scholars also try to explore the relationship between knowledge sharers' social capital and consumers' knowledge payment behavior^[16]. Therefore, this paper argues that social capital is one of the factors that affect consumers' willingness to pay.

2.2 Price

Price is the rate at which a good is exchanged with other good in the marketplace. Price has always been a central issue in economics. In economics, the price appears in the study of signal theory. Stiglitz applies the notion of signal to the market analysis of asymmetric information, believing that price can judge the quality of a good^[17]. The price can be used to measure the economic value of goods and all non-economic value. In the online shopping environment, we view price and seller reputation as the most important external cues that influence shopping behavior. With the development of market economy, the price has not only been judged as quality^[18], but also the price is considered as a sign of loss. When the price is higher, the purchase intention of consumers is relatively lower, Price is considered a financial burden^[19]. However, there is a natural difference between the knowledge product and the traditional commodity, which is original, invisible and lagging in effect.

3. RESEARCH MODELS AND HYPOTHESIS

By summarizing the past research, this paper establishes the following model framework, as shown in Figure 1. knowledge products, as experiential products, cannot directly judge their quality. In knowledge payment platform, consumers can only make judgments of purchasing decisions through information of sharers

and knowledge products. Therefore, this model discusses the influence on the willingness to pay for the knowledge consumer from two perspectives: the sharer information and the knowledge product information. The popularity of products is usually considered as a factor that influences consumers' willingness to pay, but it can only represent the degree of individual preference and cannot directly reflect the quality of knowledge products. so this research will be the popularity of knowledge content as control variables.

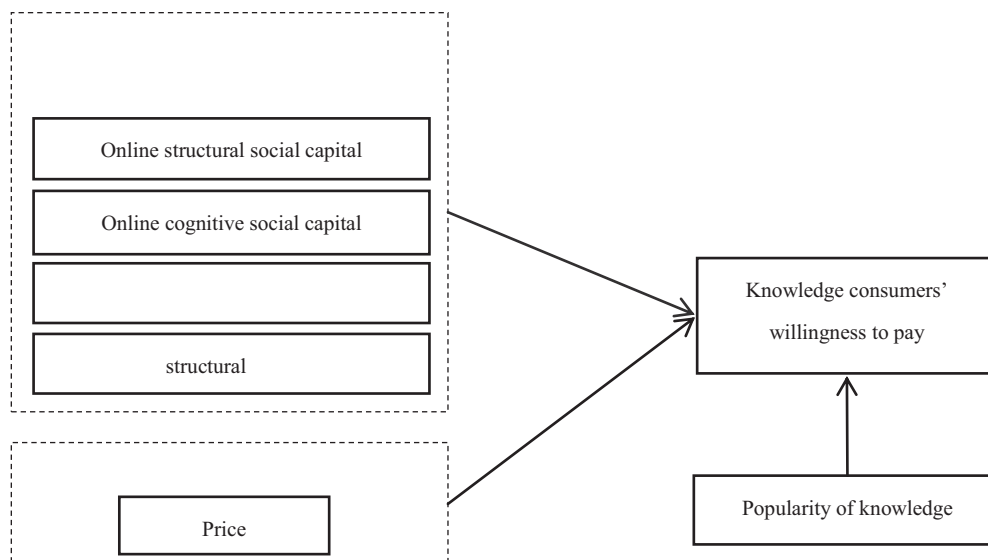


Figure 1. Research Model

3.1 Information of knowledge Sharers and the willingness to pay for knowledge consumers

Through the previous description, we know that in the network community, knowledge sharers will accumulate a certain amount of online social capital through community activities. And these social capitals are stored in the network structure in the form of the number of fans, praise and thanks. Meanwhile, offline social capital accumulated by knowledge sharers in real life through work or study can be displayed in the profile information in the form of occupation, institution, etc. Knowledge shared in “Zhihu live” automatically appears in the information flow of the followers, and more followers mean more users are hoping to get the knowledge posted by their sharers. This paper believes that this is an indirect recognition of the ability of the sharer, thereby recognizing the quality of the knowledge they share. So a large number of fans may attract more people to buy the knowledge. Meanwhile, users can get more thanks and praise through community interaction. These values reflect the high quality of the content shared by the sharer in the community. Therefore, this article puts forward the hypothesis:

H1: the online structural social capital of the sharer has a significant positive impact on the willingness to pay for the knowledge consumer.

H2: the online cognitive social capital of the sharer has a significant positive impact on the willingness to pay for the knowledge consumer.

Knowledge sharers can display personal career experiences and industry information in the community. This article scores the sharer based on Li Chunling’s Professional Prestige Rankings and measures their offline social capital^[20]. Professional reputation often shows people's power status and ability in real life. Under the same field, a university professor may be more persuasive than a worker. Because of the complexity of the industry, it is difficult for us to measure the offline structural social capital. Therefore, this paper only takes into account the offline cognitive social capital:

H3: the offline cognitive social capital of the sharer has a significant positive impact on the willingness to pay for knowledge consumers.

3.2 Information of knowledge products and the willingness to pay for knowledge consumers

Price is a very important external clue for commodity purchase intention, and it is also a sign of the quality of goods. Therefore, this paper argues that the higher the price of knowledge products, the higher the value of knowledge and thus the more people will buy. Propose a hypothesis:

H4: the price of knowledge products has a significant positive effect on the willingness to pay for knowledge consumers.

4. DATA ANALYSIS

4.1 Data collection and processing

This study used Python crawler program to grab all occupational data from the “Zhihu live” platform, and a total of 884 data are obtained. To get the data we need, we first screened out no rating data, then removed the institutional account, the price was 0, and no occupation data. And 202 effective data are finally obtained. In data processing, this article takes the arithmetic average of the number of likes and thanks obtained by the sharer to measure the online cognitive social capital. The professional prestige score was used to measure offline cognitive social capital, and the willingness to pay for the consumer is measured by the number of participants.

4.2 Descriptive statistical analysis

The results of data descriptive statistics are shown in Table 2.

Table 2. Descriptive statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Number of participants	202	56.0000	10107.0000	989.727723	1378.1893342
Online structured social capital	202	45.00	598889.00	52297.9554	99584.88460
Online cognitive social capital	202	.38	3034.01	206.1423	371.52756
Offline cognitive social capital	202	40.75	85.15	65.5530	7.45524
Price	202	9.0000	199.0000	27.668168	18.5860473
Popularity	202	4.00	11560.00	697.4158	1241.53630
Valid N	202				

4.3 Correlation analysis

Then the study conducted a correlation analysis of independent variables and dependent variable. The results showed that the online structured social capital and online cognitive social capital were significantly correlated with the number of knowledge paid participants, while the correlation between offline cognitive social capital and the number of paid participants was not significant. There is no significant correlation between the price of knowledge content and the number of participants. The popularity of knowledge content as a control variable is significantly related to the number of paid participants.

4.4 Regression analysis

The research used SPSS19.0 software to have multivariate linear regression analysis. And the results showed that the goodness of fit was good ($R^2 = 0.424$). In addition, the variance analysis result shows that the

model has passed the set test, which means that the linear relationship between dependent variable and independent variable is obvious ($F=28.868$, $P=0.00$).

In table 3, the regression analysis results show that the cognitive social capital ($\beta=5.132$, $t=0.504$, $p=0.615>0.1$) and price ($\beta=-1.273$, $t=-0.313$, $p=0.754>0.1$) not reached a significant level, so these two variables are eliminated. On the other hand, online structured social capital has a significant positive impact on consumers' willingness to pay ($\beta=0.002$, $t=1.770$, $P=0.078<0.1$). Online cognitive social capital has a significant positive impact on consumers' willingness to pay ($\beta=0.550$, $t=2.285$, $p=0.023<0.05$). Therefore, hypothesis H1 and hypothesis H2 are supported, hypothesis H3 and hypothesis H4 are rejected.

Table 3. Regression Results

Model	Unstandardized Coefficients - β	t	Sig.
(Constant)	33.316	.050	.960
Online structured social capital	.002	1.770	.078
Online cognitive social capital	.550	2.285	.023
Offline cognitive social capital	5.132	.504	.615
Price	-1.273	-.313	.754
Popularity	.658	10.903	.000

Dependent Variable: Knowledge consumers' Willingness to pay

5. CONCLUSIONS

5.1 Research conclusion

Through the analysis, we can clearly see that the online social capital of the participants has a significant positive impact on the willingness to pay of knowledge consumers. In the community, the more fans the sharer has, the higher the online structural social capital it accumulates in the community. Some scholars believe that users' high quality interaction in the online community leads to higher social capital^[11], so the high online structural social capital also means the affirmation of the sharer's ability, at the same time, The knowledge shared by sharers in the "Zhihu live" will appear in the information flow of the followers, so that more users will choose to pay the fees, which is consistent with the hypothesis. In the measurement of online cognitive social capital, we choose the number of praise and thanks of sharers got in community activities, these values can directly reflect the quality of the shared knowledge of sharers in the community. In the premise of judging the quality of the knowledge is not intuitive, consumers should judge the quality of knowledge sharing in "Zhihu live" through the previous approbation of contributive knowledge in the community. As a result, the higher the online cognitive social capital is, the more people are involved in the pay. However, there is no significant effect of offline social capitals on the willingness to pay, which may be explained by the lag in the measurement of occupational prestige. In the Internet age, people's evaluation of many professions has changed.

In the impact of knowledge products information, we see that prices have no significant impact on the number of consumers. In the hypothesis, we think that the price is the reflection of the value of knowledge, and the high price represents the better content of knowledge. But we may have overlooked some problems. Price is not only regarded as a signal of quality, but also considered as a sign of loss. When the price is higher, consumers' purchase intention may also be relatively low.

5.2 Contribution and limitation

This study mainly explores what factors influence consumers' knowledge buying products when they

cannot perceive knowledge quality through reviews, and the results help knowledge providers to understand the willingness to pay knowledge consumers. Sharer can change the strategy to attract more consumers to pay for knowledge, such as: increasing the number of fans for more community activities, improving the quality of knowledge, and obtaining more users like and thanks. Through these community behaviors, Sharer can accumulate more online social capital, so that more users can participate in knowledge payment, and the sharer can get more financial benefits. For a paid platform for knowledge, the standard of pricing knowledge products is one aspect that needs to be carefully considered.

Although in this study, the author strives for rigorous process and procedural science, but there are still some shortcomings. First, the shortage of data samples is an important problem, and the universality of the conclusion needs to be tested. Secondly, when we measure the social capital online, we adopt the occupational prestige index which is lagging behind in time, and some new occupations have not been well reflected. Finally, our research is limited to people who share knowledge for the first time in the payment platform and does not relate to the impact of ratings and auditions on their knowledge willingness to pay. Therefore, we have a lot of work to do in the future research.

REFERENCES

- [1] Hanifan, L. J. (1916). The rural school community center. *Annals of the American Academy of Political & Social Science*, 67(1), 130-138.
- [2] Jacobs, J. (1984). *The death and life of great American cities: the failure of town planning*. Penguin Books.
- [3] Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(Volume 94Number), 95-120.
- [4] Pierre. (1986). The forms of capital. *Handbook of Theory & Research for the Sociology of Education*.
- [5] Putnam, R. D. (2000). Bowling alone: the collapse and revival of American community. *ACM Conference on Computer Supported Cooperative Work (Vol.1, pp.357)*.
- [6] Nahapiet, J., & Ghoshal, S. (2000). Social capital, intellectual capital, and the organizational advantage. *Knowledge & Social Capital*, 23(2), 119-157.
- [7] Lin Nan, Zhang Lei. *Social capital: the theory of social structure and action*. Shanghai people's Publishing House, 2005.
- [8] Williams D. (2006). On and Off the 'Net: Scales for Social Capital in an Online Era. *Journal of Computer - mediated Communication*, 11(2):593–628.
- [9] Hofer, M., & Aubert, V. (2013). Perceived bridging and bonding social capital on twitter: differentiating between followers and followees. *Computers in Human Behavior*, 29(6), 2134-2142.
- [10] Uphoff, N. T. (1996). Learning from Gal Oya: possibilities for participatory development and post-Newtonian social science. *IT Publications*.
- [11] Zhang Lu, Zhang Pengyi. The Relationship Between Online/Offline Social Capital and User Behavior in Social Q&A: The Case of Medical and Health Topics in Zhihu. *Library and Information Service*, 2017, 61(17):84-90.
- [12] Zhang Wenhong. Research on social capital in China: concept, operational measurement and empirical research. *JIANGSU SOCIAL SCIENCES*, 2007(3):142-149.
- [13] Lin, N., & Dumin, M. (1986). Access to occupations through social ties ☆. *Social Networks*, 8(4), 365-385.
- [14] Wang G, Gill K, Mohanlal M, et al. Wisdom in the social crowd: an analysis of quora. *The International World Wide Web Conference*. 2013:1341-1352.
- [15] Guo, S., Guo, X., Fang, Y., & Vogel, D. (2017). How doctors gain social and economic returns in online health-care communities: a professional capital perspective. *Journal of Management Information Systems*, 34(2), 487-519.
- [16] Zhou Tao, Tan Qi. Research on the Behavior Mechanism of Knowledge Payment Users Based on Social Capital Theory. *Modern Information*, 2017, 37(11).

- [17] Stiglitz, J. E. (1987). The causes and consequences of the dependence of quality on price. *Journal of Economic Literature*, 25(1), 1-48.
- [18] Lichtenstein, D.R., Ridgway, N.M.&Netemeyer, R.G. (1993). Price preceptions and consumer shopping behavior: a field study. *Journal of Marketing Research*, 30(2), 234-245.
- [19] Völckner, F., & Sattler, H. (2005). Separating negative and positive effects of price with choice-based conjoint analyses. *Journal of Research & Management*, 27(JRM 1), 5-5.
- [20] Li Chunling. The prestige stratification of contemporary Chinese Society -- Measurement of professional prestige and social economic status index. *Sociological Research*, 2005(2):74-102.

Research on the Evaluation of Green Logistics Based on Cloud Model

Xiulian Jiang^{1*}, *Yining Song*², *Ling Ding*¹, *Zihan Jiang*¹

¹Xuzhou University of Technology, China

²School of Mathematical Sciences, Yangzhou University, China

Abstract: Businesses According to the theory of sustainable development, combining with the current development status of the social logistics industry and the characteristics of green logistics, constructing a green logistics evaluation index system. Using cloud model and Delphi method to calculate the cloud weight of green logistics evaluation index, qualitative and quantitative conversion of evaluation index is realized by cloud generator. Take Jiangsu Province as an example to do empirical research, using the cloud model and its algorithm to get the evaluation cloud of green logistics, observing the evaluation result directly and discovering problem easy by comparing the evaluation cloud chart with ruler cloud chart. The evaluation results show that the cloud model is more reasonable, and the credibility of the evaluation results is improved.

Keywords: cloud model, green logistics, evaluation index system, cloud chart

1. INTRODUCTION

The new technology and new application are emerging in the information society. The rapid development of e-commerce and cross-border e-commerce has prompted the rapid development of the logistics industry. At present, the steady and rapid development of Chinese economy has provided a good opportunity and macro environment for the development of logistics industry. In 2016, China's total social logistics amounted to 229.7 trillion RMB, up 6.1% over the previous year at comparable prices. In 2016, the total cost of social logistics was 11.1 trillion RMB, an increase of 2.9% over the previous year. The total social logistics cost-to-GDP ratio was 14.9%, 1.1% lower than in 2015 (this ratio is commonly 8-9% in developed countries, in the USA is about 8%). This shows that China's logistics market is huge and promising, but the cost is higher than that of the developed countries. Logistics has become an important pillar industry in the development of national economy and social production. In order to realize the rapid and sustainable development of the logistics industry, the sustainable green logistics development strategy is necessary.

2. CONNOTATION OF GREEN LOGISTICS

In the mid-1990s, the concept of green logistics is proposed based on the theories of logistics management, environmental science, ecological economics and ecological ethics^[1]. Green logistics uses advanced logistics information technology to reduce resource consumption in order to reduce pollution to the environment, make logistics and environment develop harmoniously but restrict each other, and realize the symbiotic development of logistics and resource environment^[2]. Domestic and foreign experts and scholars have given different definitions of green logistics from different perspectives, but all think that green logistics can effectively reduce the harm to the ecological environment in the process of logistics activities, and all emphasize the green environmental protection. Green logistics utilizes advanced logistics technology to achieve a series of green logistics activities such as green transport, green storage, green packaging, green recycling (reverse logistics) and so on^[2]. At present, green logistics is studied mainly from macro and micro aspects. On the macro level, from the perspective of supply chain, qualitative research on resource waste and environmental pollution caused by logistics is carried out, and a green logistics system is proposed to protect the environment and reduce

* Corresponding author: Xiulian Jiang. Email: jxl@xzit.edu.cn

pollution. At the micro level, the pollution of the various elements of the logistics is taken as the breakthrough point to reduce the pollution of the various elements of logistics to realize the green logistics. Logistics industry is a resource-driven and resource-consuming industry, how to use the limited resources to achieve efficient and lasting logistics activities are the problems to be solved by modern logistics. Green logistics uses information technology to achieve sustainable resources and friendly environment^[3].

3. CONSTRUCTION OF EVALUATION INDEX SYSTEM OF GREEN LOGISTICS

Green logistics is an important measure for the sustainable development strategy of logistics industry, it and green manufacturing and green consumption constitute a green circular economy that conserving resources and protecting the environment. Green logistics promotes the development of green manufacturing through the reaction of circulation. Green logistics promotes green consumption through green management, effectively allocate and utilize social resources, and realizing the sustainable development of economy society. The rapid development of logistics industry and its enormous challenges all force logistics departments and logistics enterprises to speed up the adjustment of their industrial structure and promote their own sustainable development^[4]. In order to achieve these goals, it is of great significance to carry out green evaluation of logistics. The evaluation system should use the perfect index to evaluate the development of regional Green logistics as far as possible, according to the evaluation result, find out the weak link of the development of green logistics, and provide suggestions for the development of regional green logistics^[5], promote the rapid transformation of local traditional logistics, help to promote the upgrade of environmental protection concept.

The design and construction of the index system includes combing and judging the current research status, the primary selection of the index, the establishment and improvement of the index system. At present, the construction of green logistics evaluation index system in China is still in the research stage, and there is no perfect evaluation index system. The general index system is constructed on the basis of consulting and studying the literatures, statistic indexes and information at home and abroad. After the primary selection of index system is completed, in order to strive for index system to be hierarchical, comprehensive and scientific, use the maximal non-correlation analysis and the expert consultation method to screen and optimize the index. Because of the different researchers, the direction and emphasis of the logistics evaluation index system are different, but it is basically within a certain range. Although they have different perspective on the green logistics evaluation, but the basic focus is on green transport, green storage, green packaging, green recycling, green legal system, green technology and so on. Therefore, this paper will also consider these points in the study of green logistics evaluation, but it will also add other indicators to make the evaluation of logistics green degree more realistic.

Based on the theory of sustainable development^[6], on the basis of the principles and methods of constructing evaluation index system, and drawing on the research results of experts and scholars, combining the current economic social logistics development present situation, the green logistics evaluation index system is constructed (Table 1).

4. GREEN LOGISTICS EVALUATION MODEL BASED ON CLOUD MODEL

4.1 Comparison and selection of evaluation models

The common evaluation methods are AHP^[7], fuzzy comprehensive evaluation method^[8], cloud model^[9], grey relational degree analysis, etc., different research purposes correspond to different evaluation methods.

According to the statistic results, there are many evaluation models used in the comprehensive evaluation research, the cloud model and the AHP are the highest frequently used ones, the other models and their algorithms are affected by the instability factors in the solution, and the ambiguity is eliminated in the

evaluation. According to the purpose of the study, this paper considers the representativeness of the index in the establishment and application of the model. Based on the above analysis, as well as the advantages of the cloud model, its widespread use and high frequency, this paper selects the cloud model as the model for green logistics evaluation.

Table 1. Green logistics evaluation index system and the cloud weight of C-level index relative to B-level index

Target layer (A)	Standard layer (B)	Index layer (C)	(Initial) Ex	Ex	En	He
Green logistics evaluation index system A	Green transport (B1) (0.2513)	Vehicle load (C1)	0.73	0.1613	0.0743	0.0134
		100km fuel consumption (C2)	0.78	0.1719	0.2273	0.0213
		Green vehicle usage (C3)	0.76	0.1692	0.1175	0.0107
		Transport vehicle emission level (C4)	0.80	0.1766	0.0419	0.0054
		Solid waste discharge level (C5)	0.84	0.1855	0.0559	0.0083
		Noise pollution level (C6)	0.61	0.1355	0.0260	0.0038
	Green storage (B2) (0.1608)	Storage damage rate (C7)	0.59	0.2339	0.1377	0.0223
		Turnover rate of goods (C8)	0.63	0.2489	0.0783	0.0085
		Warehouse utilization (C9)	0.71	0.2797	0.0500	0.0018
		Scientific site of storage facilities (C10)	0.60	0.2375	0.2094	0.1373
	Green packaging (B3) (0.1811)	Packaging value (C11)	0.54	0.1594	0.1939	0.0157
		Packaging material recovery rate (C12)	0.67	0.1976	0.0473	0.0044
		Recyclable ratio of packaging materials (C13)	0.75	0.2222	0.0233	0.0044
		Packaging material degradation rate (C14)	0.71	0.2104	0.0892	0.0053
		Green packaging material ratio (C15)	0.71	0.2104	0.0892	0.0053
	Green recycling (B4) (0.1914)	Recovery site coverage (C16)	0.58	0.2960	0.0718	0.0089
		Recovery rate of damaged goods (C17)	0.70	0.3597	0.1435	0.026
		Recycled product sales (C18)	0.67	0.3443	0.1629	0.0147
	Green logistics legal system (B5)(0.2154)	Laws and regulations force (C19)	0.64	0.5898	0.1067	0.0163
		Policy regulation and control (C20)	0.44	0.4102	0.031	0.0077

4.2 Cloud model

The cloud model, proposed by academician Li Deyi in 1995, can effectively realize the mutual transformation between qualitative and quantitative, and has been successfully applied in the fields of evaluation decision, data mining and intelligent control^[10]. Cloud model is the basis of this evaluation method.

4.2.1 The digital features of the cloud

Usually used Cloud (Ex, En, He) to represent the cloud model, where Ex expresses expectation, En is entropy and He is hyper-entropy, and they are the digital features of the cloud. The three numbers can be used to visualize the cloud chart composed of numerous cloud drops^[11-12]. Expectation Ex is the most representative point of the qualitative concept C in the domain space^[11]. Entropy En is used to measure the uncertainty of the qualitative concept C . The bigger the entropy, the bigger the range of values that the concept can accept, and the more fuzzy the concept becomes^[11]. Hyper-entropy He is used to measure the uncertainty of entropy En , indicating the condensation and dispersion degree of cloud drops. The bigger the He , the bigger the randomness of membership, and the more the cloud drops tend to be discrete.

4.2.2 Cloud generator and methods for calculating cloud parameters

The normal cloud generator is a special algorithm implemented by computer^[11], which establishes an interrelated mapping between qualitative concept and quantitative value, it mainly consists of forward cloud generator and backward cloud generator^[13]. The forward cloud generator is a mapping from qualitative to quantitative, which produces cloud drops based on the digital features (Ex, En, He) of the cloud. The backward cloud generator is a conversion model from quantitative value to qualitative concept, it can convert a certain number of accurate data into qualitative concept represented by digital features (Ex, En, He)^[11].

There are two main methods to determine cloud parameters: one is the backward cloud generator method, the other is the index approximation method. There are two algorithms: one is to use the determination degree

information algorithm, and one is to use the uncertainty degree information algorithm^[11,14].

In this paper, a cloud parameters calculation method with uncertainty degree information is used in the backward cloud generator: based on the known information, the 3 digital features of cloud charts are obtained by the reverse cloud algorithm.

Based on the known information, three digital features (Ex, En, He) of cloud chart are obtained by backward cloud algorithm. The methods are as follows^[11]:

Input: many cloud drops x_i , Sample point x_i ($i = 1, 2, \dots, n$).

Output: three digital features (Ex, En, He) of cloud reflecting the qualitative concept.

Step1: through cloud drops x_i compute sample mean, the expectation Ex of the cloud model:

$$Ex = \bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{Setp2: compute } S^2: S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$$

$$\text{Setp3: compute } En: En = \sqrt{\frac{\pi}{2}} \times \frac{1}{n} \sum_{i=1}^n |x_i - Ex|$$

$$\text{Setp4: compute } He: He = \sqrt{S^2 - En^2}$$

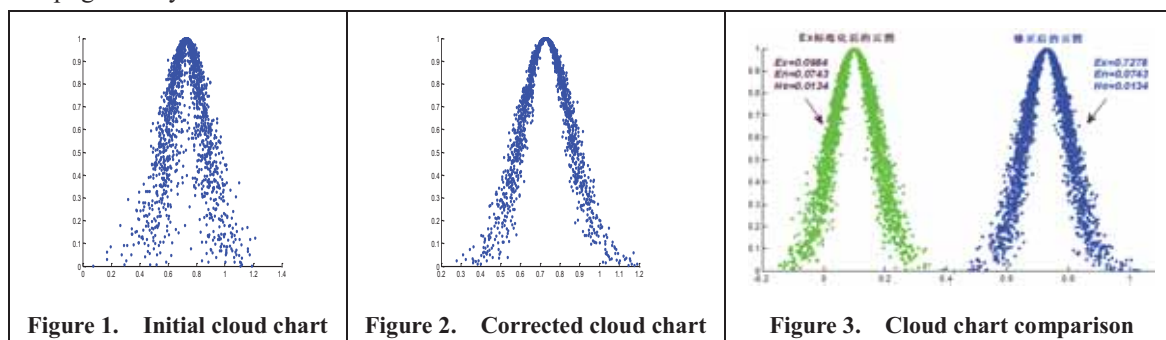
4.2.3 Weight design of evaluation index system based on cloud model

Using Delphi method, the 9 experts from logistics industry were invited to analyze and study each index of green logistics evaluation system, and then score for each index. Based on the scoring table, according to the importance of C_i for B_i and B_i for A to construct the judgment table, the corresponding cloud weight (cloud parameters) can be obtained by compute. Taking the cloud weight compute of “vehicle load index relative to green transport level” as an example, the experts’ initial scoring is shown in Table 2.

Table 2. Vehicle load index relative to the green transport level of importance evaluation table

Expert number	1	2	3	4	5	6	7	8	9
$C_1 \rightarrow B_1$	0.60	0.90	0.65	0.70	0.80	0.90	0.60	0.90	0.50

Using the backward cloud generator formula to obtain the initial cloud weight: (0.7278, 0.1640, 0.0609), the cloud chart that is programmed by Matlab is shown in Figure 1, the initial cloud chart was scattered and the clouds were thicker, indicating that the experts’ opinions were not agreed at this time. Due to differences in expert cognition scored 2-4 rounds. Finally, the perfect cloud chart is shown in Figure 2, at this time, the cloud weight of vehicle load index relative to the green transport index is: (0.7278, 0.0734, 0.0134), expressed the concept gradually formed.



Use the same method to compute cloud weights for other indexes. Because it is the weight, Ex needs to be normalized. The normalized processing method is shown in the Formula (1). According to the connotation of En and He , it is decided that the improved En and He will not be normalized, this will keep the cloud shape unchanged, but shifted left (Figure 3.), cloud chart is still good at expressing concepts. The compute results of

each index cloud weight are shown in Table 1.

$$Ex_i = Ex_i / \sum_{i=1}^j Ex_i \quad (1)$$

After the C-level index is calculated, the same method is used to calculate and improve the relative importance of the B-level index relative to the evaluation system A, the final calculation results in Table 3.

Table 3. Cloud weight integration of B-level index calculated by cloud model relative to evaluation system A

	Ex	En	He
$B_1 \rightarrow A$	0.2513	0.1347	0.0178
$B_2 \rightarrow A$	0.1608	0.1222	0.0176
$B_3 \rightarrow A$	0.1811	0.1685	0.0157
$B_4 \rightarrow A$	0.1914	0.1656	0.0152
$B_5 \rightarrow A$	0.2154	0.2695	0.0319

The Ex value of the B-level index of Table 3 shows that the degree of their impact on the total target A from large to small is the following: green transport, green logistics legal system, green recycling, green packaging, green storage. For the C-level index, the importance of the upper-level is mostly concentrated in 40%-60%, according to the half theory, it can be used as the basic index. This point also shows that in the process of index selection, the factors that cause the repetition of index caused by information sources and personal subjective are relatively small, because each index is more than half important for the upper-level index.

5. EMPIRICAL RESEARCH AND ANALYSIS

5.1 Determine the evaluation set

According to Professor Li Deyi's definition of five-layer normal cloud and evaluation index relative to the level of evaluation subject, the domain can be divided into five green logistics evaluation levels between[0,1]: deep-green (better), green (good), quasi-green (general), light-green (poor), non- green (bad)(Corresponding to Figure 5 from right to left of the cloud charts), the cloud models for five levels are: Cloud1(1, 0.1031, 0.013), Cloud2(0.691,0.064,0.008),Cloud3(0.5,0.039,0.005), Cloud4(0.309,0.064,0.008),Cloud5(0,0.1031,0.013) ^[13].

5.2 Calculation of integrated cloud model

Integrated cloud is a higher level of the parent cloud, is the synthesis of two or more of the same type of sub-cloud generated. There are many clouds $Cloud1(Ex_1, En_1, He_1)$, $Cloud2(Ex_2, En_2, He_2)$, ..., $Cloudn(Ex_n, En_n, He_n)$, composing integrated cloud model, the algorithm is as follows:

$$\begin{cases} Ex = \frac{Ex_1 + Ex_2 + \dots + Ex_n}{n} \\ En = \frac{\max(Ex_1, Ex_2, \dots, Ex_n) - \min(Ex_1, Ex_2, \dots, Ex_n)}{6} \\ He = k \quad (k \text{ is a constant}) \end{cases} \quad (2)$$

5.3 Evaluation of green logistics level in Jiangsu Province

In order to save resources and reduce environmental pollution, Jiangsu logistics industry vigorously promote green low-carbon technology, actively promote the logistics supply side reform, advocate the concept of green logistics, take effective measures to reduce the overall energy consumption and pollutant emission levels of logistics industry, improve the utilization of logistics resources, and further accelerate

the logistics industry informatization and intelligent development, promote the logistics industry green Low-carbon development. In 2016, the total social logistics in Jiangsu Province was 2458684 billion RMB, an increase of 6.5%. In 2016, the total social logistics cost in Jiangsu Province was 1.09736 trillion RMB, an increase of 5.4% over the same period of last year. The total social logistics cost-to-GDP ratio was 14.4%, 0.5% lower than the national, which is 1.1% lower than that of Jiangsu Province in 2010 and 0.4% lower than that of 2015, indicating that the quality and efficiency of logistics operation in Jiangsu Province have been continuously improved. However, there is still a gap compared with developed countries.

Using this research method to evaluate the green logistics level in Jiangsu Province. Based on the above steps to build the cloud model, the 9 experts from logistics industry were invited to score and evaluate for Jiangsu green logistics. Using the scoring method of a given evaluation range, and then select the maximum and minimum of the experts from the evaluation results to calculate the maximum cloud model and the minimum cloud model, and then using integrated cloud to enhance the concept, then getting the integrated cloud evaluation model for each index. The following is an example of a cloud model for green transport index to describe the computational process.

Table 4. The result of expert scoring under the green transport C-level index

Index	Expert 1		Expert 2		Expert 3		Expert 4		Expert 5		Expert 6		Expert 7		Expert8		Expert 9	
	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
C1	0.8	0.6	0.8	0.6	0.7	0.5	0.9	0.6	0.6	0.4	0.9	0.4	0.9	0.3	0.5	0.5	0.8	0.7
C2	0.8	0.4	0.9	0.7	0.9	0.7	0.8	0.6	0.7	0.6	0.8	0.5	0.7	0.2	0.9	0.7	0.9	0.2
C3	0.6	0.4	0.8	0.6	0.8	0.6	0.8	0.6	0.6	0.3	0.7	0.2	0.9	0.7	0.8	0.4	0.7	0.2
C4	0.6	0.4	0.6	0.4	0.9	0.7	0.8	0.6	0.6	0.4	0.8	0.2	0.8	0.5	0.7	0.2	0.8	0.6
C5	0.6	0.4	0.4	0.2	0.8	0.8	0.9	0.7	0.6	0.4	0.8	0.5	0.6	0.1	0.9	0.7	0.7	0.2
C6	0.4	0.2	0.9	0.7	0.7	0.6	0.7	0.4	0.6	0.2	0.8	0.3	0.5	0.4	0.5	0.2	0.7	0.2

Using the data of Table 4 and the compute method of cloud parameters without determining the degree information of the backward cloud generator, the maximum and minimum cloud model of each C-level index of green Transport index is computed, and the comprehensive cloud model of C-level index is obtained by using the integrated cloud computing Method (formula (2)), as shown in Table 5.

From the calculation results, the index C1, C2, C4 between the quasi-green and green states, biased to green, the mean value that is expectation E_x is more than 0.6, according to the $3En$ rules of cloud model, it is known that the value of the 99.74%^[11] falls in the interval. Using the same method to calculate the maximum and minimum cloud index of other C-level indexes, the comprehensive cloud model of each C-level index is obtained by Formula (2), the results are shown in Table 5. The B-level cloud model was calculated using each C-level integrated cloud model and the Formula (2), and the results are shown in Table 6. The results of Table 6 show that Jiangsu Province in five B-level indexes are in the “general” and “good”, in addition to green storage, the other four indexes are biased “good” grade.

Using each B-level index cloud model in Table 6 and the cloud weight of Table 3, according to the multiplication and addition rules of cloud model, the final cloud model of Jiangsu's green logistics level is calculated as:

$$AC = \sum_{i=1}^5 B_i * W_{Bi} = (0.5959, 0.2417, 0.0276)$$

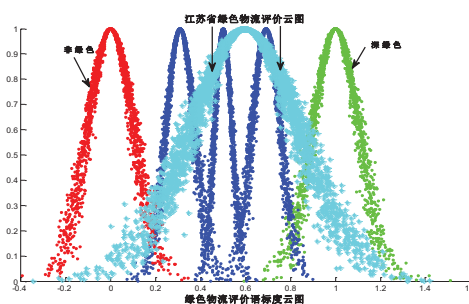
Comparing the final cloud model in Jiangsu with the cloud chart of the scale, as shown in Figure 4, we can see that the level of green logistics in Jiangsu Province is between normal (quasi-green) and good (green), slightly close to the good state.

Table 5. The max, min and integrated cloud evaluation model of each C-level index

Index	Max Cloud Evaluation Model			Min Cloud Evaluation Model			Integrated Cloud Evaluation Model		
	Ex	En	He	Ex	En	He	Ex	En	He
C1	0.79	0.10	0.03	0.48	0.11	0.03	0.6350	0.0517	0.0082
C2	0.82	0.09	0.02	0.51	0.21	0.04	0.6650	0.0517	0.0116
C3	0.74	0.11	0.03	0.44	0.20	0.07	0.5900	0.0500	0.0091
C4	0.73	0.12	0.05	0.44	0.17	0.02	0.5850	0.0483	0.0051
C5	0.70	0.17	0.02	0.44	0.26	0.06	0.5700	0.0433	0.0039
C6	0.64	0.16	0.02	0.36	0.19	0.02	0.5000	0.0467	0.0047
C7	0.63	0.15	0.08	0.54	0.16	0.01	0.5861	0.0157	0.0028
C8	0.64	0.12	0.06	0.44	0.10	0.06	0.5389	0.0333	0.0026
C9	0.66	0.09	0.04	0.50	0.17	0.06	0.5806	0.0269	0.0047
C10	0.64	0.13	0.07	0.47	0.13	0.08	0.5556	0.0296	0.0021
C11	0.66	0.13	0.06	0.49	0.16	0.09	0.5723	0.0278	0.0051
C12	0.63	0.11	0.05	0.53	0.15	0.10	0.5833	0.0167	0.0035
C13	0.73	0.13	0.04	0.59	0.16	0.10	0.6611	0.0241	0.0035
C14	0.70	0.14	0.02	0.57	0.15	0.10	0.6334	0.0222	0.0091
C15	0.63	0.25	0.09	0.54	0.26	0.14	0.5861	0.0139	0.0056
C16	0.66	0.13	0.03	0.49	0.16	0.04	0.5723	0.0278	0.0062
C17	0.63	0.11	0.02	0.53	0.15	0.02	0.5833	0.0167	0.0034
C18	0.73	0.13	0.04	0.59	0.16	0.03	0.6611	0.0241	0.0054
C19	0.66	0.18	0.05	0.61	0.19	0.03	0.6356	0.0082	0.0039
C20	0.61	0.26	0.04	0.54	0.22	0.05	0.5772	0.0109	0.0019

Table 6. The final cloud model of each B-level index

	Ex	En	He
B1	0.5908	0.0275	0.0039
B2	0.5653	0.0079	0.0022
B3	0.6072	0.0148	0.0026
B4	0.6056	0.0148	0.0026
B5	0.6064	0.0097	0.0029

**Figure 4. Evaluation level ruler cloud chart and Jiangsu Province green logistics evaluation cloud chart**

6. CONCLUSIONS

By using the research method of this paper, studying and analyzing the green logistics level in Jiangsu Province, and the results show that the distribution of weights of each index calculated by using cloud model is more reasonable, which improves the credibility of the evaluation result and proves that the green Logistics evaluation method based on cloud model has certain theoretical and practical significance.

“Environment friendly” and “low carbonization” are the important embodiment of social sustainable development. To establish a sustainable society, vigorously promote green logistics is one of the important ways^[15]. First of all, to speed up the construction of low environmental load cycle logistics system, such as multimodal transport, hanging transportation, common distribution and so on. Second, to encourage and support logistics enterprises to use green low-carbon technology, the establishment of a third-party standardized pallet

recycling network, promote the use of energy-saving, clean energy transport tools and logistics equipment. The third is to quicken the construction of the evaluation standard and cognizance system of green logistics, so as to make the green logistics evaluation norm, clear and credible. Four is to pay attention to and speed up the development of the Logistics recovery link, improve the service level of reverse logistics.

ACKNOWLEDGEMENT

This research was supported by the Research Project on Philosophy and Social Science of Universities of Jiangsu Province under Grant 2016SJD630083.

REFERENCES

- [1] Wang Juanjuan, Du Jialin.(2017). Explore the construction of green logistic system in the belt and road region and the path of that. *China Business and Market*,(6): 27-36(in Chinese).
- [2] Xie Sixin, Wang Wenfeng. (2010). The route for green logistics: the strategic choice for green logistic upgrading. *China Business and Market*, (5): 15-18(in Chinese).
- [3] AKYELKEN N. (2011). Green logistics: improving the environ-mental sustainability of logistics. *Transport Reviews*,31(4) : 547-548.
- [4] Hakan Aronsson, Maria Huge Brodin.(2006). The Environmental impact of changing logistics structures. *The International Journal of Logistics Management*,17(3):394-415.
- [5] Shao Chen. (2013). Comprehensive evaluation of green logistics of Inner- Mongolia logistics enterprises based on cloud model. *Logistics technology*, 32(9): 336-338,359(in Chinese).
- [6] Niu Erxuan, Meng Bin, Shen Siyi. (2017). Evaluation model and empirical study of port enterprise green logistics based on cloud model. *Journal of Dalian Maritime University*, 43(2): 67-74(in Chinese).
- [7] SAATY T.(2008).Decision Making with the analytic hierarchy process. *International Journal of Services Science*, (1):83-98.
- [8] Chen W, Hao X-H, Lin J, editors.(2008: IEEE). An improved fuzzy comprehensive evaluation using expanded optimization algorithm for combining index weights. *Machine Learning and Cybernetics, 2008 International Conference on*.
- [9] Meng X, Zhang G, Kang J, Li H, Li D, editors. (2008: IEEE). A new subjective trust model based on cloud model. *Networking, Sensing and Control, 2008 ICNSC 2008 IEEE International Conference on*.
- [10] Lou Hexu. (2014). Research on the Low-carbon and Green Development of Railway Based on Cloud Model. Ms D Thesis. Dalian: Dalian Maritime University, 22-23,32(in Chinese).
- [11] Li Deyi, Duiyi. (2005). *Artificial Intelligence with Un-certainty*. Beijing: National Defense Industry Press, 143-144,154-158(in Chinese).
- [12] Zhang Xudong, Jiang Fangmei, Liu Guifan. (2012). The comprehensive evaluation method of energy-saving project based on cloud model. *Oil & Gas Storage and Transportation*, 31(4): 294-296(in Chinese).
- [13] Sun Huan. (2015). Evaluation on Upgrading of Regional Logistics Industrial Cluster Based on Cloud Model. Ms D Thesis. Shenyang: Shenyang University of Technology, 21-25.
- [14] Li S H, Liu H, Wang J L. (2009). Research on evaluation method of graduates' comprehensive quality based on cloud model[C] // *Proc of ICIA'09*, 815-820.
- [15] Akyelken N.(2011). Green logistics: improving the environ-mental sustainability of logistics. *Transport Reviews*,31(4) : 547-548.

Determinants of online sellers' advertising:

An empirical study on Chinese C2C market

*Weijia You*¹

¹School of Economics and Management, Beijing Forestry University, 10083, Beijing, China

Abstract: This paper studies the determinants of sellers' paid search advertising adoption and their spending on the advertisement. Based on the theory of advertising, we proposed the hypotheses. By an empirical study on a unique dataset of the largest e-commerce platform in China, we find the U shape of the relationship between sellers' reputation and their decision on advertising. Compared to the sellers with medium reputation, the sellers with low and high reputation are more likely to use paid search advertising. And the more loyal customers a seller has, the less chance he adopts paid search advertising. Meanwhile, for those sellers that decide to use paid search advertising, the increased page views with spending on advertisement during last period makes the seller spend more in this period. This study contributes to the advertising literature by providing empirical evidence how sellers make decision on using paid search advertising. And it also provides some managerial implications to the e-commerce platform.

Keywords: Advertisement, Online Market, Heckman Selection Model

1. INTRODUCTION

Different from eBay, which charges both listing fees and sales commission from sellers, the largest e-commerce platform in China, Taobao, doesn't charge anything for listing products on the platform. The "free" policy helps Taobao beat eBay China, and millions of the sellers and buyers are active on this platform. Later on Taobao started monetizing the large traffic through paid search advertising.^[1]

Similar to Google search result, paid search advertisements are displayed alongside the organic search results based on buyers' search queries. Taobao adopted CPS (cost per click-through), which means that sellers only get charged for advertisement that are clicked on. There are conflicting results from different researches on whether advertisement helps the performance. Some studies found that on an average the performance of paid search advertisement (i.e., conversion rates and order values) are better than those from natural search^[2], while some studies found that the average returns on the advertising expenses might be negative^[3]. Although the profitability of paid search advertisement is questionable, we still observe that a lot of sellers compete for paid search advertisement on Taobao and spend significant amount of money month after month. Therefore we raise the research questions: What are the determinants of sellers' decision on advertisement? What influence the sellers, if they choose to advertise, in deciding their advertisement spending?

We collect monthly data of sellers from Taobao, the largest e-commerce platform in China. By observing whether or not the sellers opt in the paid search advertisement, and how much they spent each month, we model sellers' decision-making on advertising in two steps. First, whether to adopt advertisement is modeled as a result of its perception of its competitiveness. And then the decision on advertising spending is modeled as a result of its visitors in the last period. We found that advertising is an option for expanding the number of loyal customers because the more loyal customers a seller has, the less he tends to advertise. We also found that the sellers with low and high reputation are more likely to adopt advertising than those sellers with medium reputation. It implies the different roles advertisement play in the development of the sellers. Meanwhile, as to how much they would spend on advertising, we found that the more positive feedback from the investment during last period, the more they tend to spend on paid search advertisement in current period.

This study contributes to the literature of advertising and sellers' behavior. First, our study extends the advertising literature by empirically examining the characteristics of sellers who adopt paid search advertising. We found that, compared to the sellers with medium reputations, sellers with low and high reputation are more likely to adopt paid search advertisement. As there's evidence that paid search advertising is not profitable on average, this result provides indirect evidence that advertisement are playing different roles in different stage of sellers' development. Also we found that for those sellers that decide to use paid search advertising, positive feedback from the investment on advertisement motivate them to spend more. It reveals how sellers make decisions on the amount to spend on advertising.

2. THEORY AND HYPOTHESES

There have been some empirical studies on the effectiveness of paid search advertisement for individual online retailers on click through rates^[4], conversion performance^[5], and future visits^[6]. Also there're a bunch of studies modeling how sellers bid for the key words^[7] and how buyers interact with keyword advertising campaigns on the buying funnel^[8]. And there have been a lot of discussions on the reasons why advertisements influence consumers' behavior, among which the persuasive view^[9], the informative view^{[10][11]} and the complimentary view^{[12] [13]} are the most widely accepted points. From the persuasive view, entry deterrence happens because it changes customers' utility function by making advertisers' products first choice. Later on, the relationship between advertisement and entry invites a number of analytical studies. According to Needham^[14] and Cubbin^[15], if advertisement does connect to entry deterrence, it only happens when there is a link between pre-entry advertising and the entrant's post-entry expected profit. As advertisements create "noise" in the market and new entrants have to "shout" louder to be heard, we proposes the following hypothesis:

H1: The tendency to advertise decrease first and increase later on as reputation rise.

H2: Seller with more bookmarks will have low intention on using paid search advertisement.

As to how much sellers would spend on the advertisement, Schmalensee^[16] suggested that advertising is a durable investment and it requires a forward-looking seller with a long-term business plan and continuous profit. Therefore, when the sellers make their decision on how much they should spend on the advertisement, the new sellers, who are facing much uncertainties, tend to spend less. While the sellers with high reputation, which are established in the market, tend to spend less compared to sellers with medium reputation. Meanwhile, the increased page views resulted from spending on advertisement in last period increase the capability and intention of the sellers to spend more in this period. Hence we propose the hypothesis H3 and H4:

H3: reputation is related to the sellers' decision on advertising spending.

H4: Seller with more page views resulted from spending on advertisement in last period will spend more on advertisement in this period.

3. DATA AND MODEL

A panel dataset of 1,162 Taobao sellers of baby diapers, with monthly observations from August, 2009 to April, 2010, are used in this study. Of all the sellers, 257 of them have used advertisement at least once (22.1 %). The summary statistics of the main variables are presented in Table 1.

Advertisement (Ads) is a dummy variable showing whether a seller used paid search advertisement in a month. *Ads_Spending* is the total amount of money a seller spends on paid search ads in a month. *Tenure* is the number of months a seller had joined the platform in this month. *Page_view* is the number of the visits to the pages of the products a seller have. *Purchase* is the number of the orders a sellers have during the month. *Bookmark* is the number of times that customers clicked the button of "bookmark". *Reputation* is the cumulative reputation score a seller has. *Sales* is the number of products a seller sold during the month. *Revenue* is the amount of money a seller gain during the month. The correlations between these variables are presented in Table 2.

Table 1: Summary Statistics

Variable	Mean	Std.Dev.	Min	Max
Advertisement	0.151	0.358	0	1
Ads_spending	11235	70420	0	2.204e+06
Tenure	16.28	7.773	4	39
Page_view	6696	23667	0	853820
Purchase	145.0	422.4	0	11299
Bookmark	8.458	14.43	0	227
Reputation	3525	11676	14	322197
Sales	1890	14252	0	746606
Revenue	39603	112324	0	2.558e+06

Table 2 Correlation matrix of major variables

	Tenure	Page_view	Purchase	Bookmark	Ads_spending	Reputation	Sales	Revenue
Tenure	1							
Page_view	0.168	1						
Purchase	0.165	0.891	1					
Bookmark	0.146	0.477	0.487	1				
Ads_spending	0.0917	0.525	0.462	0.134	1			
Reputation	0.275	0.921	0.901	0.467	0.426	1		
Sales	0.0248	0.172	0.200	0.0878	0.0776	0.175	1	
Revenue	0.161	0.737	0.853	0.406	0.306	0.748	0.296	1

In this study, we apply the Heckman Selection Model to sellers' decision-making. First, we model sellers' decision on whether to adopt advertisement or not. Then, we model their decisions on spending. We assume that sellers make their first decision based on their perception of their status as follows:

$$\begin{aligned}
 Ads_{it}^* &= \omega_0 + \omega_1 Reputation_{i,t-1} + \omega_2 Reputation_{i,t-1}^2 + \omega_3 Purchase_{i,t-1} + \omega_4 Bookmark_{i,t-1} + u_{it} \\
 Ads_{it} &= 1 \text{ if } Ads_{it}^* > 0, \text{ and } Ads_{it} = 0 \text{ otherwise; } u_{it} \sim N[0,1]
 \end{aligned} \quad (1)$$

We assume that a seller's decision on whether to use paid search advertisement, Ads_{it} , to be related to their reputation, number of orders during the last month, bookmark, which shows the number of the loyal customers. To be more specific, based on the interviews to the sellers, we found that there is nonlinear relationship between reputation and their decision on advertisement. For the new sellers with little reputation, they need advertisement to gain initial customers urgently. Later on, after they have some loyal customers, the desire to use advertisement to get new customers is not as high as it was. Then when they want to expand the business, advertisement becomes an important option for them again, so we observe the tendency rise with the reputation after a threshold.

In the second stage, we focus on the sellers who have decided to use paid search advertisement. We try to model the sellers' decision on how much they would spend on advertising during the month. In this stage, the dependent variable is seller i 's total spending on advertising in period t ($Ads_Spending_{it}$). Similar to the decision in stage 1, we model the U shape of the relationship between ads spending and reputation, i.e. both reputation and reputation square are included in the model. Besides that, we also take the interaction term of number of visitors during last period and ads spending of last period into account. We expect that the increased page views resulted from spending on advertisement during last period will encourage the sellers to spend more in this period. Therefore, we estimate the model as follows:

$$\begin{aligned}
 Ads_Spending_{it} &= \\
 &\omega_0 + \omega_1 Reputation_{i,t-1} + \omega_2 Reputation_{i,t-1}^2 + \omega_3 Pageview_{i,t-1} + \omega_4 ads_spending_{i,t-1} +
 \end{aligned}$$

$$\omega_5 \text{Pageview}_{i,t-1} * \text{ads_spending}_{i,t-1} + \varepsilon_{it},$$

$$\text{Ads_Spending}_{it} \text{ observed only when } \text{Ads}_{it} = 1 \quad (2)$$

Although we model sellers' decision sequentially, we estimate the two equations simultaneously using maximum likelihood approach suggested by Greene^[17].

4. RESULT

The results are presented in Tables 3. Sellers' decisions of using advertisement are significantly affected by their reputation and the effect is not linear as we expected ($\omega_1=0.246$, $p < 0.001$; $\omega_2 = -0.009$, $p < 0.001$), suggesting that sellers who are more likely to use advertising are the sellers with very low reputation and pretty high reputation. The sellers with medium reputation are the group of sellers who are not very into advertisements. Therefore, hypothesis 1 is supported. Meanwhile, advertisement is used as a tool to gain loyal customers, because the sellers, who have a number of loyal customers, lower the intention of advertisement. ($\omega_1=-0.0936$, $p < 0.001$). It is shown that the greater the number of bookmarks, the less the seller tends to advertise. H2 is supported.

Among those sellers that use paid search advertising, we still observe the U shape between the reputation and advertisement spending. However, the direction is opposite. Sellers with low reputation and very high reputation spend less on advertisement than the sellers with medium reputation ($\beta_1 = -0.171$, $p < 0.01$; $\beta_2 = 0.008$, $p < 0.01$), so H3 is supported. Also, we find that the interaction term has a positive coefficient ($\beta_5 = 0.0176$, $p < 0.01$) which means that if the sellers observe that they have a lot of page views resulted from advertisement spending during the last period, they will increase the spending in this period. H4 is supported.

Table 3: Estimates for Heckman Selection Model

Variables	Selection model	Spending model
Constant	-2.415*** (0.0630)	10.61*** (0.671)
Reputation _{t-1}	0.244*** (0.0318)	-0.203*** (0.0617)
Reputationx ² _{t-1}	-0.00914*** (0.00136)	0.00862*** (0.00259)
Bookmark _{t-1}	-0.0919*** (0.0168)	
Purchase _{t-1}	0.351*** (0.0178)	
Page_view _{t-1}		-0.0936 (0.0624)
Ads_spending _{t-1}		0.0838 (0.0514)
Page_view _{t-1} * Ads_spending _{t-1}		0.0176*** (0.00585)

Note: ***, **, * ==>significance at 1%, 5%, 10% level. All the continuous variables are log-transformed

5. CONCLUSIONS

This study tries to reveal the determinants of sellers' decision on advertisement and we found the U shape of the relationship between sellers' reputation and their decision to use paid search advertising. We demonstrated the ads' effect of gaining loyal customers by showing that the sellers with more bookmarks will have less intention of using paid search advertising. For the sellers that decide to use paid search advertising, U shape of the relationship between sellers' reputation and their spending is also revealed, though different direction from their adoption model. And the increased page views from advertisement spending during last

period makes sellers spend more on advertisement in this period.

There are still some limitations in this study. First, the dataset is limited to the sellers of diapers. It is possible that the sellers of durable goods make decisions in another way. More empirical studies are required to ensure the generalizability of this result. Second, although we try our best to control the factors which would influence the decision of sellers, the result is limited by the innate limitations of observed data. So we explain the association cautiously rather than strong casual effects. Third, this study assumes that the sellers are stable in their decision behavior while we might also take the dynamics of sellers' behavior or the learning effect into account. Panel data with longer period will be collected to deal with this problem.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71402006 71502015, the Fundamental Research Funds for the Central Universities under Grant 2017ZY66, 2017ZY68, Humanities and Social Sciences Planning Fund Supported by Chinese Ministry of Education under Grant 17YJAZH060, and Beijing Municipal Social Science Foundation under Grant16GLC059.

REFERENCES

- [1] Chen J, Fan M, Li M.(2016) Advertising versus brokerage model for online trading platform [J]. MIS Quarterly, 40(3): 575-596.
- [2] Ghose A, Yang S.(2008) Comparing performance metrics in organic search with sponsored search advertising[C]//Proceedings of the 2nd International Workshop on Data Mining and Audience Intelligence for Advertising. ACM, 2008: 18-26.
- [3] Blake, T., Nosko, C., and Tadelis, S. (2015) Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment, *Econometrica*, 83(1): 155–174.
- [4] Yang, S., and Ghose, A. (2010) Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? *Marketing Science*, 29(4):602–623.
- [5] Agarwal, A., Hosanagar, K., and Smith, M. D. (2015). Do Organic Results Help or Hurt Sponsored Search Performance?, *Information Systems Research*, 26(4):695–713.
- [6] Rutz, O. J., Trusov, M., and Bucklin, R. E. (2011). Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead to More Future Visits?, *Marketing Science*, 30(4):646–665.
- [7] Katona Z, Sarvary M. (2010) The race for sponsored links: Bidding patterns for search advertising. *Marketing Science*, 29(2): 199-215.
- [8] Jansen B J, Schuster S.(2011). Bidding on the buying funnel for sponsored search and keyword advertising. *Journal of Electronic Commerce Research*, 12(1): 1-18.
- [9] Robinson, J. (1933). *Economics of Imperfect Competition*. MacMillan and Co., London
- [10] Ozga, S.A. (1960). Imperfect markets through lack of knowledge. *Quarterly Journal of Economics*, 74(1):29–52.
- [11] Stigler, G.J. (1961). The economics of information. *Journal of Political Economy*, 69(3):213–225.
- [12] Telser, L.G. (1964). Advertising and competition. *Journal of Political Economy*, 72(6):537–562.
- [13] Stigler, G.J., Becker, G.S. (1977). De gustibus non est disputandum. *American Economic Review*, 67(2):76–90.
- [14] Needham, D. (1976). Entry barriers and non-price aspects of firms' behavior. *Journal of Industrial Economics*, 25(1):29–43.
- [15] Cubbin J.(1981). Advertising and the theory of entry barriers[J]. *Economica*, 48(191): 289-298.
- [16] Schmalensee, R. (1983). Advertising and Entry Deterrence: An Exploratory Model, *Journal of Political Economy*, 91(4):636–653.
- [17] Greene, W. H. (2012). *Econometric analysis* (7th ed.), Boston: Prentice Hall.

Research on Poverty Alleviation of County E-Commerce

-A Case Study of Luotian County

LiLi¹, BinXu^{1}, HuimingXiao¹, YangLi¹*

¹Wuhan Donghu University, Wuhan, 430212, China

Abstract: Since the introduction of the e-commerce poverty alleviation, e-commerce poverty alleviation has been carried out in the counties. Luotian County is a demonstration county in e-commerce poverty alleviation, which is at the forefront of poverty reduction counties in Hubei Province. Playing the leading role of the county level government, Luotian County established the e-commerce industrial park and the village-level service station, developed the e-commerce industry and opened up the rural logistics. Giving full play to the main role of the market, the county also focused on developing agricultural industrialization enterprises, expanded sales channels with e-commerce, drove the development of featured industries, promoted the growth of the industrial chain, and shared the benefits of the poor households. Besides, it also promoted e-commerce tourism and helped poverty-stricken families get rid of poverty. The paper mainly analyzes the main practice and achievements of e-commerce in Luotian County and summarizes the experience of Luotian e-commerce in poverty alleviation, with a view to providing reference for other counties and regions.

Keywords: county area, e-commerce, poverty alleviation mode, luotian county

1. INTRODUCTION

In the informative era with big data, the E-commerce has become an effective means to alleviate poverty in the new economy. In August 2015, the Poverty Alleviation Office of State Council incorporated E-commerce Project into poverty alleviation system, proposing to develop and nurture the special agricultural products of the poor areas through the modern marketing channels, so as to “introduce the agricultural products into the cities and towns”. Later in 2016, “E-commerce poverty alleviation” was adopted into “the poverty alleviation project with targeted measures”. The poverty-stricken areas around the country have begun to explore and implement e-commerce poverty alleviation and achieve certain results.

2. THE CURRENT SITUATION OF E-COMMERCE POVERTY ALLEVIATION

Foreign researchers have studied the role of information technology in promoting economic development in rural poverty areas.

Adeniji(2010) ^[1]examined the potentials and benefits of Information and Communication Technologies (ICTs) as a means of enhancing food security and alleviating poverty in Nigeria. Adeniji suggested that ICTs can deliver useful information to farmers in the form of crop care and animal husbandry, fertilizer and feedstock inputs, drought mitigation, pest control, irrigation, weather forecasting, seed sourcing and market prices. Other uses of ICTs can benefit farmers, enabling them to participate in advocacy and co-operative activities.

Emmanuel and Muyingi's paper (2010)^[2] Researched mobile commerce application for rural economy development. he suggested the reasons for poverty in poor rural areas, were that these communities are often marginalized in terms of basic infrastructures; low economic activities and limited markets. Information and communication technologies can be used to develop these potentials by opening markets beyond the local borders. Emmanuel and Muyingi conducted a case study in Dwesa community , which is a rural community in the Eastern Cape coast Province of South Africa. A desktop-based e-commerce application was designed for

* Corresponding author. Email: 375446306@qq.com(BinXu).

micro-entrepreneurs in Dwesa. The survey data show that adopting m-commerce into farming and small businesses effectively helped to extend their markets beyond their immediate environment, leverage their income and enhance quality of life.

In China, e-commerce poverty alleviation has achieved outstanding results. The researchers studied the main leaders of poverty alleviation in various regions, and the leaders of different regions were different.

Xiangdong Wang (2015)^[3] explored the Longnan City of Gansu Province, and proposed the Longnan model. The government is the main leader which stimulate and guide the participation of various industries. The success of “the secretary of the walnut” shows the role of grassroots government in promoting poverty alleviation through e-commerce. Yanlong Zhang (2015)^[4] also examined the main methods of the e-commerce poverty alleviation in Longnan. He pointed out that the main role of the government is to build network infrastructure, build network service platform, and cultivate characteristic agricultural product brand.

MinSun (2016) pointed out that Lipu County of Guizhou adopted the poverty alleviation mode of “e-commerce + characteristic agricultural”, and developed the characteristic agriculture- seedling cultivation. The agricultural enterprises play a leading role, and cooperate with the Rural Co-operatives, and the poor households join the rural cooperatives.

Jingranzhe (2015)^[5] focused on the poverty alleviation of Yuan Yang e-commerce in poverty-stricken counties of Yunnan Province. Tourism is the leading industry. With its unique world cultural heritage-Hani terrace, it developed the “Internet + cultural” and “Internet + tourism.” The tourism industry will drive related industries and the whole economic development to lift them out of poverty.

Some scholars have explored the relationship between rural e-commerce participants. Hongli Cheng (2014)^[6] pointed out the role of intermediary organizations, such as agricultural association, rural cooperative and supply cooperative in the poverty alleviation of e-commerce in China. Ke Mao (2015)^[7] et al. researched on the poverty alleviation 020 Xiangxi model of “Internet + company + cooperative+ poor households” of the rural areas in the Changputang Village of Fenghuang County.

Luotian County Government took the lead and established the public e-commerce platform, developed the e-commerce industry, fully exerted the driving role of the agricultural industrialization enterprise, and promoted the perfection of the industrial chain. It will energetically develop e-commerce and tourism to help the poor households participate in the industrial chain and achieve poverty alleviation.

3. THE MAIN PRACTICE OF E-COMMERCE POVERTY ALLEVIATION IN LUOTIAN COUNTY

3.1 The government led the development of e-commerce industry

The government is leading the establishment of e-commerce public service center and e-commerce industrial park, aiming at creating a favorable environment for e-commerce, encouraging e-commerce entrepreneurship and driving the development of rural e-commerce. Luotian Electronic Commerce Public service Center is an e-commerce service organization established by the government. Through the mode of government procurement of e-commerce public service, it is entrusted to professional e-commerce service providers for operation, providing e-commerce policy consulting, technology training, marketing promotion, product control and management and other e-commerce public services.

The government has also led the establishment of county-level E-commerce pioneer park, introduced e-commerce giants such as Taobao, Jingdong and encouraged small and medium-sized e-commerce enterprises to enter the park, encouraging poor households to open online stores and promoting the vigorous development of e-commerce industry. As of September 2016, 40 enterprises and individual businesses have entered the park. From 2014 to 2016, the number of people that open Taobao shops and micro-stores in Luotian County had risen

from 160 to 1,525, which had increased nearly tenfold. The number of entrepreneurs on the Internet had increased from 193 in 2014 to 8,500 in 2016, and the number of traditional enterprises switching to e-commerce had gone up from 26 to 165^[8].

Table 1. 2014-2016 e-commerce related data in Luotian County^[9]

YEAR	2014	2015	2016
Number of individual online stores (100)	1.6	8.9	15.25
Village service station (1)	0	150	180
Training times (100 people)	0	211.25	82.83
Number of online entrepreneurs (100)	1.93	38.77	85
Number of E-commerce related employees (100)	7.6	80	170
Number of enterprises participating into E-commerce (1)	1	26	165
Online sales (10 million)	2.05	28	81

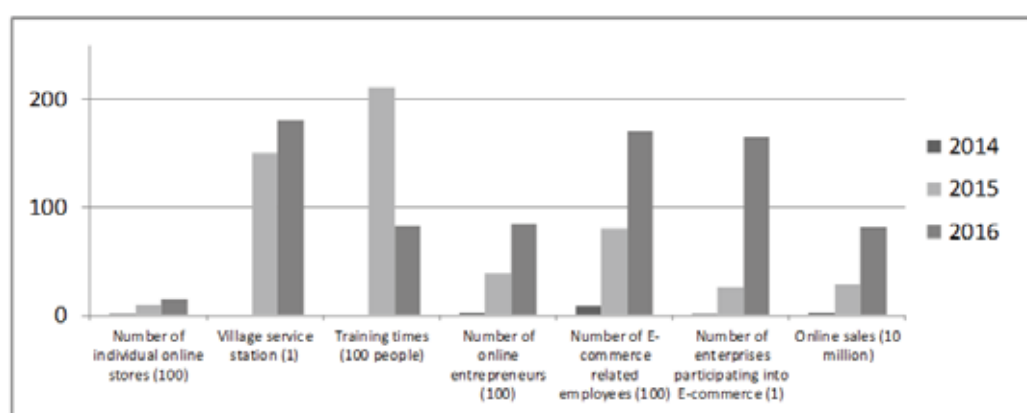


Figure1. 2014-2016 e-commerce related data in Luotian County

3.2 Developing agricultural industrialization enterprises to promote poverty alleviation

The government of Luotian County energetically cultivated the agricultural leading enterprises, participating in the e-commerce chain and realizing the “uplink of characteristic agricultural product”. The mode consists of two ways: one is that the leading enterprise directly produces the agricultural products and cooperates with e-commerce platform, thus promoting the agricultural products all over the country. The other way is that the leading agricultural enterprises establish their own e-commerce platform and sales network, which not only sell their own products, but also purchase the household products. The relationship between the leading enterprise and farmers is that through the agreement with the farmers, the enterprises help farmers package and sell agricultural products.

Table 2. 2016 Agricultural industrialization enterprises in Luotian county

	Chestnut processing enterprises.	Chinese medicinal materials processing enterprises.	Silk textile enterprises	Wine producing enterprises	Characteristic agricultural products processing enterprises.
Number of enterprises (unit)	181	18	12	7	78
Output value (yuan)	1 Billion	1 Billion	0.8 Billion	0.4 Billion	0.58 Billion
Leading enterprise	Huali Food , Lorain Food Liquan Food	Hongyuan Pharmaceutical Huayang Pharmaceutical, Huitao, Chutianshu, Shiweitian	Xinshiji, Jinluo Silk, Deili Silk Luomeng etc.	Chuxiang Wine Industry , BailianfulingWine Industry, Meigong Biotechnology	Daziran Food

Daziran Company is a local leading agricultural enterprise in Luotian County, which began the E-commerce business in 2015 under the support of the government. Daziran Company has developed 218 kinds of distinctive agricultural products including agricultural brand such as Lihuaxiang and Dogili, etc. The company's products are sold through the online platform, and network transactions reached more than RMB16.6 million in 2016^[10]. The government encourages enterprises to carry out poverty alleviation work for poor households and sign an agreement to help them. By the end of 2016, the Daziran Biotechnology Company signed the support agreement with 336 poor households across the county, established production base of 1500 mu, completed underwriting 580000 kg chestnut, 200000 kg sweet persimmon, 280000 kg vegetables, 85000 kg livestock and 65000 kg kongmy, the income of the poor households increased by 8,000 yuan^[10]. So far, only 10 leading provincial-level enterprises of agricultural industrialization such as Hubeimingyang, Daziran, Lorain etc., have signed the poverty alleviation agreement with the 13,700 poverty-stricken households in Luotian County.

3.3 Develop rural professional cooperatives to lift out of poverty.

Rural professional cooperatives expand production scale through the solidarity of farmers, and guide farmers to carry out standardized production and uniform bulk packaging of finished products. Then, online promotion and sales of finished products will be carried out by the joint e-commerce company, such as Taobao store, the flagship store of Luotian and micro-shops. By doing this, the agricultural products can be sold to consumers across the country, which will form stable sales channels and industrial chain and guarantee the income of farmers. Besides, the cooperatives have high profit, so that farmers and cooperatives realize a win-win situation. At the same time, cooperatives and farmers sign a supply and marketing agreement to protect the income of poor households. The cooperatives organize farmers to purchase insurance against the risks caused by inclement weather, bird flu and other communicable diseases in the process of their production and compensate for the losses.

The mature rural cooperatives in Luotian County include: Hubei Jinxiu Forestry and Animal Husbandry Cooperatives in Sanlifan Town, Chinese chestnut cooperatives in Shengli Town, and medicinal cooperatives in Jiuzihe and other towns, livestock breeding cooperatives in Kuanghe and other towns. Before implementing the e-commerce poverty alleviation, Jinxiu Forestry and Animal Husbandry Cooperatives only had 5 households. At present, it has had more than 4000 mu of grassland and mountain forests. There are more than 800 members in the branches, and 45 thousand of the goats in one year. The turnover is over RMB 100 billion. Shengli Town's Chinese chestnut cooperatives were established in 2006. Till now, it has possessed 1,300 members who are from 19 villages and 4000 mu of chestnut production base. The annual output of Chinese chestnut is 2200 tons, and sales volume is more than RMB 7.69million^[11]. By the end of 2017, there have been 335 market entities such as various professional cooperative organizations and family farms in Luotian County, and more than 1,600 poverty-stricken households, which have lifted more than 1,600 poverty-stricken families out of poverty^[12].

3.4 E-commerce + tourism for poverty alleviation

The government of Luotian County spared no efforts in building the e-commerce mode of “e-commerce + tourism + poverty alleviation + current,” which vigorously promoted the industrial transformation and upgrading and increased farmers' income from level of poverty line. In 2016, Luotian County vigorously promoted the integrative development of “the Internet +the whole region tourism” and built a “smart tourism” system. Through all levels of government websites such as “Luotian E-commerce”, “Luotian Tourism”, “Luotian Wenlian” and other WeChat public accounts, it made the publicity and promotion of Dabie Mountain's famous scenery, history humanities, specialty snacks and scenic spots. Through the establishment of

Luotian Tourism Flagship Store, and Characteristics Luotian Museum in Chins, it made the comprehensive layout of tourism products and agricultural products. While enjoying the special food of eco-tourism, consumers can scan the QR code to buy the tourism products and agricultural products that they have just seen or tasted.

The convenience of e-commerce tourism and the rapidness of information transmission have brought more tourists to Luotian. In Luotian, there are 75 tourist spots, and 5.35 million domestic tourists were received in 2016, an increase of 15.1 over the previous year^[13]. By means of scenic areas, hotels and rural tourism, Luotian County gave priority to the placement of the surrounding poor villages and the poor population employment, and gave priority to contract sales of its native chicken, vegetables and other agricultural products. In the first quarter of 2016, more than 350 poverty-stricken households were lifted out of poverty with the support of job placement, targeted acquisitions and sales of travel products from the scenic area. The 656 guesthouses in the county have lifted more than 500 poverty-stricken families out of poverty. The 33 rural tourism destinations have lifted 420 poor households out of poverty. The 8 tourism commodity processing enterprises have lifted 95 families out of poverty^[14].

3.5 Setting up the “Ganjie” model to improve the rural logistics system.

In 2015, Luotian adopted the government procurement method to sign with Zhejiang Suiwang Company to set up Luotian Ganjie Company, and established a rural e-commerce service station in each village for the connection between urban and rural areas. Specifically, it established 3-level logistics stations in county, town and village and signed a cooperation agreement with express delivery company. Through 3-level logistics stations of county, town and village, it fulfilled the express delivery from county to rural areas. It also takes agricultural products to the city to realize the two-way circulation function of “consumer goods into the countryside” and “agricultural products into the city”.

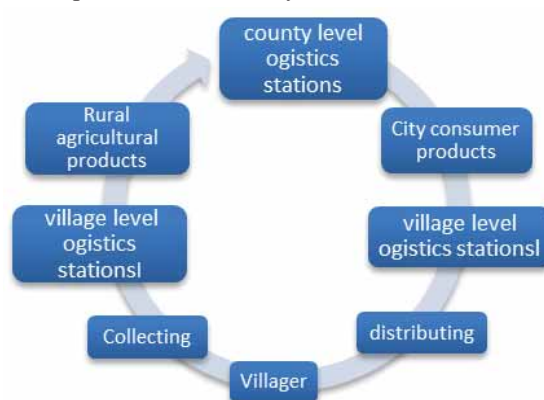


Figure 2. Flow chart of Ganjie Mode

There are more than 180 village level service stations in Luotian which connect the urban and rural areas. In 2015, the trade volume of the service station was over 6 million. In 2016, the online trade of the village service station in the county reached RMB 9 million, saving RMB 3 million for the villagers^[15].

Table 3. The growth of village service stations' sales and online sales

Year	Growth number of village service stations	Sales growth rate (100 million)
2014	0	0.2
2015	150	2.5
2016	180	8



Figure 3. village service stations

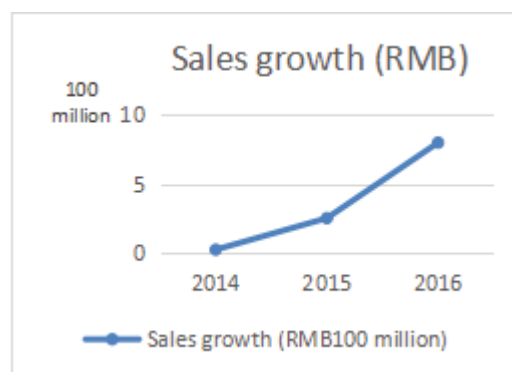


Figure 4. Sales growth (per RMB100 million)

4. THE ADVANTAGES OF LUOTIAN COUNTY POVERTY ALLEVIATION

4.1 Having a remarkable effect on poverty alleviation

By the end of 2016, 22,177 households and 72,137 people in the county got rid of “poverty hat” and 63 impoverished villages were out of the list. The third-party evaluation about the mass satisfaction was 100%. There were 165 village service stations, 85 e-commerce shops, and 5 million annual logistics packages, annual turnover amounting to 50 million yuan. The e-commerce sales volume is 560 million yuan, accounting for 12% of the total retail sales of consumer goods, with 575 e-commerce enterprises, 53 e-commerce platforms, and 8,000 employees. As of January to August 2017, the online trading volume of Luotian county had exceeded 560 million yuan, with a year-on-year growth of 20%, among which 160 million yuan was sold on agricultural products, and the income of the poor households increased by 16.5 million yuan.^[16]

4.2 Expanding the distribution channels for agricultural products

First and foremost, the export market has been expanded. The main agricultural products of Luotian County were exported to more than 20 countries, including the United States, Japan, Germany and the Southeast Asia, and the annual export volume is more than 20 million us dollars. Secondly, the domestic market effect has been increasing. Most agricultural products were mainly sold to more than 30 cities and regions such as Guangzhou Shanghai, Shenzhen, Zhuhai etc. In 2016, the deep processing enterprises of special agricultural products gained industrial output value of 4.28 billion yuan, achieved a profit tax of 250 million yuan, and resettled more than 30,000 people. Last but not least, it opened the supermarket channels, for example, the fried sugar chestnut made by Lorain Company has entered a large chain supermarket such as Walmart and Carrefour, etc in Wuhan.

4.3 Enhancing the economic and social benefits

For one thing, the farmers have increased their income. Since the implementation of the e-commerce poverty alleviation, through leading enterprises and rural specialized cooperatives, Luotian County has realized the deep processing of agricultural products, which has increased the demand for agricultural products. One-third of the county's net income comes from the planting, raising, processing and sales of special agricultural products. For another thing, the fiscal revenue has risen. Over the past three years, the tax revenue generated by the deep-processing enterprises with special agricultural products in the county has been increasing by 20% every year, accounting for 40% of the annual tax revenue of the county. Finally, the enterprise has made more benefits. Since the implementation of the e-commerce poverty alleviation project, the deep processing enterprises of agricultural products in Luotian County developed rapidly and the number of enterprises in scale

increased from 15 in 2005 to 33 in 2016, accounting for more than 30% of the total industrial output of the county. Furthermore, the industrial chain has been expanded. The leading enterprise of agricultural industrialization and rural professional cooperatives have cultivated a large number of processing enterprises of agricultural products and fostered a large number of planting bases which have achieved specialization, standardization and large-scale planting and production, so that poor households can participate in the entire industrial chain of planting, product packaging, transportation and express logistics to share profits, expand employment and achieve poverty alleviation.

5. EXPERIENCE OF LUOTIAN COUNTY IN POVERTY ALLEVIATION

5.1 Relying on local industry

Luotian County relies on the development of local industry, which plays the leading role in agricultural industrialization. These companies have produced the characteristic agricultural product. Through the publicity and sales of e-commerce channel, Luotian County expanded sales volume and strengthened the industry, which formed large-scale production, processing, marketing, and promoted the common development of processing industry, transport, tourism, service industry and other industries, thus shaping a complete industrial chain. Only the larger the scale of production, the longer the industrial chain can lead to the greater the added value of the product and the greater the income of poor households. In the following Figure 5, the chestnut industry is taken as an example, which explains the development of agricultural products in processing enterprises which promotes the development of the whole industry chain and industrial integration, so to increase the income of poor households.

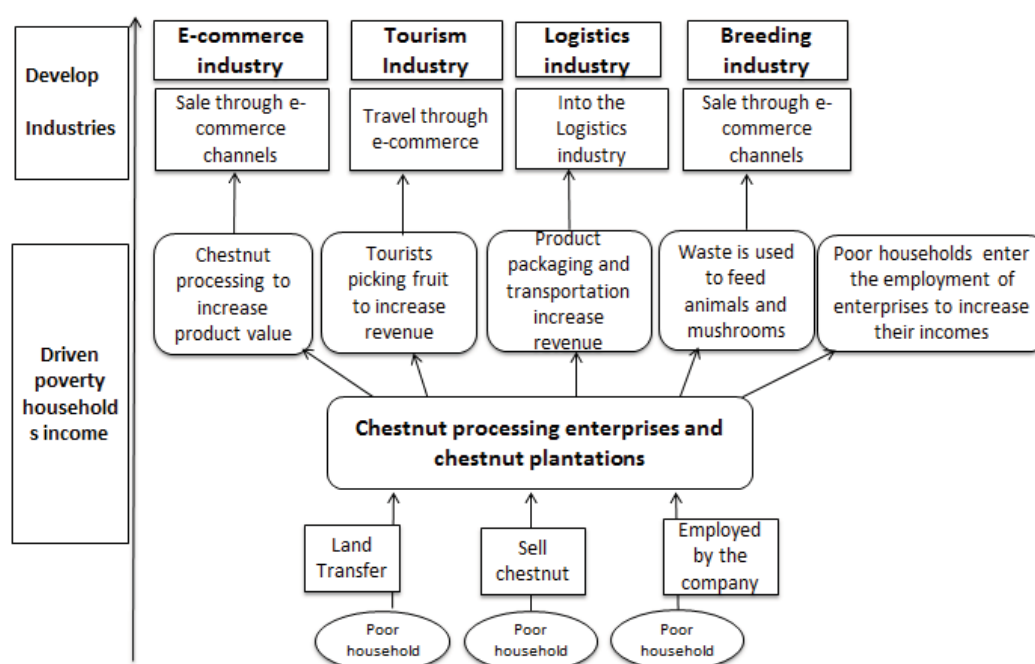


Figure5. Mechanism of Agricultural products processing enterprises to promote industrial development and poverty-stricken households income

In 2016, the five Characteristic industries of Luotian County were composed by chestnut industry, Chinese medicinal industry, tourism industry, livestock breeding industry and the e-commerce industry to poverty alleviation, as shown in Table 4.

Table 4. The situation of five Characteristic industries for driving poverty alleviation in Luotian County^[17]

2016	chestnut industry	Chinese medicinal industry	livestock breeding industry	Tourism industry	E-commerce industry
Growing area for planting	33760mu	4000 mu	33175mu		
Total income OR Transaction amount	unknown	unknown	unknown	total income 2.96billion yuan.	Online sales volume 810 million yuan.
The number of poor households that participate in	More than 10,000 households	1946 households	unknown	More than 1800 households	
The number of poor households opening online stores					64 households

5.2 Developing a market led by the government, with multi-party participating

Above all, Luotian County fully played a leading and synergizing role of the government, completely mobilized the active individual subjects in the social economy such as the initiative of leading enterprise, financial industry and put them into the e-commerce poverty alleviation system. Luotian County Government put forward the five-in- one poverty alleviation mode of “government + main market player + bank + insurance + poor household” . It gave full play to the main role of the market – the leading role of leading enterprises, gave play to the promotion role of e-commerce enterprises and strengthened the support function of the party branch. It speeded up land transfer to promote large-scale planting and production, cultivated a number of physical enterprises and built a number of scenic spots, which made the industry community of whole county integrated and promoted.

Secondly, policy support is a guarantee. Luotian County Government successively issued “the supporting methods of Luotian e-commerce,” “the implementation suggestions to accelerate the e-commerce development”, “the implementation method of Luotian County industrial precision poverty alleviation”, “the implementation plan of Luotian County e-commerce into rural work” and other policies for supporting the development of e-commerce. It formulated various incentive policies, including Incentives for online shops, incentives for enterprises (such as industry, circulation and service industries), incentives for support of online stock suppliers, incentives for supporting service systems, and advanced unit for poverty alleviation through e-commerce and advanced individual reward policy. The government also rewarded farmers to transfer their land to leading enterprises, professional cooperatives and large planters, the transfer area amounting to more than 20 mu. The government granted 100 yuan / year / mu of circulation allowance to poor households. In terms of financial support, the government provided 30,000 yuan of discount loans and 10,000 yuan of support funds for each poor family. The government gave 5000 yuan to the enterprises supporting one poverty-stricken family. As for the enterprises that need to deal with the acquisition funds, the government provided discount loans at the standard of 30,000 yuan per household. The government also guided financial institutions to help leading enterprises (specialized cooperatives) of planting agricultural products and signs a poverty reduction agreement. Additionally, the government signed the insurance agreement of poverty reduction with the poor households of leading enterprises (professional cooperatives).

5.3 Attaching importance to e-commerce service support system, and creating a favorable environment for poverty alleviation

For starters, it strengthened the infrastructure. The construction of rural roads added to nearly 800

kilometers from 2016 to 2017, and nearly 300 million yuan will be completed. In 2016, the government started to launch the upgrading of 735.96 km of rural highways in the 12 townships and 275 administrative villages of the county. In October 2017, the upgrade of rural highways in the county was completed by 95%^[18]. It will continue to invest in mobile network construction and broadband construction, expand network coverage, and increase the rate of optical fiber occupancy. As of June 2016, Luotian County had completed 228 of 2G base stations, 76 of 3G base stations, 281 of 4G base stations, and realized 100% network coverage in the county.

Secondly, the government paid attention to the construction of logistics system and strengthened the integration of e-commerce and logistics. Through reasonable reference of Suichang government mode of procuring public service by e-commerce, it established Luotian Ganjie E-commerce Co., Ltd, which not only serves as the public service platform of e-commerce, but also plays the role of e-commerce operation center and logistics transit. The “Ganjie Model” established by Ganjie Company has established a village-level service station by utilizing the densely sites such as existing convenience stores, canteens and recycling stations of in rural areas to solve the problem of “the last mile” and realize the two-way circulation function of “consumer goods into the countryside ” as well as “agricultural products into the city” .

6. CONCLUSION

With the guidance of the government, Luotian County mobilized various parties to participate in the e-commerce poverty alleviation. With the development of the industry as the root, it cultivated a number of leading enterprises by supporting the development of rural professional cooperatives to form a complete industrial chain of scale and standardization’ s cultivation and processing and sales and tourism. Farmers have been lifted out of poverty by participating in industrial chains and sharing benefits. By means of the Ganjie Mode, it established the village-level service station, which formed the three-level logistics network and expanded the distribution channels for agricultural products, increased income from poor households, and lifted poor people out of poverty. Through the establishment of e-commerce industrial park, the introduction of e-commerce giants from the outside, and the cultivation of local small and medium-sized e-commerce enterprises, the government encouraged the poor households to open stores to start businesses and realizes poverty alleviation. Through the promotion of “e-commerce + tourism” , it increased passenger traffic, promoted the poor households to participate in tourism services, and achieved poverty alleviation.

There are also some problems in the poverty alleviation of Luotian e-commerce. First, it lacks the backbone enterprises to play the role of promotion and demonstration. The scale and benefit of the main backbone enterprises are still not enough. As a result, the demonstration effect is not obvious for the development of e-commerce in the whole county. Second, the varieties of featured products are scattered and their collection and package are difficult. Besides, there are less characteristic products that are suitable for the online marketing and it lacks the characteristic products of online marketing that have registered certification. Brands are also more fragmented and less integrated. For example, there are many brands of chestnut such as Jiahui, Xinfuyuan, Hongyun and Sugar fried chestnut (Tangchaobanli) and other brands, but the reputation of these brands are not high. Thirdly, the high cost of rural logistics has affected the rapid development of e-commerce.

For the problem of brand integration, it is suggested to play the coordinating role of the trade association and coordinate the unified production standard of agricultural products. Meanwhile, it is also suggested to strictly classify the quality of agricultural products, ensure the quality, concentrate on building several key brands, and unify the price and the name. For the higher cost of rural logistics, it is suggested to further strengthen the logistics infrastructure construction, integrate the express industry and establish a logistics information sharing platform, while adopting big data and real-time information to integrate capacity, share

vehicles and personnel, and reduce logistics costs.

ACKNOWLEDGEMENT

This work was supported by the grants from Hubei Provincial Collaborative Innovation Centre of Agricultural E-Commerce(Wuhan Donghu university research [2016] No. 15Document).

REFERENCES

- [1] Adeniji,O.B.(2010). Potentials of Information and Communication Technologies for Poverty Alleviation and Food Security. *Journal of Agricultural Extension*,14(2):130-138.
- [2] Edim Azom Emmanuel, Muyingi Hippolyte N.(2010).A Mobile Commerce Application for Rural EconomyDevelopment: A case study for Dwesa. *Research Conference of the South African Institute of Computer Scientists and Information Technologists: Eastern Cape*,ACM 978-1-60558-950-3/10/10,58-66.
- [3] Xiangdong Wang, Wangxintian. (2015). E-commerce and Information of Poverty Alleviation: New Features of Poverty Alleviation Work in Internet Era. *Journal of Northwest Agricultural and Forestry University (Social Science Edition)*, (4)15:98-105. (In Chinese).
- [4] Yanlong Zhang, Gao Ke. (2015). Promoting the new path of precise poverty alleviation - Inspiration from E-commerce Poverty Alleviation of Longnan City, Gansu Province. *China Engineering Consultation*, (10): 40-41. (In Chinese).
- [5] Jing Ranzhe. (2016). *Internet + Yuanyang Mode of Poverty Alleviation+ County area*, Electronic Industry Press, 45.
- [6] Hongli Cheng. (2014). Analysis On Framework and Mode Selection of Rural E-commerce Development Model- Based on Research Perspective of Taking Farmers As Producers. *E-commerce*, (11):28-31; In Chinese. (In Chinese).
- [7] MaoKe (2015-12-26). *Ecommerce in Poverty Alleviation*, Liupanshui Daily (1).Hu Wenling, Sun Suran, Li Yan. (2017). Innovation Research of County-area's Agricultural E-commerce in Strategy of "E-commerce" Poverty Alleviation . *Commercial Economic Research*, (24):25-29. In Chinese.
- [8] YanmingMa,ZhilunLin.(2015-11-20). "Ganjie-model " the spark is a prairie fire. *East hubei Evening*,(16).
- [9] Commerce Department of Hubei Province.(2017).E-commerce Practice in Luotian County . www.hbdofcom.gov.cn/cszz/sctxjsc/ncds/45195.htm
- [10] ShiminLiu.(2016-10-11).Luotian County "thousands of enterprises to help the village" accurate poverty alleviation. *Huanggang Daily*,(01). (In Chinese).
- [11] Rural newspaper .(2014-08-14). Rural cooperative economy has occupied half of Luotian County. *Jingchu News (Rural Edition)*,(05)
- [12] The government of Luotian county. (2017)Luotian black goat: from "five-in-one" strategy to the profitable Luotian black goat [J], *Chinese rural cooperatives*, ,(8):21-23.(In Chinese).
- [13] Luotian County Bureau of Statistics.(2017).The 2016 statistical bulletin of Luotian County National Economic and Social Development. (In Chinese).
- [14] DouChen.(2017-06-21).Luotian full area tourism boost the accuracy of poverty alleviation. *East Hubei Evening*,(07).
- [15] LiXie.(2017). "Internet +" poverty alleviation "Luotian experience". (In Chinese).
http://www.financialnews.com.cn/ncjr/jrfp/201703/t20170323_114776.html
- [16] ZhengwenHe,JixianLiu.(2017-09-26).National electronic commerce into the rural demonstration county – Luotian. *Huanggang Daily*,(08). (In Chinese).
- [17] FangWang.(2016). Luotian creating a three good situation for poverty alleviation. *Economics*, (12) :31-32.
- [18] JinhuaGuo.(2017).Luotian county rural highway upgrading, benefit the villagers. (In Chinese).
<http://news.hbtv.com.cn/p/1014680.html>.

A Hybrid Model to Analyze Air Pollution Spread Scales in Xi'an and Surrounding Cities

QingZhu^{12}, TingWang¹, Yiqiong Wu¹, JianChai³*

¹International Business School of Shaanxi Normal University, Xi'an, 710000, China

²Institute of Cross-Process Perception and Control, Shaanxi Normal University, Xi'an, 710000, China

³Management and Economic School of Xidian University, Xi'an, 710000, China

Abstract: Air quality analysis and prediction are very important in environmental research as airborne pollution has become a significant health threat, especially in Chinese urban agglomerations. Most previous analysis systems have been based on direct factors, such as pollutant concentrations, wind speeds and direction, relative humidity, and temperature; however, the air quality in a city is also affected by the air quality conditions in surrounding areas. This paper proposes a novel strategy for the analysis and forecast of air quality levels, for which Artificial Neural Networks (ANNs) are employed to elucidate the complex relationships between air quality and meteorological predictor variables. The experimental results in the study demonstrated that the normalized EEMD-ANN model outperformed other models in terms of the Precise, MAE and MAPE. The proposed model, therefore, demonstrated its potential as an administrative tool for issuing air pollution forecasts and for designing suitable abatement strategies.

Keywords: air quality, EEMD, ANN, data mining

1. INTRODUCTION

Since the beginning of the economic reforms associated with the opening-up of China in the late 1970s, China's economy has grown at a remarkable speed; however, in concert with this growth, there has been a rapid rise in air pollution and poor air quality [1] due to the increased use of fossil energy. Air pollution and particularly health damaging air pollution has raised public awareness about the effects of poor air quality in both developing and developed countries. Air quality, which in China is the second worst of 180 countries according to the Environmental Performance Index (EPI) [2], can have severe negative effects on human health, the environment, agricultural quality, and ultimately, the economy. As there has been an increased focus on health in recent decades [3], the citizenry are becoming increasingly concerned about the effects of air pollution on the quality of their daily lives [4]. In fact, particulate matter (PM) or particle pollution has been linked to asthma, cardiovascular and lung diseases, and even premature death, and airborne bacteria and fungi can act as triggers for various respiratory tract diseases and allergies [5]. The World Bank reported that in 2013, air pollution caused 5.5 million deaths worldwide, 1/10 of the total deaths that year (World Bank, 2016) [6] and an earlier study reported that in 2012, around 3.7 million people died as a result of exposure to outdoor air pollution (WHO, 2014) [7] deaths due to outdoor air pollution have been rising and are set to become the main environmental cause of premature deaths by 2050 [8]. Several environmental problems also have been found to have links with PM such as corrosion, soil pollution, and vegetation damage [9]. Global warming due to the greenhouse effect also results in long term consequences [10]. Pleijel found that the present ambient ozone O_3 concentrations have negatively affected wheat grain yields, and to a lesser extent, protein yields (protein mass per unit area) [11], [12]. This relationship between environmental damage and economic development has been widely studied. Voorhees et al. claimed that if the air quality in Shanghai reached the national secondary standard, the monetary value of averted deaths could be as much as 1.7 to 12 billion CNY [13]. Saidi and Hammami also found that there was a negative influence of CO_2 emissions on economic growth [14]; therefore,

air pollution is affecting many aspects of the economy and people's lives.

It has become increasingly necessary to monitor and forecast air pollutants. In 2012, the air quality index (AQI) was proposed as the air quality indicator. Therefore, effective methods are needed to predict AQI and provide early warning systems. The AQI has been difficult to forecast because of its non-stationary chaotic original data, and because forecasting daily/hourly levels of air pollutants is difficult due to the physical and chemical processes involved in the formation, transportation, and elimination of PM [15]. Many studies have focused on air quality and pollutant prediction, but few have examined the geographical distance effect on the nearby target cities. Atakan et.al found that geographical distance influenced forecast accuracy [10]; therefore, as the air quality in a city is affected by pollutant concentrations, wind speed and direction, relative humidity, temperature and the AQI from surrounding areas, it is inefficient to take measures only in the polluted city as the geographical factors need to be added to the pollution scale prediction. If prediction accuracy can be improved, this could assist in the development of government policies for monitoring and improving air quality.

The purpose of this paper, therefore, is to propose a methodology that can measure the air pollution spread in a target city. The remainder of this paper is organized as follows. Section 2 briefly presents the related materials and methods, and in Section 3, the study region, data sources, and experimental design are described. Section 4 analyzes the data and discusses the results, and conclusions and future research directions are given in Section 5.

2. METHODOLOGY

In this section, Empirical mode decomposition (EMD), Ensemble empirical mode decomposition (EEMD), Complementary ensemble empirical mode decomposition (CEEMD) and artificial neural networks (ANN) and related methods are briefly introduced.

2.1 Empirical mode decomposition (EMD)

As air quality has been difficult to forecast because of its non-stationary, chaotic data, many studies have proposed various techniques for forecasting air pollution indicators. Jiang et al. was able to predict $PM_{2.5}$ concentrations using a hybrid method that combined high-dimensional association rules (HDAR), a modified association framework (MAF) with temporal-spatial links, an LVQ network, and an adaptive fuzzy neural network (AFNN) [9]. Conventional methods such as the autoregressive integrated moving average (ARIMA), multiple linear models (MLR), and support vector regression (SVR) are not able to fully capture the information from initial signals [16]; however, Zhu et al. predicted the AQI using a model that combined EMD and SVR [16].

Empirical mode decomposition (EMD), which was proposed by Huang et al., is a self-adaptive signal decomposition method for complex non-linear, non-stationary time series data that can decompose a complicated signal into components of different frequencies and amplitudes called Intrinsic Mode Functions (IMFs) that contain important information from the original series [17]. IMFs have two characteristics that distinguish them from other signals:

- i) The number of extreme values and zero-crossings must differ at most by one.
- ii) At any point, the mean value between the upper and lower envelope is zero.

The EMD formulation is as follows.

$$x(t) = \sum_{i=1}^N M_i(t) + R_N \quad (1)$$

where N is the number of IMFs, $x(t)$ represents the original signal, $M_i(t)$ is the i_{th} IMF component, and R_N is the final residual.

2.2 Ensemble empirical mode decomposition (EEMD)

However, EMD has been found to cause mode mixing, which can distort each IMF generated during the decomposition process. Therefore, new ensemble empirical model decomposition with added white noise has been found to resolve this problem. Ensemble empirical mode decomposition (EEMD) is an ensemble version of EMD that can decompose the copies from the original signal several times and adds different white Gaussian noise to avoid mode mixing. The formulation for EEMD is as follows.

$$x(t) = \frac{1}{I} (\sum_{i=1}^I \sum_{j=1}^N M_j^i(t) + R_N^i) \quad (2)$$

where I is the number of the copies of the original signals, $M_j^i(t)$ is the j_{th} component of the IMFs after i_{th} Gaussian white noise $\varepsilon^i(t)$ is added, and R_N^i is the i_{th} residual.

2.3 Complementary ensemble empirical mode decomposition (CEEMD)

However, as white noise cannot be completely neutralized, some residual white noise remains in the IMF components. Complementary ensemble empirical mode decomposition (CEEMD) is a member of the empirical mode decomposition (EMD) family and can be utilized to remove the noise from original data by adding reverse signals during decomposition [2]. The basic CEEMD structure is the same as for EEMD except that a collection of independent Gaussian white noise and a complementary pair are created to cancel each other out.

$$\varepsilon^i(t) \in \{\varepsilon_+^{\frac{i}{2}}(t), \varepsilon_-^{\frac{i}{2}}(t)\} \quad (3)$$

Subject to, $\varepsilon_+^{\frac{i}{2}}(t) + \varepsilon_-^{\frac{i}{2}}(t) = 0, i \in \{1, \dots, I\}$

This paper employs all three methods to decompose the AQI series.

2.4 Artificial neural networks (ANN)

Artificial neural networks (ANNs) are a mathematical model that simulate the structure and function of the neural network and have been widely used in environmental processes [18], with many studies having predicted PM concentrations using ANNs as they are able to identify the correlated patterns between the input and the objective values through learning, memory, self-adaption, and intelligent processes [19]. ANNs have an input layer for receiving and initiating data, a hidden layer for information encryption and data transfer to the connected nodes in the next layer, and an output layer for the end output. The mathematical model is developed through a training step, after which a prediction step computes the output for a given set of input values using the model created in the training. Chellali et al. proved that ANNs were capable of faster execution speeds and were able to easily work with high-dimension data and capture the complex relationships between the pollutant and with good predictive results [15]. Figure 1 represents the process, the core formulation for which is as follows.

$$h_{w,b}(x) = f(W^T x) = f(\sum_{i=1}^n W_i x_i + b) \quad (4)$$

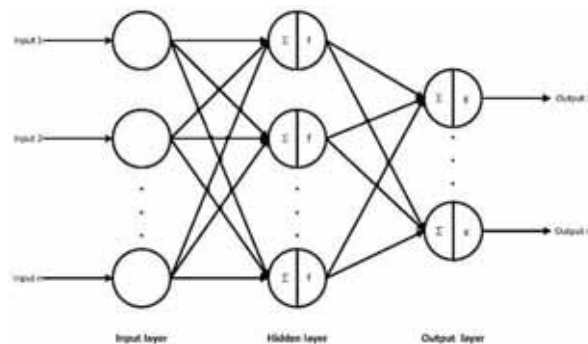


Figure 1. ANN

where f , x_i denote the active functions in the total n nodes in the hidden layers and input variables, W_i is the weight between the two neurons, and b reveals the bias.

Artificial neural networks (ANN) have obvious advantages for air pollutant concentration predictions especially in the non-linear systems traditional models are unable to handle [15]. Chellali et al. and Hurst et al. predicted PM_{10} using an ANN model and proved its success [15], [20].

3. EXPERIMENTAL DESIGN

3.1 Study domains

The general aim of this research is to forecast air quality levels and study the influence on the target city of the air quality from peripheral cities. To do this, eight typical cities in the north of the Qinling mountains; Xi' an, Xianyang, Weinan, Tongchuan, Baoji, SanMenxia, Yan' an and Zhengzhou; were selected as the study sites; all of which have similar geographical locations. The geographical distributions are shown in Figure 2.

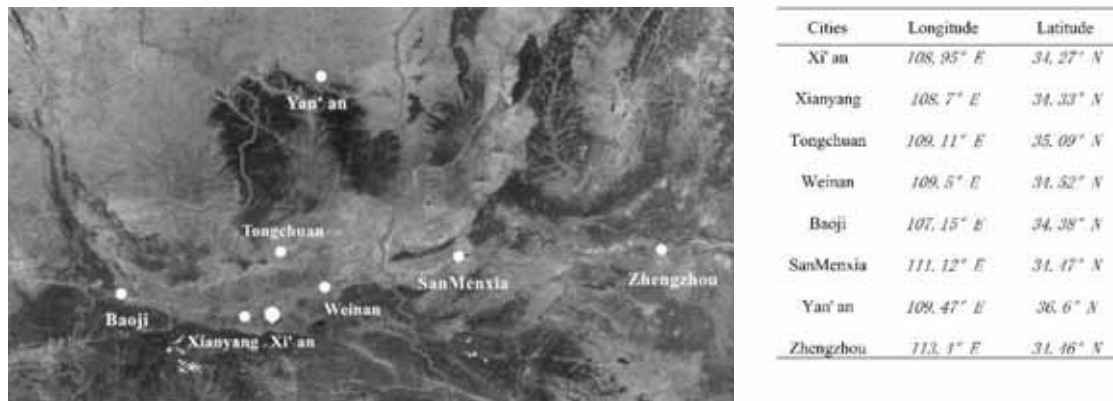


Figure 2. Geographical distribution of the study domains

Xi' an is one of China's great ancient cities; however, it has been experiencing serious air quality problems because of increased industrial development and vehicle emissions. The most affective factor is Xi' an's location under the Qinling Mountains, which prevents winds from the north. The haze from the northwestern areas accumulates above the Guanzhong Plain and is unable to disperse because of the Qinling Mountains. During winter, pollutants from the north are blown into the Xi' an region by the winter monsoon. Xianyang, Tongchuan, and Weinan are respectively located to the west, north and east of Xi' an, with Baoji and SanMenxia further to the east and Yan' an further to the north. As these cities have rapidly growing economies and rising populations because of the industrialization and urbanization, all suffer from airborne pollution. As these cities have similar geographical conditions to Xi' an, the pollutants are unable to easily disperse, worsening the regional haze. Further, increasing demand for electricity has overloaded the regional air loads. Zhengzhou, central city in Henan province is also experiencing rapid industrialization and urbanization and therefore also has a strong demand for electricity, steel, building materials and transport, all of which add to the pollutant emissions, which also exceed regional air loads. In The weather conditions in this region are characterized by high humidity and calm conditions, which prevent pollutant dispersal and exacerbate the regional haze.

These areas were chosen as the study sites for three reasons:

- i) The air pollution is relatively serious but representative of air quality problems in China. In the air quality rankings for China's 74 major cities, Xi'an ranks in the bottom ten [2]; therefore, it is important to be able to scientifically predict air pollutant concentrations.
- ii) As these cities are important, rapidly-developing urban agglomerations in China, they have better monitoring

facilities and adequate available data.

iii) The climatic, meteorological, and geographical conditions in these cities are similar, making it possible to focus on the surrounding cities' influences on the target cities.

3.2 Data set

In each region, the historical daily AQIs were collected from the Ministry of Environmental Protection of the People's Republic of China's platform (<http://datacenter.mep.gov.cn/>). In this study, as the AQI was recently proposed, the earliest data records commence on 1 January, 2014; therefore, historical data from 1 January 2014 to 9 November 2017 in the eight cities was collected. In total there were 1403 data points collected as datasets, as shown in Figure 3. Eighty percent of the datasets in each area were used as the training and validation datasets, while the remaining 20% were the test data. Null data was represented as the mean of two data points from the two nearest days. To improve the accuracy of the results, the three AQI levels were grouped based on six original classes, in which the first level was defined as when the AQI was in a 0 to 100 interval, the second was when the AQI was in a 100 to 200 interval, and the third was when AQI was over 200 the numbers 1, 2 and 3 were used to represent these three levels.

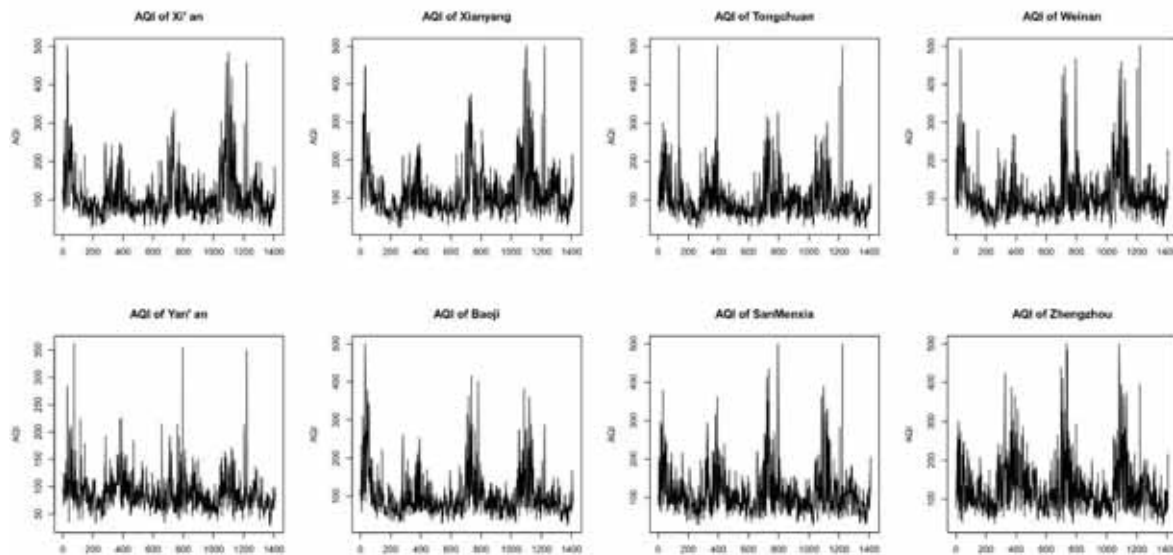


Figure 3. Datasets

In this experiment, Xi'an was deemed to be the center of the urban agglomerations. The study domains or datasets were divided into three groups based on their distances from the center city. The first was the nearest; X-T-W; which represented Xianyang, Tongchuan and Weinan, with the longest radius being 71.88 kilometers. The second was B-S, which represented Baoji and SanMenxia, which had the longest radius of 214.916 kilometers, and the third was Y-Z, which included Yan' an and Zhengzhou, with a 434.011 kilometer-radius.

3.3 Experimental setup

As this paper is focused on the air quality influences of the surrounding cities on the target city, the air quality levels each city were predicted based on the air quality conditions in nearby cities. As this paper is examining the influence in the center city of the surrounding cities, inner factors were not included in the variables. In this section, the air quality conditions in the surrounding cities are used to forecast the air quality conditions in the target city.

There were 5 stages in the experimental process:

- i) Decomposition. The EMD, EEMD, and CEEMD were utilized to decompose the AQIs in each city, after which the obtained IMFs were divided into training datasets and test datasets and bound by the date 30 January 2017 for the training data.
- ii) Normalization. Each IMF was decomposed using the EMD, EEMD, and CEEMD and then normalized within a 0 to 1 interval to serve as the input variables for the ANN training. The original IMFs were also put in the ANN net for comparison.
- iii) ANN training. During the training process, the residuals were eliminated before inputting the normalized IMFs into the ANNs; however, later experiments added these residuals to the ANN for comparison. The three target city levels, 1, 2 and 3, served as the output. In the ANN model, the number of neurons in hidden layer was determined by the number of inputs and there were two hidden layers. The number of neurons in the first hidden layer was the same as the inputs, and in the second hidden layer, one third of the neurons came from the X-T-W condition and a half came from the B-S and Y-Z conditions. To ensure model accuracy and prevent over-fitting, after several trials, the ANN was trained over 1000 repetitions using the Std-Backpropagation weight-updating function to achieve dynamic learning.
- iv) ANN testing. After training, the AQIs from the surrounding cities from 31 January 2017 to 9 December 2017 were used to test the accuracy of the proposed models trained in stage 3.
- v) Experimental results. After decomposition, training, and testing, the results were obtained. All results were analyzed by performance comparison and several error indexes R^2 , Precise, MAE and MAPE; were applied to measure the performances of the different models. The results are discussed in the next section.

4. DATA ANALYSIS AND DISCUSSION

The experimental results are presented in this section. The following three measures were used to analyze the accuracy of the air quality forecasting models; the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the squared correlation coefficient (R^2), which were computed as shown in (5) to (7), as well as Precise, which is the ratio of the number of relevant documents retrieved to the total number of documents retrieved, to measure the accuracy of the retrieval system.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|} \quad (6)$$

$$R^2 = \frac{[\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})]^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2} \quad (7)$$

where y_i ($i=1, \dots, N$) is the observed value, N is the size of the test dataset, and \hat{y}_i is the i_{th} predicted value.

In this section, the overall effect of the proposed strategy is discussed. The air quality in Xi'an, the target city, was forecast from 31 January 2017 to 9 December 2017 using the different models by predicting the air quality levels the surrounding cities. The specific results for R^2 , Precise, MAE and MAPE were calculated for each model, and are shown in Table 3, which clearly shows the superiority of the proposed EEMD-ANN model.

Experiments were also conducted on the nearest X-T-W area using the different models and by adding the decomposed residuals from the original AQI data, the results of which are shown in the first to the fifth rows in Table 1. As can be seen, if the IMFs were not normalized, the R^2 for the EEMD-ANN and EMD-ANN had no value in the R 3.4.2 version, indicating that the models did not work. The comparisons between the groups including and not including the residuals showed that the residuals contributed to accuracy improvements in the EMD-ANN model but decreased the accuracy of the EEMD-ANN model, the forecasting from which was affected by white noise; however, in the CEEMD-ANN model, after the residuals were added, the R^2 was higher

but the errors were also larger.

Table 1. Model Performance. * represents the error index obtained by the models adding the residuals

		EMD-ANN		EEMD-ANN		CEEMD-ANN	
		Normalized	Raw	Normalized	Raw	Normalized	Raw
X-T-W*	R square*	0.626	0.029	0.625	-	0.469	-
	MAE*	0.160	0.631	0.170	0.645	0.287	0.675
	MAPE*	0.090	0.271	0.159	0.276	0.275	0.276
	Train Accuracy*	0.876	0.625	0.862	0.620	0.867	0.620
	Precise*	0.840	0.501	0.830	0.494	0.713	0.494
X-T-W	R square	0.541	0.259	0.626	0.424	0.467	0.424
	MAE	0.209	0.383	0.149	0.238	0.259	0.259
	MAPE	0.109	0.301	0.095	0.168	0.131	0.156
	Train Accuracy	0.873	0.908	0.870	0.909	0.844	0.890
	Precise	0.791	0.617	0.851	0.762	0.741	0.741
B-S	R square	0.481	0.242	0.536	0.366	0.359	0.303
	MAE	0.284	0.369	0.184	0.284	0.284	0.316
	MAPE	0.232	0.280	0.134	0.232	0.147	0.200
	Train Accuracy	0.802	0.876	0.807	0.841	0.800	0.868
	Precise	0.720	0.667	0.816	0.723	0.720	0.684
Y-Z	R square	0.230	0.133	0.247	0.125	0.255	0.278
	MAE	0.351	0.411	0.348	0.433	0.340	0.326
	MAPE	0.223	0.235	0.195	0.246	0.196	0.200
	Train Accuracy	0.738	0.809	0.768	0.821	0.744	0.789
	Precise	0.652	0.610	0.652	0.582	0.670	0.677

After eliminating the residuals from the EMD-ANN model for the normalized X-T-W, the highest R^2 and Precise at 0.541 and 0.791 and the lowest MAE and MAPE at 0.209 and 0.109 were achieved. In the EEMD-ANN model, the MAE and MAPE for the normalized X-T-W were the lowest of all models at 0.149 and 0.095, and the R^2 and Precise were the highest at 0.626 and 0.851, which indicated that the normalized EEMD-ANN model was able to predict relatively accurate air quality levels in Xi' an and could be used by meteorological departments to predict haze and pollution levels. From the two models, X-T-W condition, were more accurate; and for the CEEMD-ANN model, the B-S condition was found to be the same as the previous model. For the X-T-W condition, the MAE and Precise for the normalized models were equal to the models not normalized and the R^2 was higher at 0.467 and the MAPE was lower at 0.131. In the Y-Z condition, the normalized model's Precise were lower but the MAE was higher.

From Table 1, it can be seen that for all models, the normalized models better prevented over-fitting compared to differences between the training set accuracy and the test set accuracy and also had a lower error index and a higher R^2 in most cases, which indicated good model fitness and demonstrated the benefits of using normalization to process the data. Table 1 indicates that the air pollution in Xi' an influences the surrounding cities and that the further the distance, the smaller the influence.

While many studies have achieved accuracies over 90 percent, this paper only achieved accuracies over 85 percent, which may have been because this model did not consider inner variables such as wind speed or humidity. However, the predictions were found to agree with the actual measured data, and the performance of the developed normalized EEMD-ANN model was superior to the performances of the others.

From the error index analysis above, the following conclusions can be made. First, the EEMD-ANN model performed better than the EMD-ANN and CEEMD-ANN models when the data was normalized. Second, the differences in the model error values was possibly related to the changing distances; therefore, it is important to consider geographical factors when using decomposition methods. Third, it is important to select the proper model when seeking to predict different objects for different occasions.

It could be inferred that EEMD is more capable of decomposing signals according to the different physical characteristics, which means that each decomposed series denotes a sort of component with important information of the original signal. That makes the inputs of the ANN model are more representative so as to improve the accuracy compared with the ANN model used alone.

From the above analysis, it can be concluded that the air pollution regional effect increases as regional economic integration progresses. Coal generated electricity, motor vehicles, industrial processes, and dust are the main causes of regional air pollution, the severity of which depends on location, as weak wind, calm conditions and low humidity can prevent flow and cause high regional air pollution.

5. CONCLUSIONS AND FUTURE WORK

If properly designed and implemented, air quality forecasting model are important for air quality management and control. This work developed an air pollution spread scales measuring strategy.

Over the past few decades, China has seen a dramatic increase in airborne pollution, which has become a substantial threat to the environment and resident health. In the Chinese urban agglomerations that have intense industrialization, the air pollution problem is even worse. As diseases, environmental problems, and economic growth have been shown to be associated with air pollution problems, air pollution analysis and prediction have become an important focus in current environmental research.

This paper developed an effective method to measure air pollution spread scales that could be used by the relevant authorities to develop appropriate policies. Because air pollution in one place is affected by the air pollution in surrounding cities, it is necessary to understand the extent of the spread scales to be able to develop cooperative city government mitigation approaches. In this paper, an accurate model was developed; the normalized EEMD-ANN model; to forecast air quality in Xi'an from air pollution information from the surrounding cities.

From a monitoring and modeling point of view, the strategy proposed in this paper has three important applications

- i)* First the normalized EEMD-ANN model outperformed all others models considered in this paper.
- ii)* Second, the proposed models were able to determine the air pollution spread scales (Figure 4) and obtain a valid relationship between the surrounding cities.
- iii)* Finally, this model proved that it could be used by city planners to develop cooperative cross-city policies to manage regional air pollution and reduce the air pollution problems associated with regional economic integration.

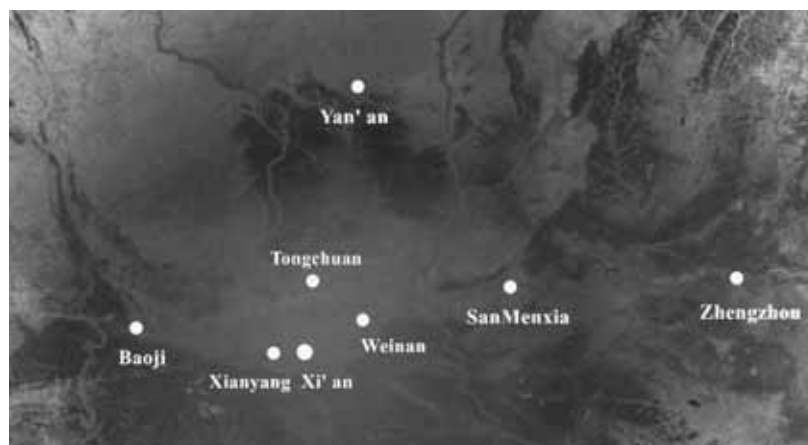


Figure 4. Spread Scales

At the same time, the proposed hybrid model called the normalized EEMD-ANN model can also be used in other conditions as the AQI is applied popularly in almost all cities of China except some small towns. Therefore, although different regions possess various datasets, the model is proper to analyze and predict the air quality considering the strong learning ability of this method.

As pollution content and the long-term effects of regional air pollution need to be considered when managing air quality, in the future, the interactions between the pollutants from the different areas and the long-term effects need to be assessed so as to develop more focused regional air pollution abatement plans.

ACKNOWLEDGEMENT

This research was supported by the Science and Technology Department of Shaanxi Province 2011KRM108, the Chinese Ministry of Science and Technology 2012GXS2D027, the Shaanxi Land Construction Group SXDC 201208.

REFERENCES

- [1] Lu Z, Chen H, Hao Y, Wang J, Song X, Meng M T. (2017). The dynamic relationship between environmental pollution, economic development and public health: evidence from china. *Journal of Cleaner Production*, 166: 134-147.
- [2] Yang Z, Wang J. (2017). A new air quality monitoring and early warning system: air quality assessment and air pollutant concentration prediction. *Environmental Research*, 158: 105-117.
- [3] Deng Z, Liu S. (2017). Understanding consumer health information-seeking behavior from the perspective of the risk perception attitude framework and social support in mobile social media websites. *International Journal of Medical Informatics*, 105: 98-109
- [4] X Zhang X, Liu S, Deng Z, Chen X. (2017). Knowledge sharing motivations in online health communities: a comparative study of health professionals and normal users. *Computers in Human Behavior*, 75: 797-810.
- [5] Jahne M A, Rogers S W, Holsen T M, Grimberg S J, Ramler I P. (2015). Emission and dispersion of bioaerosols from dairy manure application sites: human health risk assessment. *Environmental Science & Technology*, 49(16): 9842-9849
- [6] Ministry of Environmental Protection of the People's Republic of China, (2007). Action Plan on Environment and Health (2007-2015). www.china.org.cn/english/environment/238275.htm (Accessed 25 July 17)
- [7] WHO, 2014. 4. Burden of disease from the joint effects of household and ambient air pollution for 2012. Report WHO (3 p)
- [8] Cariolet J M, Colombert M, Vuillet M, Diab Y. (2018). Assessing the resilience of urban areas to traffic-related air pollution: Application in Greater Paris. *Science of the Total Environment*, 615: 588-596.

- [9] Jiang P, Dong Q, Li P. (2017). A novel hybrid strategy for pm2.5 concentration analysis and prediction. *Journal of Environmental Management*, 196: 443-457.
- [10] Kurt A, Oktay A B. (2010). Forecasting air pollutant indicator levels with geographic models 3days in advance using neural networks. *Expert Systems with Applications*, 37(12): 7986-7992.
- [11] Broberg M C, Feng Z, Xin Y, Pleijel H. (2015). Ozone effects on wheat grain quality - a summary. *Environmental Pollution*, 197(1): 203-213.
- [12] Grünhage L, Pleijel H, Mills G, Bender J, Danielsson H, Lehmann Y, et al. (2012). Updated stomatal flux and flux-effect models for wheat for quantifying effects of ozone on grain yield, grain mass and protein yield. *Environmental Pollution*, 165(6): 147-157
- [13] Voorhees A S, Wang J, Wang C, Zhao B, Wang S, Kan H. (2014). Public health benefits of reducing air pollution in shanghai: a proof-of-concept methodology with application to benmap. *Science of the Total Environment*, 485-486(1): 396-405
- [14] Saidi K, Hammami S. (2015). The impact of energy consumption and CO₂ emissions on economic growth: fresh evidence from dynamic simultaneous-equations models. *Sustainable Cities & Society*, 14(1): 178-186.
- [15] Chellali M R, Abderrahim H, Hamou A, Nebatti A, Janovec J. (2016). Artificial neural network models for prediction of daily fine particulate matter concentrations in algiers. *Environmental Science & Pollution Research*, 23(14): 14008-14017..
- [16] Zhu S, Lian X, Liu H, Hu J, Wang Y, Che J. (2017). Daily air quality index forecasting with hybrid models: a case in china. *Environmental Pollution*. 231: 1232-1244
- [17] Furlaneto D C, Oliveira L S, Menotti D, Cavalcanti G D C. (2017). Bias effect on predicting market trends with emd. *Expert Systems with Applications*, 82(C): 19-26.
- [18] Park S, Kim M, Kim M, Namgung H G, Kim K T, Cho K H, et al. (2017). Predicting pm10 concentration in seoul metropolitan subway stations using artificial neural network (ANN). *Journal of Hazardous Materials*, 341: 75-82.
- [19] Shi B, Wang P, Jiang J, Liu R. (2017). Applying high-frequency surrogate measurements and a wavelet-ANN model to provide early warnings of rapid surface water quality anomalies. *Science of the Total Environment*, 610-611: 1390-1399.
- [20] Hrust L, Klaić Z B, Križan J, Antić O, Hercog P. (2009). Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations. *Atmospheric Environment*, 43(35): 5588-5596.

Fine-grained Aspect Extraction for Online Reviews of E-commerce Products Based on Semi-supervised Learning

Huosong Xia*, Yitai Yang

School of Management, Wuhan Textile University, Wuhan, 430073, China

Abstract: The accuracy of online review mining for e-commerce products is of great value to customer and product matching portrait. Mining the fine-grained aspect in reviews is a key indicator. It can better analyze the emotion tendency of online reviews and understand the advantages and disadvantages of evaluation objects. In this paper, we propose a semi-supervised learning method to extract product aspects and description of aspects. Specifically, we firstly construct word vector space model of large scale reviews with deep learning, then get the list of similar words based on the model. Finally, the fine-grained aspect sets are obtained by classification algorithm. The results of the study show that the efficiency of fine-grained extraction is improved by using semi-supervised method.

Keywords: Fine-grained aspect, car reviews, aspect extraction, deep learning

1. INTRODUCTION

Online reviews are important information sources for business data analysis in e-commerce matching portrait between customer and product. Mining sentiment information in online reviews is of great value for e-commerce. Aspect extraction is a key part of online review mining. Extracting the aspects of users' concern in online reviews can be more effective in mining useful information. As the amount of online review data is growing, it is possible to understand the global view of the evaluation object by analyzing the more finer attribute sentiment. In order to solve this problem, a PageRank algorithm to extract product features has been proposed in Yan et al.^[1], a method with the syntactic parsing and K-means to extract fine-grained aspect has been proposed in Qiu et al.^[2], a LDA method has been proposed to get fine-grained aspect in Jia et al.^[3]. But these studies can only extract the main aspects of evaluating objects for the limitations of adopting methods and the stage of technological development.

Owing to the development and application effect of deep learning which has been generally consented by both academia and industry, this paper proposes a semi-supervised method to get fine-grained aspects in online reviews with deep learning, which aiming at solving the global view and efficiency of this problem. Our method is experimented on car reviews, and compared with the new method in Zhou and Zhang^[4]. The results show that this method can obtain more fine-grained attributes while ensuring higher efficiency. The remaining main structures of the paper are as follows: the second part is related research and theoretical methods, the third part is the fine-grained aspect extraction method of Chinese online review, the fourth part is the experimental results and discussion, the last part is the discussion and conclusion.

2. RELATED WORK

Fine-grained aspect extraction is an important part of online review mining. The methods of extracting fine-grained aspect are divided into two categories: fine-grained aspect extraction based on ontology and fine-grained aspect extraction based on machine learning^[5]. The fine-grained extraction method based on ontology refers to extracting fine-grained based on setting up the concept of evaluation object aspect and the

* Corresponding author. Email: bxxhs@sina.com (Huosong Xia), 994140279@qq.com (Yitai Yang)

relationship between object aspect and itself^[6]. Lau et al. construct a fuzzy domain ontology tree with emotional dictionary and the evaluation object^[7]. Lau et al. propose a method using LDA and Gibbs to analysis fine-grained aspect emotion^[8]. Tang et al propose a method to get fine-grained aspects of product with frequent item algorithm, establish product feature ontology to get feature emotion word pairs, and then obtain implicit product features based on feature emotion word pair^[9].

The fine-grained extraction method based on ontology can only get the main evaluation object attributes, and it is difficult to excavate more fine-grained attributes. Therefore, more researches extract fine-grained attributes using machine learning. Fine-grained attribute extraction methods based on machine learning can be divided into two categories, one is traditional machine learning method, the other is deep learning. In the traditional method of machine learning, extracting fine-grained aspects are mainly based on the rule based learning method. Some methods use building dictionary to get fine-grained aspects, some methods use LDA, CRF and other models to get fine-grained aspects according to contextual information, and some of them use clustering to get fine-grained attribute^[10, 11]. Amplayo et al. propose a model named EBTM to extract fine-grained aspects in short text^[12]. Poria constructs aspect extraction rules with domain knowledge and syntactic dependency to obtain fine-grained aspects^[13]. In contrast with supervised methods, unsupervised methods mainly use clustering methods to extract product attributes.

As the number of online reviews is increasing, the available online reviews data is getting bigger and bigger. The results of extracting fine-grained aspect with deep learning in the large data set are often better than the traditional machine learning methods. Zhou et al. obtain word vectors with deep learning, and then obtain fine-grained aspect based on the AP clustering algorithm and domain knowledge. The method proposed by Zhou et al. has more fine-grained aspects than traditional machine learning methods. But this method still needs after clustering while manually getting obtain fine-grained aspect. Therefore, this paper proposes a semi-supervised method to extract fine-grained aspect. First, we build the word vector model of online reviews with deep learning, and then get the list of similar words based on the word vector model. Finally, we use the classification algorithm to get the fine-grained aspect set automatically.

3. FINE-GRAINED ASPECT EXTRACTION IN CHINESE ONLINE REVIEWS

In this paper, a semi-supervised method is proposed to extract the aspect of the user's attention in the online reviews. Firstly, we construct a word vector model of online reviews with neural network, and then get the list of similar words of the candidate aspect words based on the model. Finally, the words in the list are divided into three categories with a classification algorithm. The three categories are the same aspect words, the fine-grained words and the emotion words. The same words and the fine-grained words are the final fine-grained aspect set. As shown in Fig. 1, the first step is to preprocess the online reviews, it is followed by a neural network model, which is a training word vector model with skip-gram. The second step takes the key aspect words as the seed words, and obtains the similar words list of the key aspect words according to the word vector model. Finally, the words in the list are divided into three categories with a classification algorithm. The three categories are the same aspect words, the fine-grained words and the emotion words. The meaning of words in the same aspect words is the same as key aspect word. The fine-grained words contain the fine-grained aspect words of the key aspect word in the user reviews. The emotion words include the emotion words on the key aspect of the evaluation objects and their fine-grained aspect in reviews.

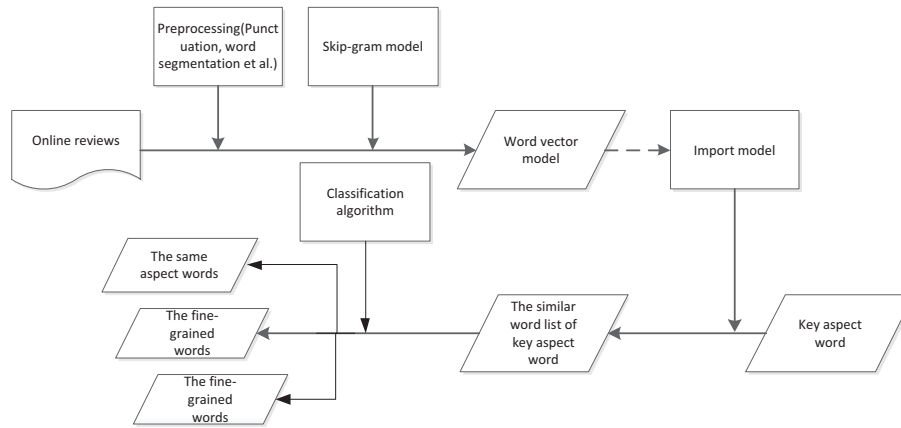


Figure 1. Fine-grained attribute extraction method

3.1 Word vector model

There are many computing methods for text semantic similarity. The way of word vector is to map words as features into word vectors, and then calculate the similarity between words and words by calculating the distance between vectors and vectors. This paper trains word vector model of online reviews by skip-gram neural network. The skip-gram model is a three layer neural networks, it can predict the probability of its adjacent words by each word itself, as shown in Fig. 2.

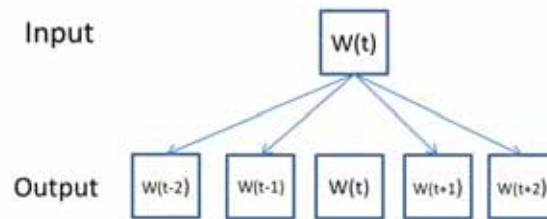


Figure 2. Skip-gram model

Skip-gram model takes the word as features, mines deeper features to get word vector of words by mapping features to K dimensional vector space, and uses cosine similarity to calculate the similarity between the words, the approach can be expressed via Eq.(1). The greater result of $\cos(W_i, W_j)$, the higher the similarity between words. In this paper, the skip-gram model in Word2Vec is used to train online reviews to obtain the word vector model, and each candidate aspect word can get the list of similar words based on the word vector model.

$$\cos(W_i, W_j) = \frac{\sum_{k=1}^N (W_i^k \times W_j^k)}{\sqrt{\sum_{k=1}^N (W_i^k)^2} \times \sqrt{\sum_{k=1}^N (W_j^k)^2}} \quad (1)$$

Here, W_i^k and W_j^k denote the weight of the i -th and j -th word vector in k dimensional respectively. In addition, N denotes the number of word vector dimensions.

3.2 The fine-grained aspect classification algorithm

There are three categories words in the list of similar words. The three categories are the same aspect words, the fine-grained words and the emotion words. The meaning of words in the same aspect words are the same as key aspect words. The fine-grained words contain the fine-grained aspect words of the key aspect word in the user reviews. The emotion words include the emotion words on the key aspect of the evaluation objects and their fine-grained aspect in reviews.

Many online reviews have different appellations for the same aspect because of the high degree of freedom

and randomness. Such as word “内饰” has written in “内室”, this phenomenon of error writing is very common in online reviews. This paper proposes a way to mining the words with same meaning to key aspect word in review. First, words in the list of similar words are transformed into the phonetic alphabet, this is followed by calculating weight of each word based on the Levenshtein distance. Then the words is classified as the same aspect word if the weight of word is less than the threshold. The Levenshtein distance between two words is the minimum number of single-character edits required to change one word into the other one, the edit operations include insertions, deletions and substitutions. The same aspect words algorithm as shown in Table 1.

Table 1. Classification algorithm for the same aspect words

Name: getCateA()
Input: simList, keyword, L //simList is the similar word list of keyword, L is the threshold of distance
Output: cateA //cateA includes words with the same mean of keyword
Init: cateA=∅
Begin:
pyKeyWord = keyword.pinyin() //get pinyin of keyword
for word in simList:
pyWord = word.getpinyin()
simDis = Levenshtein.distance(pyword,pyKeyWord)
if simDis < L:
cateA.add(word)
return cateA
End

Another part of the fine-grained aspect classification algorithm is to obtain the fine-grained words and the emotion words. The idea of the algorithm in this paper is as follows: The first step is to construct an emotional word dictionary based on the emotional words of HowNet and the emotional vocabulary ontology library of Dalian University of Technology^[16]. The second step is to classify words into the emotion words if the words appeared in the emotional word dictionary. The third step, if the word do not appear in the emotional word dictionary, the word is used as the key word to obtain its list of similar words. The score of each word in the list is calculated, the score of word is the similarity weight between the word and it's key aspect word if word is not an emotional word, otherwise the score will be opposite, then score of the word is calculated by sum weight of its list of similar words. Word is classified in the fine-grained words if score is negative, otherwise classified in the emotional words if score is positive. The algorithm to obtain the fine-grained words and the emotional words as shown in Table 2 and Table 3.

Table 2. Recognition algorithm for emotional words

Name: get_emotion()
Input model, keyword, score, deepFlag, emotionDict. //deepFlag is the depth hreshold of recursion, emotionDict is a directory of //emotional words. model is the word vector model. score is the similar score of keyword
Output score //return 1 if keyword is an emotional word otherwise -1 if not
Begin
deepFlag = deepFlag-1
if deepFlag < 0:

```

    return -1*score
if keyword in emotionDict:
    return 1*score
else:
    simList = model.most_similar(keyword,num=10)//get 10 words most similar to keyword and their scores
    for word in simList:
        score = score+get_motion(word,model,emotionDict, score, deepFlag)
    if score <= 0:
        return -1*score
    else:
        emotionDict.add(word)
        return 1*score

```

Table 3. Classification algorithm for the fine-grained words and the emotional words

```

Name: get_cateBC()
Input simList, model, keyword, emotionDict, deepFlag
Output cateB, cateC //cateB includes fine-grained aspect words of keyword, cateC includes emotional words
Init: cateB=∅, cateC=∅
Begin
    For word in simList:
        If get_motion(word,model, emotionDict, word.score, deepFlag)
            cateB.add(word)
        else
            cateC.add(word)
    return cateB, cateC

```

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experimental data

The datasets used in this study were obtained from www.autohome.com.cn. We extracted 815557 Chinese online reviews from the website, and divided the reviews into sentences by punctuation because the reviews are too long. We get 50505086 sentences after preprocessing. Part of data as shown in Table 4

Table 4. Sample of car reviews

Number	Review
1	四门具备一定实用性
2	小众车满足装 X 欲望
3	油耗说实在的都买这排量的车了
4	座椅的包裹性强
5	这种车没有性价比一说

4.2 Experimental procedure and results

Reviews are divided into sentences by punctuation in our method. It is followed by segmentation with Jieba, a tool to separate a Chinese sentence into words. And then the skip-gram model in Word2Vec is used to train online reviews to obtain the word vector model. In this paper, the results of fine-grained aspect words are extracted from two key aspect words, “内饰”(Interior) and “外观”(Appearance), the fine-grained attribute extraction method proposed by Zhou Qingqing et al. is used as a contrast^[4]. Part results of fine-grained aspects are shown in Table 5.

Table 5. Part results of fine-grained aspects

Aspect	Method	Fine-grained aspect	Emotion word
Interior	Our method	'内设', '内适', '黑内', '总体设计', '用料', '棕黑', '风格', '做工', '配色', '料子', '前 中控'...	'大气', '简简单单', '简洁', '时尚', '素雅', '赖看', '新潮', '不拉风', '前卫', '大大方方', '上档次'
		'内设', '内适', '木纹', '黑内', '总体设计', '用料', '做工', '配色', '料子', '材料', '前 中控'	∅
	Compared method	'外怪', '外感', '外款', '屁股', '很靓', '前脸', '红颜色', '车脸', '咖版', '外表', '大胡子', '侧身'...	'犀利', '大气', '帅气', '可爱', '不娘', '时尚', '赖看', '霸气', '漂亮', '大气磅礴', '高大威武'
		'外怪', '外感', '外款', '屁股', '很靓', '前脸', '红颜色', '车脸', '咖版', '外表', '大胡子', '侧身'...	∅
Appearance	Our method		
	Compared method		

4.3 Experimental analysis

As the data set used in this article is large, it is costly to annotate for all online reviews, so this article builds the aspect set as follows: We search for twenty videos which introduce or recommend cars, and treat the description words of a car in the video as a fine-grained aspect when it is used to introduce the interior and appearance details. In the final result, the fine-grained aspect set of the interior contains 47 attributes, with 39 attributes in the appearance, and part result of aspects as shown in Table 6.

Table 6. Part result of aspects of interior and appearance

Aspect	Fine-grained aspect
Interior	'前中控', '材料', '造型', '质地', '工作台', '中控台', '布局'
Appearance	'屁股', '前脸', '侧身', '侧脸', '外形', '流线形', '线条'

This paper uses recall score, accuracy score and F1 value to evaluate the results of fine-grained properties, the approach can be expressed via Eqs. (2)–(4):

$$P = \frac{TP}{TP+FP} \quad (2)$$

$$R = \frac{TP}{TP+FN} \quad (3)$$

$$F_1 = \frac{2P \cdot R}{P + R} \quad (4)$$

Here, TP denotes the number of extracted aspects correctly, FP denotes the number of extracted aspects inaccurately, FN denotes the number of non-extracted aspects, P denotes accuracy score, R denotes recall scores. The evaluation results are as shown in Table 7.

Table 7. The evaluation results on car reviews

	Accuracy		Recall		F_1	
	Our method	Compared method	Our method	Compared method	Our method	Compared method
Interior	78.6%	81.1%	93.6%	91.4%	85.4%	85.9%
Appearance	77.8%	83.3%	89.7%	89.7%	83.3%	86.3%

In Table 7, the accuracy and F_1 of our method is lower than the compared method. The reason is that our method obtains fine-grained aspect automatically by a classification algorithm, when the fine-grained aspect sets are obtained, but the compared method uses the artificially selected method to filter the noise vocabulary and reduces the number of the aspect of the error extraction. The recall scores in Table 7 show that our method is better than the compared method, the result demonstrates that our method can be more comprehensive to extract the fine-grained aspect from online reviews. In contrast with compared method, the results of two methods are in the same level, but our method extracts fine-grained aspect with more efficiency by a semi-supervised method.

5. CONCLUSIONS

With a heavier demand on global view of e-commerce matching portrait between customer and product, this paper proposes a method to extract fine-grained with technical background of the development of deep learning. Our method improved the problem of efficiency and granulation in the existing research. Firstly we construct a word vector model of online reviews with neural network, and then get the list of similar words of the candidate aspect words based on the model. Finally, the words in the list are divided into three categories with a classification algorithm. The three categories are the same aspect words, the fine-grained words and the emotion words. The same words and the fine-grained words are the final fine-grained aspect set. In contrast with the traditional method based on seed word or LDA, this method can obtain more meaningful fine-grained attributes. Our method improves the efficiency of fine-grained extraction with a semi-supervised approach, In contrast with a similar new methods in literature^[4].

However, this method also has some shortcomings. Although we can get more fine-grained aspects, some of the emotional words are divided into the final fine-grained aspect set because of noise. The future work will reduce noise in fine-grained aspect extraction with unsupervised or semi-supervised algorithms. At the same time, it can expand to the more common e-commerce product reviews mining in the application area, promote the pixel of the global view of matching portrait between customer and product, and achieve the precise mining.

ACKNOWLEDGEMENTS

This research has been supported by the National Natural Science Foundation of China (71571139); We deeply appreciate the suggestions from fellow members of Xia's project team and Research center of Enterprise Decision Support, Key Research Institute of Humanities and Social Sciences in Universities of Hu Bei Province(DSS20150215 &DSS20150108).

REFERENCES

- [1] Yan Z, Xing M, Zhang D, Ma B. (2015). Exprs: an extended pagerank method for product feature extraction from online consumer reviews. *Information & Management*, 52(7), 850-858.
- [2] Qiu Y, Ni X, Shao L. (2015). Research on extracting method of commodities implicit opinion targets. *Computer Engineering and Applications*, 51(19):114-118(in Chinese).
- [3] Jia W, Zhang H, Yang C, Zhao X, Li Bo. (2016). User sentiment model oriented to product attribute. *Journal of Computer Applications*, 36(1):175-180(in Chinese)
- [4] Zhou Q, Zhang C. (2017). Fine-grained aspect extraction from online customer reviews. *Journal of the China Society for Scientific and Technical Information*. 36(5):484-493(in Chinese)
- [5] Tang X, Liu G. (2017). Research review on fine-grained sentiment analysis. *Library and Information Service*, 61(5):132-140(in Chinese)
- [6] Silveira Chaves M, Laurel A., Sacramento N, Drebes Pedron C. (2014). Fine-grained analysis of aspects, sentiments and types of attitudes in restaurant reviews. *Encontros Científicos - Tourism & Management Studies*, 10(1): 66-72.
- [7] Liu L, Nie X, Wang H. (2012). Toward a fuzzy domain sentiment ontology tree for sentiment analysis. *International Congress on Image and Signal Processing*, IEEE, 1620-1624.
- [8] Lau R Y K, Li C, Liao S S Y. (2014). Social analytics: learning fuzzy product ontologies for aspect-oriented sentiment analysis. *Decision Support Systems*, 65(5):80-94.
- [9] Tang X, Lan Y. (2016). Sentiment analysis of microblog product reviews based on feature ontology. *Library and Information Service*, 60(16):121-127,136(in Chinese)
- [10] Sängner M, Leser U, Klinger R. (2017). Fine-Grained Opinion Mining from Mobile App Reviews with Word Embedding Features. *Natural Language Processing and Information Systems*, pp3-14.
- [11] Chen Z Y, Mukherjee A, Liu B. (2014). Aspect extraction with automated prior knowledge learning. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 347-358.
- [12] Amplayo R K, Song M. (2017). An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews. *Data & Knowledge Engineering*, 110(3):54-57.
- [13] Poria S, Cambria E, Ku L. (2014). A rule-based approach to aspect extraction from product reviews. *Proceedings of the 2nd Workshop on Natural Language Processing for Social Media*, 28-37
- [14] Suleman K, Vechtomova O. (2015). Discovering aspects of online consumer reviews. *Journal of Information Science*, 42(4):492-506.
- [15] Zhao L, Huang M, Sun J, Luo H, Yang X, Zhu X. (2015). Sentiment extraction by leveraging aspect-opinion association structure, *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*. New York: ACM. 343-352
- [16] Xu L, Lin H, Pan Y, Ren H, Chen J. (2008). Constructing the Affective Lexicon Ontology, *Journal of the China Society for Scientific and Technical Information*. 27(2):180-185(in Chinese)

The Relationship Between Investor Sentiment and Stock Market Volatility: Based on the VAR Model

Ge Zhang^{1,2}, Jishun Wang¹, Hao Guo^{1}, Xin Zhang¹*

¹School of Management Science and Engineering, Shandong University of Finance and Economics,
China

²School of Information, Capital University of Economics and Business, China

Abstract: Using web crawling technology crawls investors' comments of SANY stock(Stock Code: 600031) and Fujian Expressway stock(Stock Code: 600033) from February 11, 2015 to August 16, 2017. Then using semi-supervised machine learning method construct investor sentiment index. Moreover, collecting the daily closing stock price and trading volume data from Qianlong software explore the relationship between investor sentiment and stock market volatility based on VAR model and Granger Test Method. The results show that the rate of return and trading volume have a two-way Granger causality, while negative emotion and the rate of return have a one-way Granger causality. Furthermore, with the impulse response function and variance decomposition, the results show that trading volume has significant effects on rate of return and negative emotions of investors have significant negative effects on rate of return and trading volume.

Keywords: big data, machine learning, investor sentiment, VAR model, stock market volatility

1. INTRODUCTION

In the domestic stock market, there still exist some unreasonable factors, which caused the stock market to rise and fall and bring huge losses to countless investors^[1]. Furthermore, the irrational factors of investors increase the uncertainty of investors' decision-making. The traditional standard finance cannot explain the investor's irrational factors influence on the investor well, while the behavior finance explains the irrational behavior of the investor better from the angle of investor's behavior and the psychology of generating such behavior. Investor sentiment as a representative of irrational factors, studying the relationship between investor sentiment and stock market volatility is very important. Domestic and foreign scholars have also made lots of achievements in the field of investor sentiment using traditional financial data^{[1]-[5]}.

Recently, with the development and applications of Computer Technology, investors have more opportunities and channels to comment on a particular stock on Internet financial platforms, such as guba.eastmoney.com and guba.sina.com. Those comments are more accurate in reflecting investors' emotional tendencies and gradually becoming an important source of investor sentiment initial data. So how to make better use the comment data and study its role in the stock market has become a focus for many scholars.

This paper uses web crawling technology to collect investors' comments, and quantify the comments to construct investor sentiment index which is an important concept index in finance, then study the interaction between investor sentiment and stock market volatility. The remainder of the paper proceeds as follows. Section 2 introduces the literature review. Section 3 presents the research design. Section 4 analyzes the results. And the last section provides some conclusions and prospect.

2. LITERATURE REVIEW

Many scholars have taken into account the fact that the presence of a large number of emotional-driven investors can lead to price deviations from fundamental value as early as 1936. Research on investor sentiment

* Corresponding author. Email: 422140921@qq.com(Hao Guo)

begins in 1990s, Barberis N etc.^[2] aimed at two defects: the empirical model can not effectively response to price information and stock price on the news over reaction, presented a thrifty model based on investor sentiment to study how to form the investors' expectations of future earnings. After 21th century, the research on investor sentiment was gradually increasing, Lee W Y^[3] studied the role of investor sentiment in stock market volatility and excess earnings, and concluded that the change of mood was negatively correlated with market volatility and was positively correlated with excess returns. The study of Schmeling M etc.^[4] found that emotion has a negative impact on the average returns of stock market in countries. Jiang Yumei etc.^[5] discussed the overall and cross-sectional effects of investor sentiment on stock returns. Wang Chun etc.^[6] using GARCH-M model to study the influence of investor sentiment on stock market returns and volatility found that there is a positive feedback effect between investor sentiment and stock market returns. Hence, we can know that the impact of investor sentiment on stock returns has been agreed by scholars, but the research conclusions based on different emotional measurement indicators are differently in direction and extent.

Investor sentiment measures can be divided into three parts by foreign scholars: direct, indirect and meta methods^[7]. Since meta method is a hybrid version of direct method and indirect method, this paper adopts the viewpoint that there are two methods: direct method and indirect method, which are accepted by most domestic scholars, to measure investor sentiment. The first method based on investors' survey data about future market trends, e.g. American Association of individual investors Index, CCTV system index, HaoDan Index. Based on above index, Xiong Wei etc.^[8] adopt the HaoDan Index analyzing the dynamic relationship between stock trait volatility, stock returns and investor sentiment from the perspective of theory and practice. Zhang Ziqiong etc.^[9] using the investor opinion survey data of Sina Finance and Economics stock column calculated the investors' sentiment, then took the stock of the Shanghai 180 index as the research object to carry out an empirical study. Although emotional data which was gotten though this method needs less to process, and was convenient to use, the cost of obtaining is high and it is hard to process. Currently, the research on investor sentiment is mainly based on the second methods, the proxy index of which includes CEFD, retail selling and buying stocks ratio, CCI^[10]. Liu Weiqi etc.^[11] select the monthly new accounts of individuals and institutions as individual and institutional investors' emotional proxy indicators, making a comparative study of the two effects to determine the role of the two emotions in the market.

With the popularization of big data technology, some scholars use crawler program to grab investor' comments on Internet platform, get investor sentiment index by quantification, and then study the relationship between investor sentiment and stock market. Lai Kaisheng etc.^[12] grabbed comments from Sina Weibo, forming Weibo emotion comprehensive index through a series of processes, and discussed the relationship between emotion and stock market. Shi Yong etc.^[13] based on the relevant module data of excellent mining financial platform, the investor attention degree, snowball investor attention degree, news attention degree and news sentiment index were constructed respectively, and used correlation and VAR model to explore the relationship between these indicators and the CSI 300 Index.

In summary, the research hot spot of investor sentiment has not been reduced, and the measurement method has been developed more scientific with the development of technologies. With the literature review, the result shows that using big data technology to build investor sentiment is still in the initial stage. The papers are not common by using big data machine learning to quantify comments as investor sentiment indicators to study related issues. Therefore, this study intends to use the big data machine learning method to build investor sentiment, regarding www.eastmoney.com as data sources, to study the relationship between investor sentiment and stock market volatility based VAR model.

3. RESEARCH DESIGN

3.1 Model introduction

Through verification of the initial data found that the VAR model can explain the relationship between each variable and its own lag better. In VAR model, each endogenous variable in the system is regarded as a function of the lag value of all the endogenous variables in the system to construct the model. At the same time, it has less requirements for the characteristics of variables, and could better describe the dynamic behavior in the fields of economy, finance and so on, and it has a good prediction effect. The model will be established in this paper as follows:

$$Y_t = \sum_{i=1}^p A_i Y_{t-i} + \sum_{j=0}^r B_j X_{t-j} + \varepsilon_t, t = 1, 2, \dots, n \quad (1)$$

Where Y_t is a k dimensional endogenous variable, Y_{t-i} ($i = 1, 2, \dots, p$) is a lag endogenous variable vector. X_{t-i} ($i = 0, 1, \dots, r$) is the lag order of a d dimensional exogenous variable or a lag exogenous variable vector, p, r , respectively, is the lag order of the endogenous variable and the exogenous variable. A_i is $k \times k$ dimension matrix, B_j is $k \times d$ dimension matrix, ε_t is a vector of the k dimensional perturbation term.

3.2 Indicator description

Yao Jun etc.^[14] pointed out that the volatility of the stock market could be measured by the daily returns of high frequency data. Li Dan etc.^[15] drawn an conclusion that the trading volume was positively related to stock market volatility based on the theory of mixed distribution hypothesis. Therefore, this study selects the rate of returns(R), trading volume(V) as proxy indicators for market volatility, and selects the positive investor sentiment (PE) and the negative emotion (PN) as proxy indicators for investor sentiment. The formula for calculating R we adopted as follows:

$$r_t = \ln(close_t) - \ln(close_{t-1}) \quad (2)$$

where, $close_t$ represent the closing price of current trading day.

As for investor sentiment, using big data platform get the index of PE and NE. The processing flow of the big data platform is shown as follows:

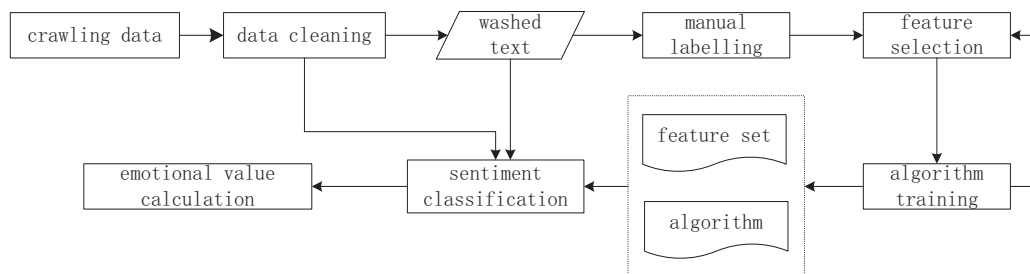


Figure 1. The process and function of the investor sentiment big data platform

3.3 Data source and processing

The data of V and closing price of stock market come from Qianlong software which relying on its products be a pioneer in domestic securities industry. Qianlong software has a long history and contains a variety of time sharing data and technical data.

Dongfang Wealth has the largest access and is the most active website in China's financial and economic securities portal, according to iResearch. Therefore this article takes www.eastmoney.com as the source of initial emotional data. In order to ensure the representative, reliable of data, two stocks (including 600031, 600033) are randomly selected. Using the Python language to write crawl program based on the Pycharm environment crawled comment contents, comment time, amount to be read, amount of comment, commentator's message of two stocks from February 11, 2015 to August 16, 2017, finally crawled 90342 comments. The process of

crawling comments is divided into the following steps :

Step1: data cleaning. In this step, the main task is cleaning the third party institutions' comments and some other messages. After data cleaning, the remaining comments have 87693 pieces of valid.

Step2: sentiment classification. If there is no emotion classifier, an emotion classifier needs to be formed by a series of processes: (1)Inviting 5 investors with many years of stock investment experience manually mark 10,296 valid comments on 600033 stocks. Then, (2)electing of characteristic words has great influence on the accuracy rate of the machine learning algorithm, so it's particular important to choose the right feature words. Next, (3)The training set and test set are selected according to the tagged text and its features. (4)Finally, the emotion classifier is formed according to algorithm and the corresponding feature set, then the 600031 comments are classified.

Step3: emotional value calculation. The emotional index is calculated by using the ratio of the number of comments on one day's positive evaluation (negative evaluation) and the total number of comments on the day, and the emotional index is calculated to get the daily emotional index. Computational formula are as follows:

$$posindex_t = \frac{poscount_t}{total_t} \quad (3)$$

$$negindex_t = \frac{negcount_t}{total_t} \quad (4)$$

Through above methods, the data of each proxy index is obtained, but in order to make every index data with lateral comparability and feasibility, the index value needs to be standardized by Range Standardization Method. At the same time, considering the practical significance of the positive and negative value of the index, the calculation method of the positive index is as follows^[16]:

$$x_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \quad (5)$$

When the index is negative, the method of calculation is:

$$x_{ij} = \frac{\max(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})} \quad (6)$$

4. EMPIRICAL ANALYSIS

4.1 Unit root test

After processing of each indicator, the trend of the four indicators are shown in figure 2:

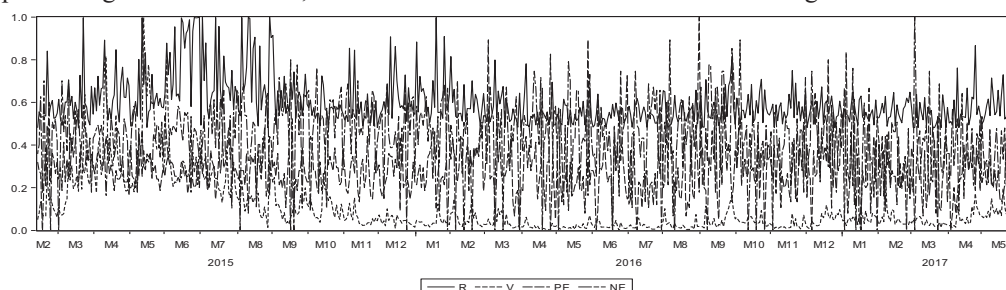


Figure 2. Trend chart of four indicator data

The reliability of the established VAR model depends on the stability of the variables. Stationary time series can be directly constructed non-constraint VAR model. Conversely, non-stationary time series need to check by cointegration test whether it exists a long-term equilibrium relationship between variables. If there is a long-term equilibrium relationship, the vector error correction model could be constructed, otherwise, it is necessary to make the difference of each variable into a stable variable. For further analyze the model, Eviews8

software is used to test the stability of each variable by unit root test. The result as table 1 shows:

Table 1. The result of unit root test

variable	type of models	DF-GLS	ADF	result
R	containing constant term, excluding trend	-2.345**	-6.8236***	stable
V	containing constant term, excluding trend	-3.5491***	-3.805***	stable
PE	containing constant term, excluding trend	-21.8361***	-22.0543***	stable
NE	containing constant term, excluding trend	-6.5702***	-7.7514***	stable

***represents a significant level of confidence at 1%, **represents a significant level of confidence at 5%.

The results of unit root test show that the test statistic values in above four time series are all less than the critical values under the 1% test level, so the four sequences do not contain unit roots, that is, all variables are $I(0)$ stationary variables.

4.2 Construction of VAR model

In this study, four variables are used as endogenous variables to establish VAR model. Through “Lag Structure” function in “Lag Length Criteria”, select lag number 7 to carry on analysis, then get table 2.

Table 2. Optimal delay order

Lag	LogL	LR	FPE	AIC	SC	HQ
0	1404.545	NA	1.04e-07	-4.723591	-4.694012	-4.712071
1	1840.797	865.1471	2.53e-08	-6.140967	-5.993069*	-6.083363
2	1890.219	97.34414	2.26e-08	-6.253689	-5.987472	-6.150002
3	1932.550	82.80667	2.07e-08	-6.342497	-5.957960	-6.192726*
4	1952.167	38.10910	2.04e-08	-6.354695	-5.851840	-6.158841
5	1989.987	72.96080*	1.90e-08*	-6.428286*	-5.807112	-6.186349
6	1999.969	19.12167	1.94e-08	-6.407988	-5.668495	-6.119967
7	2009.692	18.49655	1.98e-08	-6.386821	-5.529009	-6.052717

Clearly, the optimal delay order is 5 according to five criteria for determining the optimal delay order. Then, VAR(5) model is established. In order to prove whether VAR(5) model is stable, we make an AR root graph. From the result of AR root graph, as figure 3 indicates, all points are in the unit circle, which is to say the model is stable.

4.3 Granger causality test

After the VAR model passes the stability test, it needs to test the result of the model estimate using granger method. The final result is shown in Table 3.

Table 3. Granger causality test

Dependent variable: R				Dependent variable: V			
Excluded	Chi-sq	df	Prob.	Excluded	Chi-sq	df	Prob.
V	11.2535	5	0.0466	R	13.0327	5	0.0231
PE	3.3291	5	0.6494	PE	0.1514	5	0.9996
NE	10.6068	5	0.0498	NE	2.58051	5	0.7643
All	27.0212	15	0.0286	All	17.1353	15	0.3108

From the data in Table 3, under the test level of 5%, V and R are mutual Granger reasons. This relationship confirms the phenomenon of stock market. No matter the rise and fall of trading volume and stock price are the actual reflection of supply demand and mapping of investors' psychological behavior. NE is the Granger reason of R. The reason is that when the investor's mood is low, it would affect the act of selling and buying stocks, and

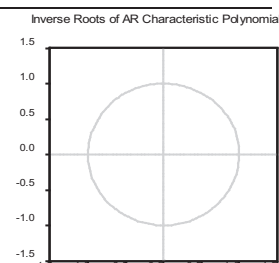


Figure 3. Stability text of VAR model

then affect the supply and demand relationship of the stock market, finally affect R. Endogenous variable R is jointly significant relative to the V, PE and NE lag terms, which indicates the lagged variables of V, PE, NE have strong influence on the R. Similarly, the endogenous variable PE is significantly associated with the lagged variables of R, V and NE.

4.4 Pulse response

In order to understand the dynamic characteristics of VAR model, the impulse response function can be used to analyze it. Impulse response function is used to measure the impact of a standard deviation impulse on the current and future values of all endogenous variables in the VAR model. That is to consider how the effect of the residual term spreads to other variables. The real line, in the graph 4(unit: day, lag interval set 30 periods), represents the impulse response function, and the dashed line represents the deviation zone of positive and negative double standard deviation.

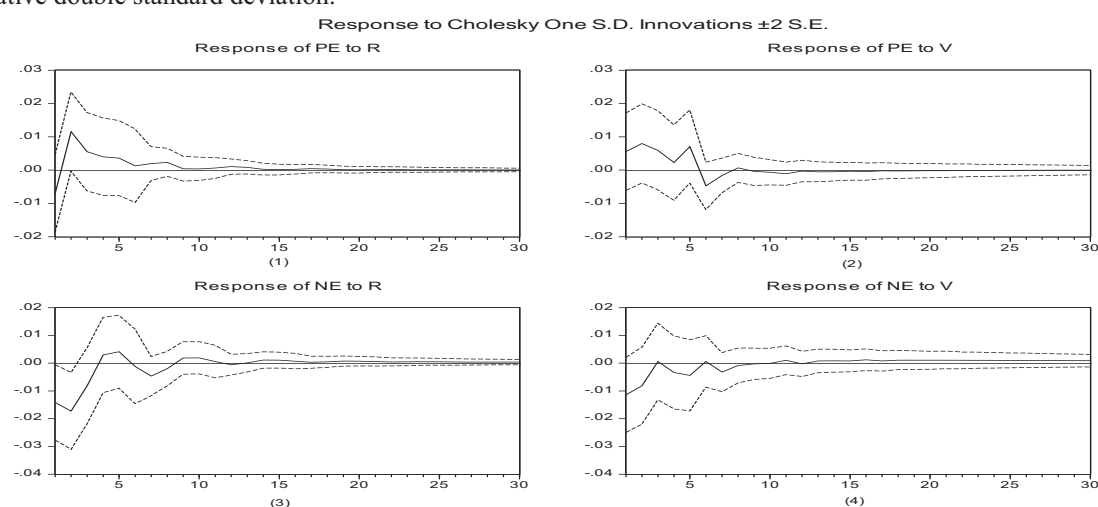


Figure 4. Pulse response of R, V to PE, NE

Figure 4(1) displays that PE will increase in the short term and then return to quiet gradually with the rate of return increasing. Figure 4(3) shows that when R is high, NE decreases in the first two period, but then increases again. It suggests the reason for NE is more complex, while the increase of R is the main reason for PE, which is more clear than the reason for NE. At the same time we find that NE has a greater impact on R and lasts longer.

Figure 4(2) the response of PE to a standard deviation impacting on V is also timely, the value of the response in the first period is about 0.007, and then gradually decreases with the period increasing. From figure 4(4), we can find that the NE in response to a standard deviation shock of V in the first period is -0.11. In a few fluctuations period, it tends to be around 0. The above two cases indicate that the short-term increasement of V will cause emotional differentiation, both positive and negative emotions will increase. Then after going through a large fluctuation, PE return calm.

4.5 Variance decomposition

The impulse function can only analyze the response of some variables to the disturbance of one variable, but can not analyze the importance of exogenous variables to endogenous variables. Therefore, variance decomposition is needed to analyze the contribution of each structural shock to the change of endogenous variables (measured by variance). Figure 6 shows the results of variance decomposition of R and V in the 30 period of delay. The transverse axis represents the number of lag periods, and the longitudinal axis represents the index contribution.

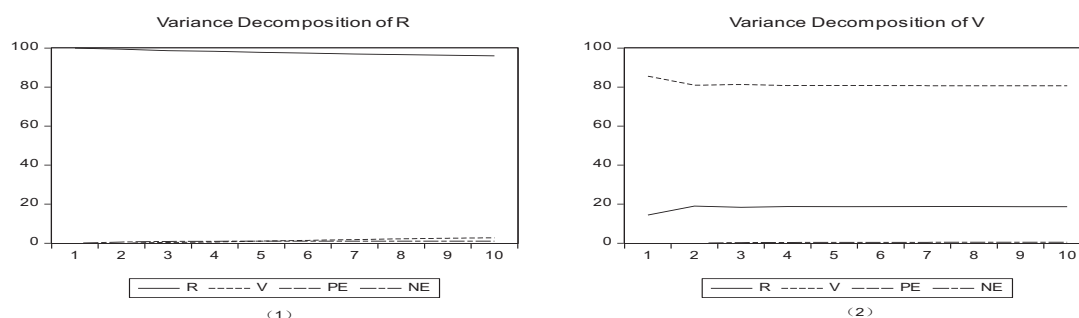


Figure 5. Variance decomposition of R, V index

From Figure 5 (1) we found the contribution of return to itself is 100%, then it decreases gradually in second period while the contribution of trading volume, positive emotion and negative emotion increase gradually. After the fifth period, V, PE and NE's contribution to return arrived 2%, 1%, 3%, respectively. This suggests that the contribution of V, NE to R is significant. From Figure 5 (2), R is highly contribute to V in the first five period at around 23%, PE and NE contribution to the V is insignificant. This phenomenon is consistent with the China stock market which do not allow investors to sell short. When positive emotions are high, on the one hand, there will be a phenomenon of reluctant to sell, on the other hand, the risk will increase, and investors who are willing to buy stocks at a high level will also be reduced causing positive emotions to have little contribution to volume.

5. CONCLUSIONS

Through the method of machine learning construct the investor sentiment index, we studied the mutual influence of investor sentiment and stock market volatility. The results show that there is a two-way Granger causality between trading volume and returns and the negative emotion and the rate of return have one-way Granger causality. Through the impulse response function we find the trading volume to the rate of return is positive, while the response of rate of return to volume is negative.

The main contribution of this paper is that using machine learning method to construct investor sentiment index, under the background of big data, combined with VAR model of econometrics to study the relationship between investor sentiment and stock market volatility. The results confirm the conclusions of previous scholars' research, indicating that the method of constructing investor sentiment index is feasible and effective. However, there are several points in this paper that need to be continued in the future research.

1. In this paper, big data technique is used to analyze stock comments, quantifying the text into emotional values. In the process of research, emotional values may deviate from actual values due to the subjectivity of manual annotation. There is a certain impact on the results of the research. In the future research, the comments of investors will be dealt with more persuasively. For example, more senior stock investors and stock researchers will be invited to manually annotate comments to form a more effective training set.

2. From the results of this paper, we can know that the positive and negative emotions have a significant impact on the stock returns and trading volume in different ways. The next step is to expand the sample size, adopt the data of different stock market and more stocks.

3. Though the impulse response function find the change of positive emotions and negative emotions are not the relationship of one is rising and the other is falling, sometimes it can also be strengthened or weakened at the same time. In the future research, sentiment of investors will be subdivided more detailedly to explore the stock market investors emotional impact on stock market volatility.

ACKNOWLEDGEMENT

This research was supported by the Natural Science Foundation of Shandong Province under Grant ZR2012GQ006 and the the National Social Science Foundation of China under Grant 17BGL198 and 16AJY003.

REFERENCES

- [1] Lai Kaisheng, Chen Hao, Le Guoan. (2014). Can mood predict the stock market? *Advances in Psychological Science*. 22(11):170-1781.
- [2] Barberis N, Shleifer A, Vishny R W. (1997). A Model of Investor Sentiment. *Journal of Financial Economics*. 49(3):307-343.
- [3] Lee W Y, Jiang C X, Infro D C. (2002). Stock Market Volatility, Excess Returns, and the Role of Investor Sentiment. *Journal of Banking & Finance*. 26(12):2277-2299.
- [4] Schmeling M. (2009). Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*. 16(3): 394-408.
- [5] Jiang Yumei, Wang Mingzhao. (2010). Investor sentiment and stock return : an empirical study of overall effect and cross-sectional effect. *Nankai Business Review*. 13(3):150-160(in Chinese).
- [6] Wang Chun. (2014-9). The influence of investor sentiment on the returns and volatility of the stock market-An Empirical Study Based on the net inflow of open stock fund money.*Chinese Journal of Management Science*. 22(9):49-56(in Chinese).
- [7] Kim T, Ha A. (2010). Investor Sentiment and Market Anomalies. 23rd Australasian Finance and Banking Conference 2010 Paper.
- [8] Xiong Wei, Chen Langnan. (2015-9). Stock Price Volatility , Stock Returns and Investor Sentiment. *Journal of Management Science*. 28(5):106-115(in Chinese).
- [9] Zhang Ziqiong, Qian Guoming, Li Yijun. (2013-7). An empirical study on the interaction between investor sentiment and Chinese stock market based on online voting. *Soft Science*. 27(7): 123-127(in Chinese).
- [10] Zhang Qiang, Yang Shue, Yang Hong. (2007-7). An empirical study on investor sentiment and stock returns in Chinese stock market. *Systems Engineering*. 25(7): 13-17(in Chinese).
- [11] Liu Weiqi, Liu Xinxin. Individual and institutional investor sentiment and stock returns: a study based on Shanghai Stock Market. *Journal of Management Sciences in China*. 17(3):70-87(in Chinese).
- [12] Lai Kaisheng, Chen Hao. (2014-5). Weibo emotion and Chinese stock market : Based on Cointegration Analysis. *Journal of Systems Science and Mathematical Sciences*. 34(5): 565-575(in Chinese).
- [13] Shi Yong, Tang Jing, Guo Kun. (2017-7). The concern of social media investors and investor sentiment influence on the Chinese stock market. *Journal of Central University of Finance & Economics*. 45-53(in Chinese).
- [14] Yao Jun, Yin Kairui, Li Qi. (2017-11). The impact of Macro fundamentals and institutional investors on the volatility of stock market : a review and prospect. *Modernization of Management*. (6): 118-121(in Chinese).
- [15] Wen Xi, Wang Guoshun. (2015-3). Research on the dynamic relationship between liquidity and returns of fuel oil futures market. *Systems Engineering*. 33(3): 38-44(in Chinese).

E-commerce Service and Supply Chain Finance in Rural China: A Value Captured Perspective

Yi Chen

Center for Fiscal and Financial Development in Haixi, Minjiang University, Fuzhou, China

Abstract: From a value captured perspective, we aim to examine e-commerce service facing the problem of a lack of supply chain finance. Based on multiple case studies, we explore how to capture value via e-commerce service efficiently. Using the data from a randomly sampled questionnaire survey in rural China, we would address the value co-creation process within supply chain finance. The e-commerce industry is in need to collaborate and combine with supply chain finance in order to maintain the value transformation.

Keywords: e-commerce, supply chain finance, value captured

1. INTRODUCTION

In terms of market development, China's urban e-commerce market is approaching to the saturation after recent years of explosion and some e-commerce giant's groups have already turned towards the rural area for new market exploring. In 2016, the number of online shoppers in less developed areas in China has reached 77.1 million, which resulted in 40.6% increase against the number of 16.9% in urban areas ^[1]. E-commerce industry are experiencing a booming stage and has gradually become the new growth driven. However, some rural inland counties still face the severe challenge of accessing internet, service and supply chain finance, etc.

Government sectors have released related policies and documents, together with financial support of 20 million RMB per perished county for five years to accelerate the development procedure. With the comparatively lower income level, infrastructure deficiency and dispersed logistics, majority of less developed counties are the main reasons which rural areas have failed to attract investors and brick-and-mortar retailers whereas the limited products' variety and inflated prices also influenced demand of rural consumers ^[2].

In view of the above analysis, this paper engages with the discussion around e-commerce service to address two research questions and represents the following theoretical contributions: (1) What's the role of supply chain finance in e-commerce service during the value co-creation process? (2) How does e-commerce service capture values?

The research would address the value co-creation process within supply chain finance. During the field works and investigations, service entities were found to be the value connectors who take responsibility for integrating customers' demand, delivering political announcements, transforming one-time financial support into long-term value creation opportunities and precisely match all aspects of information. It is an extremely comprehensive process which requires the high level of operating skilled labors, the large volume of data collection, professional data analyzing and sufficient financial fund.

2. THEORETICAL BACKGROUND

Researchers have examined the E-commerce service and supply chain finance along two dimensions: value networks and service supply chain. Scholars in the field of supply chain management have brought the concept of supply chain network structure with direct interactions and indirect interactions of suppliers and customers ^[3].

Lusch, Vargo and Tanniru believe that the process of marketing and supply chain management will converge with value network, which is embedded in S-D logic^[4]. The value network connects loosely coupled value proposing actors to interact with institutions and technology to co-create values and exchange service offerings. It contains multiple supply chains where firms with competitive relationships often have the same suppliers and result in forming an even larger and more encompassing net. Through establishing a highly-structured network economy, organizations which seem unrelated to each other can obtain the opportunity to build connections^[5]. Therefore, supply chains in network might become more adaptable, fluid and dynamic.

Achrol and Kotler have also stated that only vertically integrated hierarchy with knowledge enriched and turbulent environment is not enough for firms to grow since they are overcommitted to upstream and downstream technologies^[6]. Instead, value network provides a wider stage with more complex situations for companies to learn and adapt, which may also encourage and urge them to become more critical about survival and growth.

However, when firms adopt the perspective of value network, supply chain as the traditional focal construct have not been replaced but enhanced by better integration between customer and marketing^[7]. In general, supply is a product-centric or firm-centric concept, whereas value is more likely to be customer- and service-centric. Based on S-D logic, customer is the only entity who can co-create value and assess value by using market offerings practically in reality. Other researchers hold different perspectives where company exist as a collection of commerce services, which can be integrated to satisfy consumer's needs^[8]. The term of value has also been considered as the connection between suppliers and customers^[9].

Concluded by Burton, et al.^[10], firms make appropriate value from rest of actors in value network and thus create tensions that can offset servitization efforts or even destroy values for all participants. Tensions are reviewed between each actor in network, as well as groups and individuals within those actors. The study identified five broad groups of tensions: direct challenge to expertise, pressure to learn, cost-focused challenges, process-change risk Aversion and external actor impacts on value creation and from which they can be categorized as internal tensions (e.g. threats to personal dominance of activity) and external tensions (e.g. regulatory issue). Most of tensions operate at the micro-environmental level, restricted to firms' servitization value network. In reaction towards the potential issues, actors might invest in new resources, build stronger relationships or share a wider service vision across the network. Although the study was based on manufacturing industry, the findings can still offer constructive suggestions to e-commerce value network buildings.

In sum, the theories above are complementary to each other, forming an explanation framework for theorising about e-commerce service in China. Based on the theoretical frame, explanations of e-commerce service require to illuminate the mechanism of it.

3. METHODOLOGY

3.1 Case study methods and sampling

The paper adopted a multiple case study methods since the existing researches on county-level e-commerce service platform are still in the early stage and few theories can become supportive. Conger^[11] also stated that quantitative research itself is insufficient to investigate the phenomenon which consist of multiple levels with dynamic and symbolic components, especially in studying the links among behavioural, interpersonal, organizational and environmental levels. Moreover, multiple cases strengthen the result by replicating the patterns, therefore improve the robustness of findings. With the aim of addressing a conceptual framework or theory of value co-creation model of e-commerce service platforms within the niche area. Patton^[12] has conducted further theories that there are no rules of determining the size of the sample and the validity, meaningfulness and insights have deeper relationships with the analytical capabilities.

Our unit of analysis is value co-creation process of the case study companies. By combining data from four SMEs which specialized as e-commerce service providers in various counties, the study attempted to extract the unknown added and marginal value creation theory. SMEs tend to build contract with the third-party service providers more often since they may not have sufficient resource and scale to operate their own sectors while strategic plans are in need. Having values created through diversified channels proves that the e-commerce ecosystem has entered the more mature stage and that the online commerce service model has gradually stepped into offline model (O2O). The investigations of each case growth path have justified the assumption as well, from which most of e-service corporation have attempted to take control of several crucial products' supply chains operations; the selection of the cases reflects the diversity of the counties in China: All CITY (Nanchang), SOWIN (Lishui), WENYUAN (Linan), GANJIE (Suichang) and LECUNTAO (Taiyuan).

Established in, 2014, ALLCITY E-commerce Group is a comprehensive county-level e-commerce service provider with 139 current employees in Nanchang, Jiangxi. SOWIN Group has established a subsidiary company called Lishui Rural E-commerce Co.Ltd. with local government with the aim of offering public service in 2013, tendering and bidding for subject of project operations. WENYUAN was established in 2009 and has already founded 26 branches in China. Under the operation of specialized rural ecosystem model, the company has contributed to 18 counties and 5 impoverished villages' economic and cultural development. GANJIE was the first comprehensive e-service company in China and started operating from 2010. Until 2015, under control of 88 local county operation stations cross China, 6000 service centres were set up in each village providing public service, data collection and goods distribution. LECUNTAO Group is the only one that located in North-western area during the investigation among the five cases and all the other companies are distributed in South-eastern area.

All the companies are different in size, functions and growth history. Beside the core commerce unit of e-service platform, the investigations have also covered at least one commerce unit within the ecosystem circle, which include government department, individual e-commerce companies, traditional companies and third-party online platforms.

3.2 Data collection

The research has conducted 15 interviews working at all the levels across the 5 cases, which including a) board directors, commerce unit director, customer service director, creative directors, quality managers and branch managers from e-service platforms; b) county level government departments and public associations' managers; c) individual production companies; d) e-commerce corporations who involved in the value creation and exchange relationships.

Data collection procedures adopted qualitative triangulation techniques. Denzin has suggested that triangulation is highly convincible in terms of combining data sources to study the same social phenomenon^[13]. Instruments included face to face interviews, private seminars, telephone semi-structured open ended interviews, company documents and archival data from credible organizations and websites. All interviews were recorded, translated and transcribed from Chinese to English for the interviews organized in oral mandarin.

3.3 Data analysis

Both within case analysis and cross case analysis were generated for the 5 cases. Within case analysis is a form of pattern recognition, where emerged themes can be categorized and incorporated with data-drive inductive approach of Boyatzis^[14]. Therefore, we might conclude a pattern of information that organizes and explains the possible observations to the minimum and interprets aspects of phenomenon to the maximum. The themes emerged from within case analysis then might form a broader concept for synthesis in cross-case

analysis. Integrations of data then start collapsing, merging and filtering themes that best represented the value creation model, which leave the stand-alone themes for further discussion.

4. FINDINGS

4.1 Value Missed for being lack of supply chain finance

In general, the missed values here refers to the value creation opportunities which are not realistic to be created with a specific company's current situation or have been left out when formulating operation strategies. However, in theory all the miss value opportunities can effectively contribute to value co-creations and are still possible to be utilized when the corporations become mature enough to enter the next stage^[15]. All the cases under evaluation are facing the same issue of acquiring sufficient financial funding or capital raising sources to a certain extent, from which LECUNTAO have the least problem. After analysing the information, the research found that proper implicated strategies and the large scale the company has covered might be benefited from its start-up strategy that focused on taking control of supply chains finance. As for innovative company like GANJIE, who originated the e-commerce service firm, has contrived multiple projects at the early stage and could be able to generate sufficient support in forms of crowd funding.

4.2 Cross-case Value Co-Creation Analysis

In general, the missed values here refers to the value creation opportunities which are not realistic to be created with a specific company's current situation or have been left out when formulating operation strategies. However, in theory all the miss value opportunities can effectively contribute to value co-creations and are still possible to be utilized when the corporations become mature enough to enter the next stage. All the cases under evaluation are facing the same issue of acquiring sufficient financial funding or capital raising sources to a certain extent, from which LECUNTAO have the least problem. After analysing the information, the research found that proper implicated strategies and the large scale the company has covered might be benefited from its start-up strategy that focused on taking control of both uplink and downlink supply chains management. As for innovative company like GANJIE, who originated the e-business service firm, has contrived multiple projects at the early stage and could be able to generate sufficient support in forms of crowdfunding.

Table 1 Cross-Case comparison of value missed and destroyed

Value Co-creation Corporations	Missed Value	Destroyed Value
ALLCITY	Manufactured products downlink; Supply chain finance	Cold-chain raw food supply
WENYUAN	Supply chain financet, Cross boarder E-commerce	Manufactured products downlink;
SOWIN	Manufactured products downlink; Marketization operation	Social network marketing; B2C retailing
GANJIE	Crowd funding; Supply chain finance	Goods dominant logic value system; Manufactured products management
LECUNTAO	Supply chain financet; Cold-chain transportation	B2C dominant agricultural products uplink; Systematically public training program

4.3 Core Value Captured by adopting supply chain finance

There are three innovative value co-creation approaches concluded by adopting supply chains finance: online to offline (O2O) commerce model, social network and big data analytics in terms of disadvantaged group alleviation

The value might be delivered and transferred to both side. On the one side, service providers might receive extensive brand recognition and reputations which go along with actual financial support to practice the program. Funds are then converted into special training plans; household hardware facilities support or even available paid jobs afterwards. Instead of direct limited financial support, the service provided towards those disadvantaged groups have benefited wider range of people and offered them the chance to live a self-sufficient life which also deepen social value creation and result in a more stable social environment. On the other side, vulnerable groups finally have the chance to realize self-fulfilment—which is considered as the highest-level value creation from individual's view. Abilities and skill will be the most valuable tool to sustain their life without attaching to others and all the benefit might partially reflected from monthly income and actual transactions.

Our analysis shows that value co-creation from big data occurs not from the data itself or individual data scientists, but mainly through the proper methods of data management, where companies are able to contextualize, experiment and execute data on a regular base. Therefore, the e-commerce service providers play the significant role along value creation chains as who might be capable of extracting delivering valuable information to upstream-downstream industries.

Value added and personalized service is recommended through data analyzing, such as, account management, online store design and transactions preference to assist sellers and public sectors.

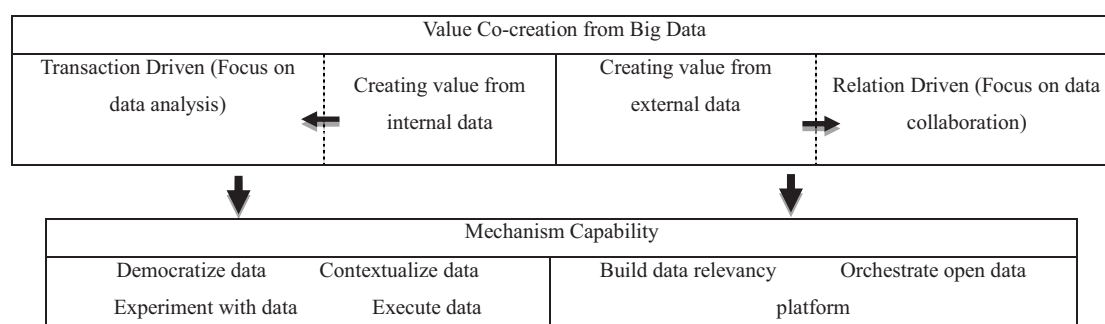


Figure 1. Value Co-creation Procedure from Big Data

Thirdly, the data generated might assist firms to understand supplier or customer's needs. The utilization of external data network mainly refers to higher co-created value through collaborations in comparison to the value harnessed through internal sources. It is all about mixing and matching, different trends and patterns are discovered follow various orders and with the aim of grow the whole market, all the service platforms can create synergistic bundles and achieve higher value premiums.

Base on above finding, we propose that: First, the less dependent of e-commerce service entities on public and government sectors, the more opportunities of value creation the company might have. Second, most financial value captured by service providers is created from controlling supply chains and business extension to offline real economic sectors. Third, the more information and big data resource an internet service entity controls, the better opportunity it may have to create more values in both terms of volume and quality.

5. DISCUSSION

5.1 The less dependent of e-commerce service entities on public and government sectors, the more opportunities of value creation the company might have.

To further illustrate the assessments from findings, four propositions are concluded for theory-building from which the research findings are extracted and purified. Value is a comprehensive concept that could be created in many various forms and increasing or diminishing along the transformation. The study has generated the complemented version of the original conceptual framework from which three propositions are concluded in the following paragraphs.

The first proposition will be developed and concluded from the participant's perspective of value creation procedure with each internet service company. According to the findings in Section 4, some of the service platforms operate as “the third-hand” of local government and public sectors, instead of operating as a market-oriented enterprise. The strategies of those companies intended to rely more on government financing or public funding which then being transferred into different forms of values and benefit target groups live in less developed areas, such as knowledge and skills' training, individual store' s opening, e-commerce related job offering, etc.

It can be seen from the cross-case analysis table above that government and public institutions have been the main participant during the general value co-creation procedure, as the investor and supporter for disadvantaged groups alleviation and online to offline business extensions, where values are transferred from centralized economic values towards decentralized tangible and intangible values. Meanwhile, it is the value receiver with the implication of big data analysis, who obtain the transferred database or statistics and then reuse them to promulgate updated versions of policies and regulations, thus consist the benignant circle for sustainable value co-creations. However, with the relatively tight regulatory control and manipulations, the value co-creation process is limited to a certain amount and involvements of other organizations are restricted due to potential political and social obstacles.

In terms of rural banking institutions effect on value' s generating process, a comparatively heavier proportion of government supervision will probably result in less rural commercial institutions' financing which could be given to e-commerce start-ups or service entities. It might be plausible under a circumstance that with the strong interference and control from local government, the overall efficiency is reduced through tedious processes with related administrative functions. Typically, the above estimations are currently only applicable to rural financial institutions rather than large state-banks. With further evaluation, our study has reached the same conclusion. Among the five cases, LECUNTAO and SOWIN Group are the only two companies that started with comparatively independent value creation strategy, which result in generating relatively larger value than other three firms within the same time span. Among the four key values captured, value co-creation through disadvantaged group alleviation has the most intimate connection with public sectors, where ALLCITY and WENYUAN have devoted the most into. Consequently, the theoretical inference precisely matches with data analyzing results. Therefore, we may propose that value co-creation is a process which requires connections through multiple channels rather than focusing too much on national public support and be “the third-hand” of government sector.

5.2 Most financial value captured by service providers is created from controlling products' supply chains and business extension to offline real economic sectors.

Taking control of products' supply chains and other offline real economic sectors may realize marketers' long-term objectives from creating social values by building brand awareness to the financial values reflected by actual payments transfer. Through the process, retailers and e-commerce platforms will have the ability to ‘close

the loop' on marketing investment and obtain a more accurate return on investment (ROI) on customer' s retention and acquisition, especially by collaborating with service providers in a multi-channel environment. With 54% of all the transactions were made over mobile applications, value creations through social networks are proved to be one of the main trend in the future.

With the purpose of strengthening value generating process, the study concludes three key points to become further constructions. First of all, the interaction between online and offline business operations. It has been found that the most often appeared missed value is the value could be produced with supply chain management of the products which service platforms are in responsible to promote or build their own brand with. Companies such as SOWIN, ALLCITY and LECUNTAO have focused relatively more on products offline management. They trace the production to the source to ensure the overall quality, adjust logistics strategies and improved its efficiency, together with online promoting or marketing plans. During the process, interaction and fluent contacts with each entity is significant to ensure values can be sustained transferred to next link. The interaction seems to be the key to integrating O2O industrial chain with the main objective of matching with customer' s needs and create a more comfortable environment, which indirectly accelerates the value creation process.

Secondly, only surface and low technical services are not enough to maintain the operation, instead, the more comprehensive business model could be developed, such as house rental service, socializing car rental services and news media sharing services (Zhang, 2011). Flexible profit mode might also contribute to the higher efficiency. The user-oriented fees might be split between customers and merchants, depending on specific conditions for a win-win situation. As both previous academic papers and our research findings have shown the similar results, the proposition 2 here might become a useful view for companies to refer to.

6. CONCLUSIONS

The research has addressed the value co-creation process within supply chain finance. During the field works and investigations, service entities were found to be the value connectors who take responsibility for integrating customers' demand, delivering political announcements, transforming one-time financial support into long-term value creation opportunities and precisely match all aspects of information. It is an extremely comprehensive process which requires the high level of operating skilled labors, the large volume of data collection, professional data analyzing and sufficient financial fund. All the requirements are compulsory to be involved in the process of generating useful values and ensure those values are effectively itinerant.

The e-commerce industry is in need to collaborate and combine with supply chain finance in order to maintain the value transformation. Together with a rational and efficient implication of information processing systems, higher volume and quality of values could be created.

ACKNOWLEDGEMENT

This research was supported by the Fujian Provincial Department of Science & Technology under Grant 2017R0081.

REFERENCES

- [1] China Internet Network Information Center (CINIC). (2017) Statistical Report on Internet Development in China. Available from: <https://cnnic.com.cn/IDR/ReportDownloads/201706/P020170608523740585924.pdf> ,2017-02-01.
- [2] Organisation for Economic Co-operation and Development.(2013). The Role and Measurement of Quality in Competition Analysis. Retrieved from: <http://www.oecd.org/competition/Quality-in-competition-analysis-2013.pdf>.

- [3] Lambert, D. M., Cooper, M. C. and Pagh, J. D.(1998). Supply chain management: implementation issues and research. *International Journal of Logistics Management*, 9(2): 1-19.
- [4] Lusch, R., Vargo, S. and Tanniru, M. (2010). Service, Value Networks and Learning. *Academy of Marketing Science*, 38:19-31.
- [5] Flint, D. J., & Mentzer, J. T. (2006). Striving for integrated value chain management given a service-dominant. *The Service-Dominant Logic of Marketing: Dialog, Debate, and Directions* Armonk, NY: ME Sharpe, 139-149.
- [6] Achrol, R. S., & Kotler, P. (1999). Marketing in the network economy. *The Journal of Marketing*, 146-163.
- [7] Coe, N. M., Dicken, P., & Hess, M. (2008). Global production networks: realizing the potential. *Journal of economic geography*, 8(3): 271-295..
- [8] Zhao, J. L., Tanniru, M., & Zhang, L. J. (2007). Services computing as the foundation of enterprise agility: Overview of recent advances and introduction to the special issue. *Information Systems Frontiers*, 9(1): 1-8.
- [9] Vargo, S. L., & Lusch, R. F. (2004). The four service marketing myths: remnants of a goods-based, manufacturing model. *Journal of service research*, 6(4): 324-335.
- [10] Burton, J., Story, V., Zolkiewski, J., Raddats, C., Baines, T. S., & Medway, D. (2016). Identifying Tensions in the Servitized Value Chain: If servitization is to be successful, servitizing firms must address the tensions the process creates in their value network. *Research-Technology Management*, 59(5): 38-47.
- [11] Conger, J. A. (1998). Qualitative research as the cornerstone methodology for understanding leadership. *The Leadership Quarterly*, 9(1):107-122.
- [12] Patton, M.D. (2001). *Qualitative research and evaluation methods*, Sage Publications. *Proceedings of the First International Conference on Knowledge Capture*, Victoria, Canada, 60-67.
- [13] Denzin, N. K. (1973). *The research act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.
- [14] Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. sage, 1998, 28(1): 210-214.
- [15] Bocken, N., Short, S., Rana, P., & Evans, S. (2013). A value mapping tool for sustainable business modelling. *Corporate Governance*, 13(5): 482-497.

Research on Zipf's Law of Hot Events in Search Engines

XU Ying-fan^{1}, SHI Ming-liang¹*

¹School of Business, Sichuan Agricultural University, China

Abstract: This paper focuses on the amount of searching and browsing of hot events in China and finds that the searching index sequences of daily hot events and weekly hot events are in line with Zipf's law. Through continuous collection of large data samples of multiple dates, We find that the Zipf index of the searching index series for daily hot events fluctuates in a very small range. Through Zipf analysis, we find that only a few events maintain long-term heat. A few events will be the focus of most people, while a few will focus on some directional events. So Zipf distribution describes the balance of economic propensity of sender and receiver during the transmission of information. This research is of some reference to commercial activities that make use of hot events for e-commerce.

Keywords: Zipf's law, hot events, searching sequence, Zipf distribution

1. INTRODUCTION

Right now, we have entered an era of rapid development of the Internet, and the Internet has become the main way for people to access a large amount of information. As a kind of creative technology, Internet technology innovation has brought great changes to all aspects of economic society^[1]. According to "Internet Moore's Law", the amount of Internet information is growing at an extremely high rate, the factors of production tend to be virtualized, and a new factor of production is produced - big data^[2].

Network hot events arise spontaneously, which are formed by social members in accordance with specific logic and value requirements^[3]. It has the characteristics of fast propagation, strong diffusion, strong interactivity and strong linkage with the real society^[4]. The Internet-based hot event is the relatively fixed part of the updated network information, which can present the important events, the focus and the direction of public opinion in the Internet^[5]. At the same time, through the statistics of big data samples, it also accords with a phenomenon of herd behavior in the network hot events discovered by Cass R. Sunstein, which provides value guidance for our business activities in e-commerce^[6].

So, as one of the most important web information retrieval tools, the search engine has developed rapidly. According to its characteristics of fast retrieval speed and high accuracy, Zipf analysis will be used to study the Zipf distribution of search index sequences of single-day and multi-day hot events and the reason of the fluctuation of Zipf index on different dates^[7].

2. CONTENTS OF ZIPF'S LAW

Zipf's law was proposed by George K. Zipf, a professor of linguistics at Harvard University in 1948, who conducted a large number of statistics on the frequency of occurrences of words in English documents to test the quantitative formulas of predecessors^[8,9]. Zipf's law is mainly used in natural language courses, its content is: If the frequency of occurrence of each word in a given article is counted, each word is arranged in decreasing order of frequency of occurrence as a sequence, and each word after arrangement is labeled with a natural number increasing from 1. With r that serial number, with $f(r)$ that frequency, there is the following power law holds:

$$f(r) \sim r^{-\alpha} \quad (2-1)$$

* Corresponding author. Email: xu850765216@163.con(XU Ying-fan)

In the formula, the index α is called the Zipf index. In the Zipf distribution, the index α is a positive constant, which depends on the distribution and has no relation with other parameters. In most countries' languages, the index $\alpha \approx 1$. It shows that in most countries, only a few words are frequently used, while most words are rarely used.

At present, Zipf's law is not only used in the analysis of the phenomenon of natural language, but also used in the research of petroleum price, town scale, biological engineering, medicine and other fields. In recent years, Zipf's law has also started to play its role in the field of finance and e-commerce. For example, through Zipf's law, N. Vandewalle, M. Ausloos analyzed the stock price index^[10], Y. Fujiwara analyzed bank bankruptcy and its causes^[11].

3. ZIPF DISTRIBUTION OF HOT EVENTS IN SEARCH ENGINES

In view of the fact that the main way for people to access information is to search the Internet at the present time, and the retrieval of hot events needs the help of search engines, it is very important to study the general law of frequency of hot events in search engines. By using the ideas and methods of metrology, we regard all the information that can be retrieved by search engine as a whole, take the search index of hot events that are positively correlated with the popularity of hot events as the research object, and explore the Zipf distribution of the search index sequence of hot events.

Because Zipf's law is universally applicable to the distribution of word frequency in natural languages, we adopt the strategy of changing data statistics or adjust the statistical objects to verify the statistical results of Zipf distribution, so that we can study Zipf's law in more depth and have more meaningful expansion.

Then, if the search index of each hot event in search engines is counted, the search index of each hot event is arranged in decreasing order of frequency of occurrence as a sequence, and each hot event after arrangement is labeled with a natural number increasing from 1. We use r to denote the sequence number of the search index sequence, $f(r)$ the search index, and β as the Zipf index for this distribution, which should be satisfied:

$$f(r) \sim r^{-\beta} \quad (3-1)$$

In order to make the relationship more intuitive, we simply transform the above formula. After taking the logarithm of the above formula, we obtain:

$$\lg f(r) = -\beta \lg r + C \quad (3-2)$$

In the above formula, C is a constant. It is easy to see that the slope of a straight line in logarithmic coordinates is the Zipf index.

This paper selects the hot events and their search indexes in search engines on December 29, 2017 as the data set and arranges them according to decreasing order of search indexes.

Table 1. Relevant data on hot events in search engines on December 29, 2017

Hot events	S/N r	Index $f(r)$	$f(r) * r$	$\lg r$	$\lg f(r)$
Twenty-eight vegetarian dishes on the wedding	1	343999	343999	0.000000	5.536557
Digging out Eight-Diagram tactics when refurbishing	2	333818	667636	0.301030	5.523510
Shared boyfriends appeared in Haikou	3	295163	885489	0.477121	5.470062
Cashing out by Ant-Check-Later was sentenced	4	227910	911640	0.602060	5.357763
A rich second generation hurt people with the death reprieve	5	224724	1123620	0.698970	5.351649
Netizens bumped into Jay Chou	6	212413	1274478	0.778151	5.327181
Ran Yingying was exposed speculation	7	149435	1046045	0.845098	5.174452

The crowd ticketed the police car	8	125267	1002136	0.903090	5.097837
The employee was fired because of sick dozing	9	109429	984861	0.954243	5.039132
Qianbao-Net Zhang Xiaolei surrendered	10	103238	1032380	1.000000	5.013840
.....
Treasury reverse repurchase	47	9296	436912	1.672098	3.968296
Pan Yueming take selfies in the background	48	9179	440592	1.681241	3.962795
Netizens bumped into Yue Yunpeng	49	8751	428799	1.690196	3.942058
Six people were sentenced to death in Bahrain	50	8394	419700	1.698970	3.923969

In the statistical analysis, I treat each set of data as a separate observation. This approach not only reduces the risk of contingency or asymmetry in large data studies, but also allows me to explore the considerable within-study variation in Zipf estimates^[12]. More on this below.

Looking at the term $f(r) * r$ in Table 1, we find that the data in a certain section of the middle is stable within a certain range of values, that is, there is a section with stability and the section gradually decreases to both sides. These two phenomena show that the data we collected and studied basically satisfy Zipf's law^[13].

Next, we describe the collected data and its logical relationships in figures, which are more intuitive. In the following figure, we use $\lg r$ as the abscissa and $\lg f(r)$ as the ordinate. Based on the data shown in Table 1, we can predict that, except for the beginning part, the points in the figure should approximate a straight line^[14].

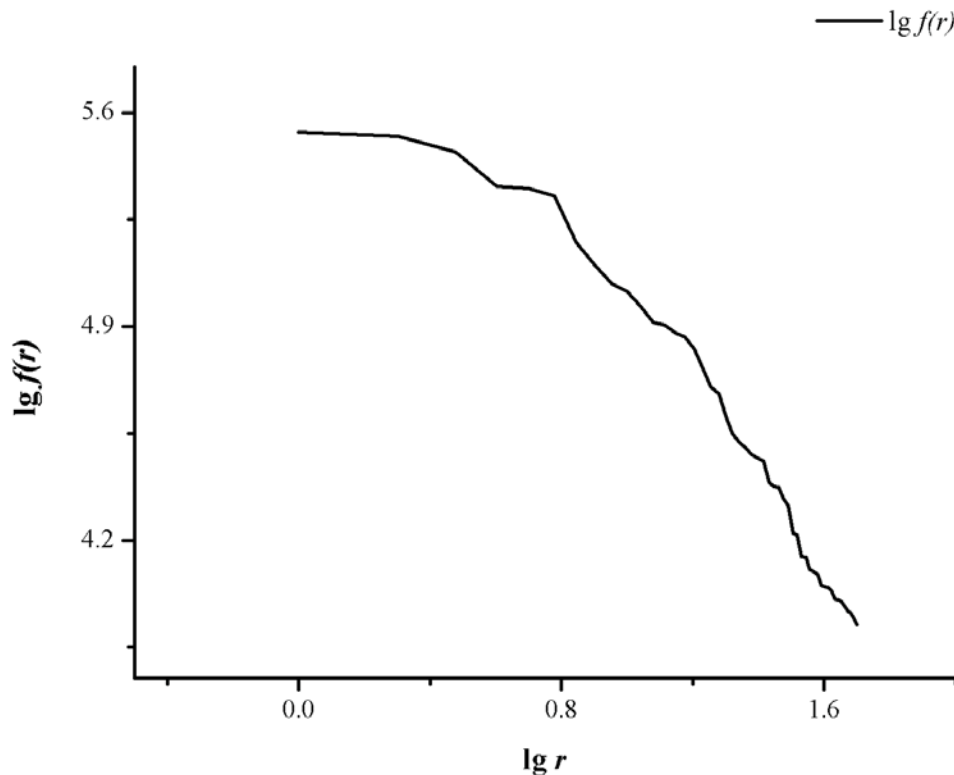


Figure 1. Relevant data on hot events in search engines on December 29, 2017

At the same time, we make a power-law fit to the points in Figure 1 to find the Zipf index. According to the power-law fitting line in Figure 2 below, we can get the Zipf index $\beta = 1.21419$.

From these two figures, we can easily see that the index points except for the first few index points converge to a straight line, which is in line with the prediction and is in line with Zipf's law. For the first few index points, let's explain below.

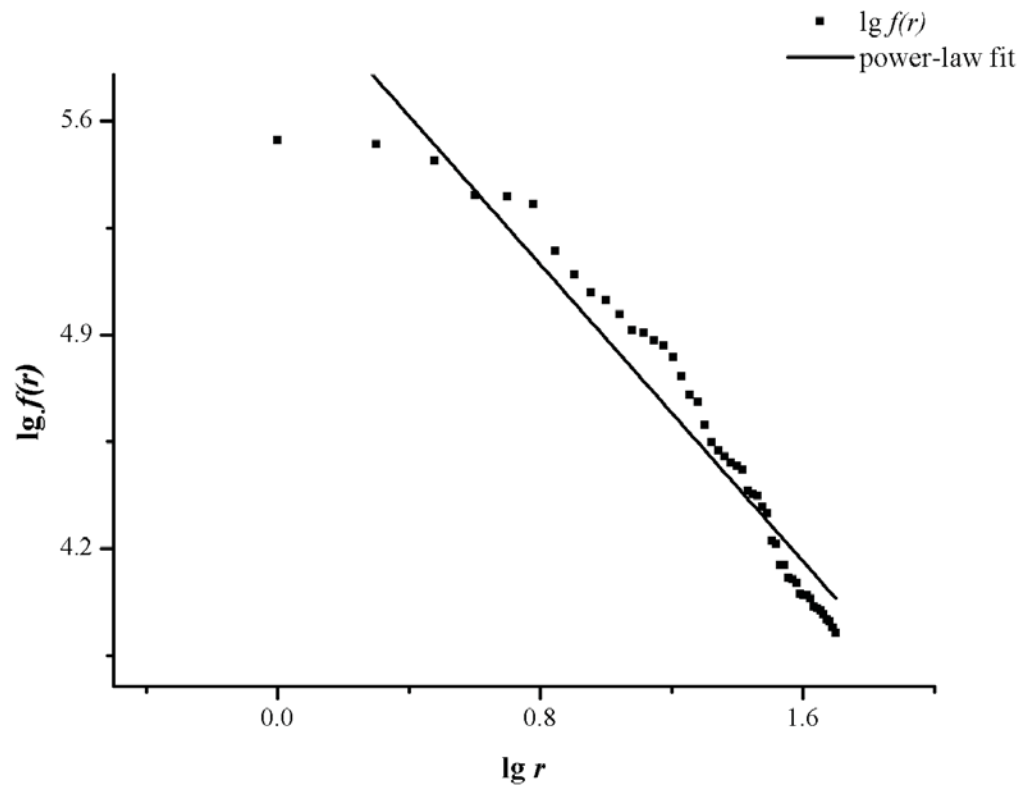


Figure 2. Relevant data on hot events in search engines on December 29, 2017

The distribution of the first few index points is not very regular, because the search engine will recommend the hottest hot events in a variety of ways, which directly affects the search index of these hot events, so this is an inevitable accidental phenomenon without objectivity.

4. ZIPF DISTRIBUTION OF WEEKLY HOT EVENTS IN SEARCH ENGINES

In order to eliminate the risk brought by the asymmetric situation, this paper also takes a week as a unit to collect the comprehensive search indexes of hot events in search engines. Similarly, the search index of each hot event is arranged in decreasing order as a sequence, and each hot event after arrangement is labeled with a natural number increasing from 1. We study this data set in Zipf estimates.

This paper selects the hot events and their search indexes in the 53rd week of 2017 as a data set and sets up the following table according to decreasing order of search indexes.

Table 2. Relevant data on weekly hot events in search engines in the 53rd week of 2017

Hot events	S/N r	Index $f(r)$	$f(r) * r$	$\lg r$	$\lg f(r)$
Didi-Motorcycle was stopped	1	764087	764087	0.000000	5.883143
Ma Rong questioned Wang Baoqiang	2	738114	1476228	0.301030	5.868123
Shared boyfriends appeared in Haikou	3	552193	1656579	0.477121	5.742091
Huaxi Village debt 38.9 billion	4	514508	2058032	0.602060	5.711392
Workers removed bones from chicken claws through their mouths	5	490205	2451025	0.698970	5.690378
Ma Su ' s playing Yang Yuhuan is amazing	6	469353	2816118	0.778151	5.671500
Tencent computer housekeeper apologized	7	447991	3135937	0.845098	5.651269
Hu Ge knelt on the ground to sign	8	431208	3449664	0.903090	5.634687

The event of subway rolling people in Shenzhen	9	424482	3820338	0.954243	5.627859
Zhang Han broke up with Nazha G.	10	423504	4235040	1.000000	5.626858
.....
Father accidentally crushed the son while reversing	47	146380	6879860	1.672098	5.165482
The Imperial Palace wall was blown down by the wind	48	134405	6451440	1.681241	5.128415
Two-year-old boy hit the wall and died	49	132501	6492549	1.690196	5.122219
A three-no mp3 exploded suddenly	50	131790	6589500	1.698970	5.119882

Similarly, we use the $\lg r$ as the abscissa, $\lg f(r)$ as the ordinate, and the collected data and its logical relations are described by a figure to construct a Zipf distribution model with a week as a unit.

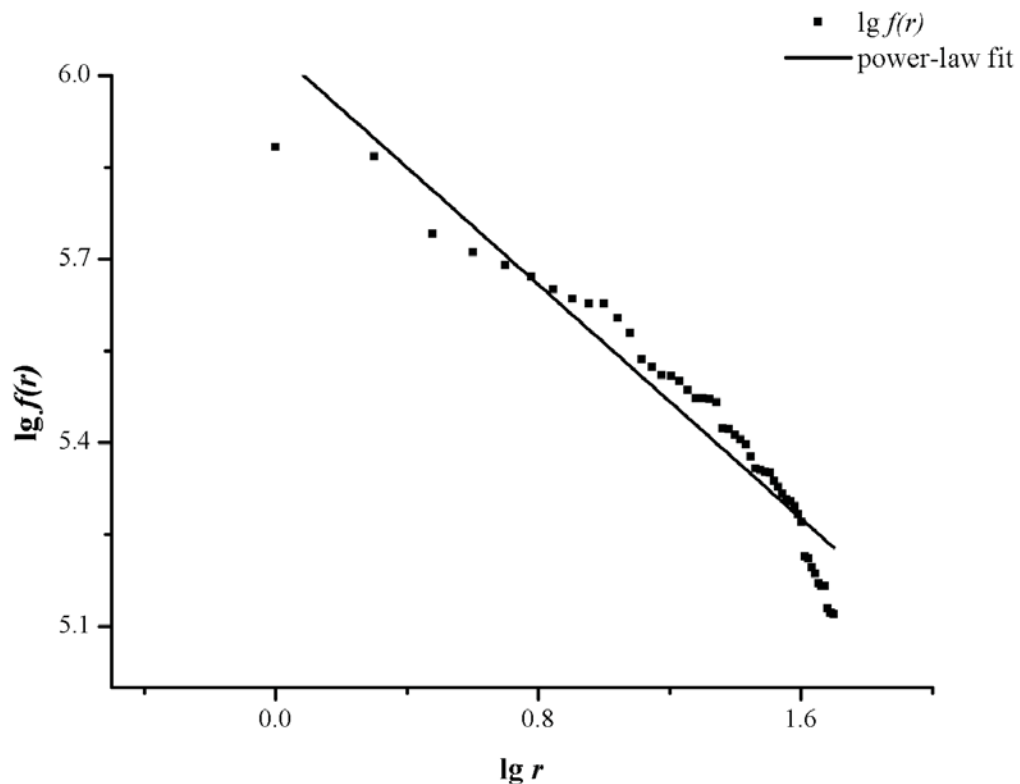


Figure 3. Relevant data on weekly hot events in search engines in the 53rd week of 2017

According to the power-law fitting line in Figure 3, we can get the Zipf index $\beta = 0.47826$. We find that this figure structure approximates the Zipf distribution of the search index sequence of daily hot events. Even better, the first few index points of the figure are closer to power-law fitting straight line because the amount of data collected in a week as a unit to describe the comprehensive search index is much larger, avoiding the asymmetry of the data distribution.

5. PRINCIPLE OF LEAST EFFORT

Zipf discovered Zipf's law in a number of unrelated phenomena and proposed the Principle of Least Effort to explain the causes of this regularity^[15]. Zipf considered that the economy of words need to be discussed from perspectives of both the speaker and the listener. From a speaker's point of view, it is economical to express various meanings in a single word. On the contrary, a listener wants the exact correspondence between the forms and meanings of words^[16]. These two principles are contradictory.

The Principle of Least Effort applied to the field of e-commerce, the sender in the transmission of information and the receiver in the acquisition of information both have propensities for the economy. Zipf distribution describes the balance of economic propensity of sender and receiver during the transmission of information, and this is quite useful for improving the efficiency and effectiveness of commercial activities in e-commerce.

6. ANALYSIS OF THE FLUCTUATION OF THE ZIPF INDEX

Based on the big data sample of twenty consecutive days, we make the following fluctuation figure of the Zipf index and find that the search index sequences of daily hot events in the observation period all conform to the Zipf distribution. At the same time, their Zipf indexes fluctuate between a small range, mainly between 1.00 to 1.22. However, one of the Zipf indexes is particularly high, which we will analyze later. The average Zipf index in the observation period is calculated as 1.12636.

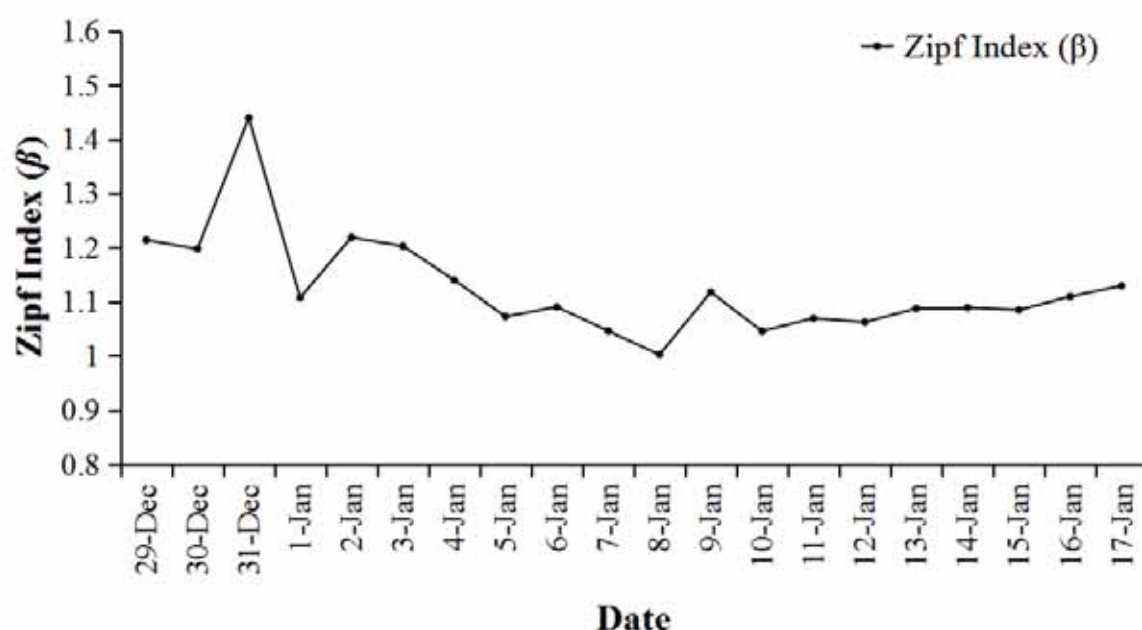


Figure 4. Fluctuation of the Zipf index of daily hot events in search engines in the observation period

As shown in Figure 4, we can visually see the fluctuation of the Zipf index.

According to the change of data in the observation period, we can not judge whether this fluctuation is cyclical or not. However, we suspect that the formation of fluctuation is affected by the non-network information medium, which is equivalent to being forced to vibrate by an irregular driving force.

Next, we focus on the peak of the fluctuation of the Zipf index, which is the Zipf index of the search index sequence of hot events on December 31, 2017. Obviously, the hot event on the top of the sequence on that day is different from that on previous few days, and the search index of this hot event is much higher than the search index of the top hot event on other dates.

The Zipf index of the data set for that day is much higher than the average Zipf index of the entire observation period, however, when we remove the top-ranked hot event, the Zipf index of the search index sequence consisting of the remaining forty-nine hot events is very close to the average Zipf index over the entire observation period.

Because that day is the last day of 2017, the hot event on the top of the sequence that day is related to the

New Year, which is of special significance and receives a great deal of attention. This indicates that hot events with unusually high search index have a direct effect on the Zipf index of the sequence.

7. CONCLUSION

This paper takes the hot events in search engines as the research object, and verifies that the search index sequences of daily hot events and weekly hot events accord with Zipf's law. A few hot events are the objects of most people's attention at the same time, and most of the hot events are those of a few people, which shows that there are similarities and differences in the people's attention. Based on the statistics of big data samples lasting twenty dates, we find that the search index sequences of daily hot events in the observation period all conform to the Zipf distribution, and their Zipf indexes mainly fluctuate between 1.10 and 1.26. Only a small number of events can maintain long-term heat, and most of them are short-term hot events.

This paper also takes a week as a unit to collect the comprehensive search index of hot events in search engines so as to eliminate the interference of the asymmetric situation on the research. At the same time, this paper also collects big data samples for 20 consecutive days to eliminate the interference of contingency.

REFERENCES

- [1] Manyika J, Chui M, Bughin J, et al. Disruptive Technologies: Advances That will Transform Life, Business, and the Global Economy[R]. McKinsey Global Institute, May, 2013.
- [2] Hu Beibei, Wang Shengguang. The New Production Function of the Internet Age[J]. Studies in Science of Science, 2017, 35(9): 1308-1312. (in Chinese)
- [3] Feng Shiqi, Zhu Dedong. On the Dual Impacts of Internet Hot Events on Cybercitizens' Social Mentality[J]. Journal of Chongqing Institute of Technology, 2016, 30(8): 67-71. (in Chinese)
- [4] Wang Shiyong. Understanding Network Culture[M]. Chongqing: Chongqing Publishing Group, 2011: 47-50. (in Chinese)
- [5] Zeng Yiling, Xu Hongbo. Research on Internet Hotspot Information Detection[J]. Journal on Communications, 2007, 28(12): 141-146. (in Chinese)
- [6] Cass R S. Going to Extremes: How Like Minds Unite and Divide[M]. New York: Oxford University Press, 2009.
- [7] Fang Xi, Li Na, Ge Yuefeng. Evaluation System for Chinese Search Engines Based on the AHP-TOPSIS[J]. Science & Technology Review, 2012, 30(14): 49-54. (in Chinese)
- [8] Zipf G K. Human Behavior and the Principle of Least Effort[M]. Cambridge: Addison-Wesley Press, 1949.
- [9] Zipf G K. The Psycho-Biology of Language: An Introduction to Dynamic Psychology[M]. Cambridge: Addison-Wesley Press, 1968.
- [10] Vandewalle N, Ausloos M. The n-Zipf Analysis of Financial Data Series and Biased Data Series[J]. Phys Rev E, 1995, 52(1): 446-452.
- [11] Yoshi Fujiwara. Zipf Law in Firms Bankruptcy[J]. Physica A: Statistical Mechanics and its Applications, 2004, 337(1): 219-230.
- [12] Nitsch V. Zipf Zipped[J]. Journal of Urban Economics, 2005, 57(1): 86-100.
- [13] Zipf G K. Human Behavior and the Principle of Least Effort[M]. Cambridge: Addison-Wesley Press, 1949.
- [14] Li Yujian, Xiao Chuangbai. Zipf's Law Probably Existing in Protein Sequences[J]. Journal of Beijing University of Technology, 2005, 31(4): 366-368. (in Chinese)
- [15] Zipf G K. Human Behavior and the Principle of Least Effort[M]. Cambridge: Addison-Wesley Press, 1949.
- [16] Jian Wangqi. Zipf and the Principle of Least Effort[J]. Tongji University Journal(Social Science Section), 2005, 16(1): 88-95. (in Chinese)

Quality Prediction of Answers in Community of Questions and Answers of Zhihu

Ming Li, ¹Yi Zhang^{1,}, Kaijuan Xing², Xiaoyu Qi¹*

¹School of business administration, China University of Petroleum-Beijing, China

²School of information, University of Texas at Austin, USA

Abstract: The participation of Web2.0 in collaboration makes the Q&A social platform develop rapidly, but the problem that the answering quality is difficult to distinguish has gradually become apparent, even it hindered the healthy development of social platforms. In order to predict the Answers in Zhihu, which is the famous Community of Questions and Answers website, the quality prediction method based on Logistic Regression is proposed. Firstly, we collect dataset from Zhihu. Based on the factors, the training set is constructed. After that the two answer quality prediction models based on Logistic Regression is constructed. With the comparison of the two models, the final model is determined. It can be used to predict the answer quality of Zhihu.

Keywords: Community of questions and answers; Answer quality prediction; Zhihu; Logistic Regression

1. INTRODUCTION

With the coming of Web 2.0, the way people obtain information has changed from the original one-way transmission to user-centered, which is a network communication mode emphasizing collective sharing [1,2]. Various Community of questions and answers (CQA) platforms emerged. Their emergence greatly improved the efficiency of the Internet use, and to a certain extent, changed the inherent mode of the Internet. Q&A social networking sites use questions and answers as the primary mode of communication, where users can obtain the similar question and answer with that users posted previously, or invite users on the site to answer the question, and then choose an answer with high quality [3,4]. Zhihu is such a Q&A community platform. In recent years, the development of Zhihu is very fast, and registered users have also increased significantly [5,6]. The information on Zhihu is mainly based on questions and answers of various topics, and the questioner and the responder can communicate through the platform. The main service of Zhuhu is to provide users with high quality answers, but with the amount of users and information increasing, the quality of the answer becomes more and more difficult to distinguish, so the quality of the information provided to users is difficult to guarantee, which becomes a challenge for the development of community websites like Zhihu [7-9]. Therefore, it is essential to build a scientific method of evaluating and predicting the quality of the answers.

The evaluation of answer quality is one of the hot topics. Shah C took Yahoo Answers as the research object, let five manual scorer from Amazon score from thirteen dimensions including informative, polite, complete readable, relevant, brief, convincing, detailed, original, objective, novel, helpful and expert, and then analyzed the data and set up the model based on Logistic Regression. Then he further extracted questions, answers and users characteristics and used the same method for analysis and modeling [10,11]. Agichtein E [12] is also a study of Yahoo Answers, based on the analysis to variable factors associated with high quality contents, ultimately created an icon based quality classification framework with the algorithm model of the relationship between questioners and respondents, text features of the content and features of the usage. Toba et al [13] used Support Vector Machines, Logistic Regression and other basic classification methods to analyze the extracted

text and non-text features, and then by comparison, a classification method with the best effect was obtained. They also analyzed the relationship between the types of questions and quality of answers. Hoang, Lee et al extracted the characteristics of four aspects: the answer, the rigor, the readability and the subjectivity, then used the classification method of maximum entropy to analyze the data, and then established the quality classification model [14]. Jia Jia et al. assessed the model by the answer quality with thirteen dimensions, and evaluated the answering quality of the two domestic social Q&A platforms—Zhihu and Baidu knows through the questionnaire [15].

These researches mostly focused on the famous CQA website Yahoo! Answers, but less research is targeted at the Chinese CQA website. In this paper, we take the Zhihu as the main research object and establish the experimental training set. Different classification methods are used to analyze the data. An optimal algorithm model is determined through comprehensive analysis and comparison. The second part of the paper is a detailed introduction to the adopted Logistic Regression. The third part introduces the related contents of the experiment, including the introduction to the specific process of the experiment, the analysis of the experimental results and the quality prediction model established. The fourth part is the summary of the research, the significance of the study, the analysis of the existing problems and the prospect of the future research work.

2. LOGISTIC REGRESSION

As one of the most important classification models in pattern recognition and machine learning, Logistic Regression is an interpretable model and has a good generalization ability [16]. Logistic Regression is a regression model in which variables are classified and are applied to various fields including machine learning, medical field and social science [17]. For the two-element Logistic Regression function (that is, there are only two types of the dependent variable) it can take any actual input T and the output always values between 0 and 1. Logical Regression is a typical probability statistical classification model. It fits logarithm by the linear function and can be expressed as:

$$\ln\left(\frac{p(y=1|x)}{p(y=2|x)}\right) = w^T x \quad (1)$$

w is a fitting parameter, x is an instance, y represents the label of the class and the p represents the conditional probability.

Because the sum of the probability of each class is 1:

$$\sum_{j=1}^2 p(y = j|x) = 1 \quad (2)$$

Combine formula (1) and formula (2), available:

$$p = (y = 1|x) = \frac{\exp(w^T x)}{1 + \exp(w^T x)} \quad (3)$$

$$p = (y = 2|x) = \frac{1}{1 + \exp(w^T x)} \quad (4)$$

The value of w is often estimated by the maximum likelihood method, and its logarithmic likelihood function is:

$$L(w) = \sum_{j=1}^2 \sum_{i=1}^{N_j} \ln p(x_i^{(j)}|y = j; w) \quad (5)$$

N_j represents the number of class j instances[16].

In conclusion, it is a generalized linear model with sigmoid function and can also be expressed as:

$$P_i = \text{Logit}^{-1}(\beta \cdot X_i) = \frac{1}{1 + e^{-\beta \cdot X_i}} \quad (6)$$

$\beta \cdot X_i = \beta_0 + \beta_1 \cdot x_{1,i} + \dots + \beta_m \cdot x_{m,i}$, β is the Regression coefficient, x_i is the explanatory variable^[17].

3. EXPERIMENT

3.1 Logistic Regression [1]

We collect 500 answers from Zhihu. Then the quality of the answers is evaluated manually. The standard of assessment is whether the content of the answer is related to the question, whether the information provided in the answer is true and reliable, whether it can solve the questioner's question and so on. We refer to the thirteen dimensions scoring model of Community answer quality proposed by Shah C et al [11]., let each scorer scores for answers respectively from the angles of Informative, Polite, Complete, Readable, Relevant, Brief, Convincing, Detailed, Original, Objective, Novel, Helpful, and Expert in the range of 0-5, and according to those scoring points and the assessment data of answer quality, we set up a training set. The prediction of answer quality is actually a classification problem. We will divide the answer quality into two categories, including good quality and poor quality, which are expressed by 1 and 0 respectively. 1 corresponds to good quality, while 0 represents poor quality.

The essence of Logistic Regression is a classification model, the input value is random, but there are only two types of output result, and Shah C et al [11] also used Logistic Regression to model. So we analyze the data of the training set with the method of Logistic Regression using SPSS statistical analysis software. The results are as follows:

Table 1 Variables in the Equation

Step1	B	S.E.	Wald	df	Sig.	Exp(B)
Informative	-.001	.087	.000	1	.993	.999
Polite	.008	.087	.009	1	.923	1.008
Complete	.107	.089	1.426	1	.232	1.113
Readable	.065	.098	.443	1	.506	1.067
Relevant	.711	.106	44.659	1	.000	2.036
Brief	.086	.098	.772	1	.379	1.090
Convincing	-.017	.103	.026	1	.871	.983
Detailed	.917	.110	68.962	1	.000	2.501
Original	.007	.092	.006	1	.940	1.007
Objective	-.057	.092	.381	1	.537	.945
Novel	-.106	.097	1.206	1	.272	.899
Helpful	-.020	.089	.052	1	.819	.980
Expert	.040	.092	.188	1	.664	1.041
Constant	-5.052	.775	42.454	1	.000	.006

In table 1, B represents the coefficient of variables. Sig. reacts to the significance of variables. If the Sig. < 0.05 of a equation's variable, it means that this variable has statistical significance, and vice versa. According to the data in the table, it can be found that only "Relevant" and "Detailed" variables have statistical significance. Therefore, the poorly significant variables are removed and the data are analyzed again, and the results are as follows:

Table 2 Variables in the Equation

Step 1	B	S.E.	Wald	df	Sig.	Exp(B)
Relevant	.734	.101	52.440	1	.000	2.084
Detailed	.875	.106	68.185	1	.000	2.398
Constant	-4.803	.415	133.710	1	.000	.008

The model of Logistic regression that can be obtained from the above table is as follows:

$$P = \frac{1}{1 + e^{-(4.803 + 0.734 * \text{Relevant} + 0.875 * \text{Detail})}}$$

When the probability of prediction is $P > 0.5$, the quality of the answer is good; when $P < 0.5$, the quality of the answer is poor.

Table 3. Classification Table

Observed			Predicted		
			The quality of the answer		Percentage
			NO	YES	Correct
Step 1	The quality of the answer	NO	313	38	89.2
		YES	65	84	56.4
	Overall Percentage				79.4

This table responds to the accuracy of the model. Known from the table, the accuracy rate of the answer predicted with pool quality is 89.2%, and the accuracy of the answer predicted with good quality is 56.4%, and the total accuracy is 79.4%.

Table 4. Model Summary

-2 log likelihood		Cox&Snell R Square	Nagelkerke R Square
Step			
1	432.135 ^a	.296	.420

Table 4 reflects the fitting degree of the model. The closer the value of Cox&Snell R Square and Nagelkerke R Square is to 1, the better the model fits. According to the table, the two values are small, so the fitting degree of the model is not very ideal. The requirements of the model algorithm established by the thirteen dimensional method are relatively high for the user itself, because users' scoring on different aspects directly affects the prediction of answer quality, so the final predictive results are highly uncertain and the precision and the fitting degree of the model is not high.

4.2 Logistic Regression [2]

In order to improve the practicability of the model, we decide to establish a new model, Toba [7], Agichtein E[12], Shah C et al [10,11], analyzed directly to the question text and non-text features, so we decide to extract relevant features from questions and answers on Zhihu. Because the quality of the answer is not only related to the answer itself, also may be associated with the problem and answer users, we decide to extract the

characteristics of questions, answers and respondents. According to the characteristics of Zhihu operating mechanism, we finally decide to extract the following features: the length of the question, the number of labels of the question, the number of comments on the question, the number of answers to the question; the order of the answer, the length of the answer, the number of sentences in answers, the number of comments of the answer, the number of approval of the answer; the number of answers of the respondent, the number of share of the respondent, the number of questions of the respondent, The number of respondents' attention, the number of concerns of the respondent, the number of endorsed of the respondent, the number of thanks to the respondent and the number of collection to the respondent. The above characteristics almost involve all aspects of questions and answers. Therefore, the models built on this are more practical and objective.

We have collect another 500 answers from Zhihu. The criteria for judging the quality of answers remain unchanged, and then a new training set is set up. We still use Logistic Regression to analyze the relevant data of the training set, and the final results are as follows:

Table 5. Variables in the Equation

Step1	B	S.E.	Wald	df	Sig.	Exp(B)
The length of the question	.034	.013	6.460	1	.011	1.034
The number of labels of the question	.036	.015	5.684	1	.017	1.036
The number of comments on the question	-.010	.005	3.307	1	.069	.990
The number of answers to the question	.000	.000	3.988	1	.046	1.000
The order of the answer	.048	.053	.836	1	.361	1.050
The length of the answer	.001	.000	26.993	1	.000	1.001
The number of sentences of the answer	-.008	.008	1.017	1	.313	.992
The number of comments of the answer	-.001	.001	1.117	1	.291	.999
The number of approval of the answer	.000	.000	6.643	1	.010	1.000
the number of answers of the respondent	.000	.000	.061	1	.805	1.000
the number of share of the respondent	-.003	.003	.684	1	.408	.997
the number of questions of the respondent	.001	.001	.461	1	.497	1.001
The number of respondents' attention	.000	.000	.002	1	.962	1.000
The number of concerns of the respondent	.000	.000	6.362	1	.012	1.000
The number of endorsed of the respondent	.000	.000	.443	1	.506	1.000
The number of thanks to the respondent	.000	.000	.165	1	.685	1.000
The number of collection to the respondent	.000	.000	.031	1	.861	1.000
Constant	-3.860	.605	40.708	1	.000	.021

From the Table 5, we can see that variables of Sig.<0.05 include: the length of the question, the number of the question, the number of answers to the question, the answer length, the answer agreement and the respondents' attention. Therefore, removing the variables with poor significance, the data of the training set are analyzed again and the results are as follows:

Table 6. Variables in the Equation

Step1	B	S.E.	Wald	df	Sig.	Exp(B)
The length of the question	.024	.012	4.035	1	.045	1.024
The number of labels of the question	.034	.014	5.820	1	.016	1.035
The number of answers to the question	.000	.000	.389	1	.533	1.000
The length of the answer	.001	.000	60.036	1	.000	1.001
The number of approval of the answer	.000	.000	10.422	1	.001	1.000
The number of concerns of the respondent	.000	.000	17.658	1	.000	1.000
Constant	-3.389	.484	48.961	1	.000	.034

The table shows Sig=0.533>0.05 of "the number of answers to the question", so it is necessary to remove "the number of answers to the question" variable and reanalyze the data.

The result is the following table:

Table 7. Variables in the Equation

Step1	B	S.E.	Wald	df	Sig.	Exp(B)
The length of the question	.024	.012	3.893	1	.048	1.024
The number of labels of the question	.033	.014	5.585	1	.018	1.034
The length of the answer	.001	.000	60.077	1	.000	1.001
The number of approval of the answer	.000	.000	12.013	1	.001	1.000
The number of concerns of the respondent	.000	.000	17.969	1	.000	1.000
Constant	-3.348	.480	48.741	1	.000	.035

A new Logistic Regression model can be obtained from the table as follows:

$$P = \frac{1}{1 + e^{-(-3.348 + 0.024a + 0.033b + 0.001c + 0.000146d + 0.000020e)}}$$

Among them, a, b, c, d and e respectively represent the length of the problem, the number of labels of the question, the length of the answer, the number of answers to the question, and the number of concerns of the respondent.

Table 8. Classification Table

Observed			Predicted		Percentage
			The quality of the answer		Correct
			No	Yes	
Step1	The quality of the answer	No	321	21	93.9
		Yes	68	90	57.0
	Overall Percentage				82.2

From Table 8, the total probability of the prediction of the new Logistic Regression model is 82.2%, which is better than the previous one.

Table 9. Model Summary

-2 Log likelihood		Cox&Snell R Square	Nagelkerke R Square
Step			
1	425.126 ^a	.328	.460

From the table, the values of Cox&Snell R Square and Nagelkerke R Square are all increased, and the fitting degree of the model is also improved.

Therefore, the effect of the established algorithm model is quite good, no matter the accuracy rate or the degree of fitting is ideal, and the variables are directly acquired characteristics eliminating the influence of human factors.

According to the results of the above experiments, the accuracy of each algorithm model is as follows:

Table 10. Accuracy

Algorithm model	Logression ^[1]	Logression ^[2]
accuracy rate	79.4	82.2

According to the above table, the best model in the modeling is the Logression^[2] model.

5. Conclusions

In this paper, the method for quality prediction of answers in Community of Questions and Answers is proposed. Based on the collected dataset from the famous Community of Questions and Answers website Zhihu and the factors that have influence on the quality of answers, the quality prediction method based on Logistic Regression is proposed. After that the two answer quality prediction models based on Logistic Regression is constructed. With the comparison of the two models, the final model is determined. It can be used to predict the answer quality. Our experiments also have some shortcomings and improvements, such as the training dataset is too little. We can further explore the optimal model; the quality and the type of problems may affect the quality of the answer; respondents' mood may also be related to the quality of the answer and all of these can further be researched.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71571191, the Humanity and Social Science Youth Foundation of Ministry of Education in China (Project No. 15YJCZH081), National Natural Science Foundation of China under Grant 91646122, the Science Foundation of China University of Petroleum, Beijing (No. 2462015YQ0722).

REFERENCES

- [1] Kamel B, M. N, & Wheeler S.(2007).The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. *Health Information & Libraries Journal*, 24(1): 2-23.
- [2] Deng Z, Luo L. (2007). An exploratory discuss of new ways for competitive intelligence on Web2. 0. *Integration and Innovation Orient to E-Society Volume 2*, 597-604.
- [3] Liu J, Shen H, Yu L. (2017). Question quality analysis and prediction in community question answering services with coupled mutual reinforcement. *IEEE Transactions on Services Computing*, 10(2): 286-301.
- [4] Figueroa A.(2017).Automatically generating effective search queries directly from community question-answering questions for finding related questions. *Expert Systems with Applications*, 77: 11-19.
- [5] Zhang J., Ackerman M S, Adamic L. (2007). Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web* ,221-230
- [6] Dror G, Koren Y, Maare k , Szpektor I.(2011).I want to answer; who has a question?: Yahoo! answers recommender system. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* , 1109-1117.

- [7] Toba H, Ming Z Y, Adriani M, Chua T S.(2014).Discovering high quality answers in community question answering archives using a hierarchy of classifiers. *Information Sciences*, 261: 101-115.
- [8] Liu Y, Agichtein E. (2008).On the evolution of the Yahoo! Answers QA community. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*: 737-738.
- [9] Shtok A, Dror G, Maarek Y, Szpektor I.(2012).Learning from the past: answering new questions with past answers. In *Proceedings of the 21st international conference on World Wide Web*: 759-768.
- [10] Shah C,Kitzie V,Choi E.(2014).Questioning the Question-Addressing the Answerability of Questions in Community Question-Answering.IEEE Xplore,1530-1605.
- [11] Shah C, Pomerantz J.(2010).Evaluating and Predicting Answer Quality in Community QA.ACM SIGIR conference ,411-418.
- [12] Agichtein E,Castillo C,Donato D,Gionis A,Mishne G.(2008). Finding High-Quality Content in Social Media. International Conference on Web Search and Data Mining,183-194.
- [13] Toba H, Zhao Y M, Adriania M,Chua T-S.(2014).Discovering high quality answers in community question answering archives using a hierarchy of classifiers. *Information Sciences* 261,101-115.
- [14] Hoang L,Lee J-T,Song Y-I,Rim H-C. (2008). A Model for Evalating the Quality of User-Created Documents. *Information Retrieval Technology*496-501.
- [15] Jia Jia,Song Enmei,Su Huan.(2013). The quality assessment of the social question and answer platform - "Zhihu", "Baidu Zhidao" as an example [J]. *Journal of information resources management*, 02: 2095-2171. (in Chinese)
- [16] Guo Huaping.,Dong Yadong.,Wu Changan&Fan Ming.(2015). Logistic Regression Method for Class Imbalance Problem.*Pattern Recognition and Artificial Intelligence*,8.
- [17] Le L T, Shah C, Choi E.(2016).Evaluating the Quality of Educational Answers in Community Question-Answering[J],129-138.

A Study on Online-Shopping Opinion Leaders Based on Massive Data

Chong Zhang^{*}

College of Information Sciences, Beijing Language and Culture University, China

Abstract: The online communities are important channels for consumers to distribute, transmit and access word-of-mouth information. However, the influence power of the users has a significant difference. This article selects about 200 Online Communities (taobao wangwang group), based on more than one million taobao real transaction data provided by nearly fifty thousand members of the communities, from the perspective of social network to build weighted and directed graph of online shopping interpersonal influence relations and users' influences model, mining online shopping opinion leaders and exploring the community online communication, spread and effect regularities to provide theoretical basis and practice guidance for word-of-mouth marketing.

Keywords: virtual community, word-of-mouth marketing, online shopping, opinion leaders

1. INTRODUCTION

With the rapid development of e-commerce, the phenomenon of online information asymmetry is growing. Numerous studies have demonstrated that word-of-mouth can reduce the degree of information asymmetry and influence consumers decision-making. However, a lot of word-of-mouth information release in online community such as BBS, QQ group. The online community is an important channel for word-of-mouth information to distribute, transmit, and access.

In the process of interpersonal interaction and word-of-mouth behavior, the existence of opinion leaders has become a consensus. Compared with normal members of groups, the opinion leaders are considered to have greater influence on consumers. In the past study of opinion leaders' recognition and feature research, the vast majority of research has focused on online community such as BBS or SNS, using the post or replies to map the relationship between users, but mining opinion leaders or highly influential users in online-shopping field are very rare, especially the research of online shopping opinion leaders based on actual purchase behavior and transaction data is blank. This article selects about 200 TaoBao wangwang groups as the research object, based on more than one million TaoBao real transaction data provided by nearly fifty thousand members of the communities, from the perspective of social network to build weighted and directed graph of online shopping interpersonal influence relations and users' influences model, mining online shopping opinion leaders and exploring the community online communication, spread and effect regularities to provide theoretical basis and practice guidance for word-of-mouth marketing.

2. CONCEPTUAL MODEL AND RESEARCH DESIGN

2.1 Conceptual model

Model's basic idea: there are a few members in the community, at first ,member 1 purchase the goods; the entire process(before-purchase, purchase and post-purchase) can produce various kinds of information related to goods and services; there are two kinds of forms of information, the first is the cognition, attitude, evaluation about goods and services, that is word-of-mouth information; The second is behavior information, the purchase behavior is also a kind of information, which can be others perceived and impact on others.

^{*} Chong Zhang. Email: zhangchong@blcu.edu.cn

These information are created in these communities, and spread them to other members, such as member 2 member 3 and member 4, through community platforms. the information receiver will be affected, showing the change of attitude and purchase behavior about goods. The most direct effect shows as following purchase, such as members 2 also purchased the same goods after member 1 purchasing them, therefore, we can say there's obvious interpersonal influence relation between member 1 and 2. Similarly, the interpersonal influence relation also can exist between others members, such as between member 3 and member 4. The interpersonal influence relation presents as net-like structure.

2.2 the Identify of the Interpersonal Influence Relationship

In behavior influence, the most direct expression is the relationship between the purchase and following purchase as member 1 firstly buy a commodity, then member 2 in the same group was affected and followed to buy the same commodity over a period of time. From this view to measure consumer influence relationship will be highly accurate and objective. Therefore, this article's empirical research will take "purchase - following purchase" as important factors affecting the relationship between group members to construct the shopping influence relationship network.

3. BUILD THE INFLUENCE RELATIONSHIP NETWORK

3.1 Introduction of the Data

This research uses the real transaction data of taobao users from taobao website (www.taobao.com), which is the largest c2c e-commerce website of China. Each data recorded a transaction. The contents of each record is as follows:

- Trading time: xx month xx day xx hour xx minute
- Buyer ID: the buyer's unique identifier in a transaction, it can be a string variable,
- Buyers' wangwang group: the number of wangwang group that buyers joined
- Seller ID: the seller's unique identifier in a transaction,
- Trading goods ID: the goods' unique identifier in a transaction.
- Trading goods' name: goods display name, such as "super fantasy small-cake sleepshirts"

These variables above describes the buyer, the seller (trade subjects) and trading goods' information (the object) in any transaction in detail. "Buyers' wangwang group" is to determine whether a buyer belongs to the same online community.

3.2 Online interpersonal influence' measuring and judging

Based on the transaction data above, if the trading behavior between two users satisfy the following three conditions, then we can determine that there exist influence relationship. Judging conditions are as follows:

- Firstly, the users must have influence/transmission approach, which means that the users belong to the same wangwang group. The community's members possess the online transmission and communication approaches.
- Secondly, the users can affect others' behaviors in the same community, that is "purchase - following purchase" behavior, if the users in the same community buy a goods of the same stores, so there's may exist influence relationship.
- Finally, the two successive purchasing behavior must satisfy a certain time range, and the interval is

generally less than 30 days.

Having the three conditions, for example: User A and B belong to a same online community. A bought of goods 1 in a store .A week after, B also bought the same goods, or buy other goods of the same stores. Then we say B followed A to purchase, so this article judged that B's purchasing behavior is influenced by A.

We calculate the interaction relationship between any two members within community, and part of the calculation results shown in table 3.1 . The first column in the table is the name of the goods; the second column is influencers' ID; the third column is the affected' ID; the fourth column is the number of influence.The first record in the table shown that in "Tea life" want want group, users "97649038" influence user "70568760" on shopping behavior by 11 times.

Table 3.1 The sample of online shopping interpersonal influence determination results

Tea life	97649038	70568760	11
	70022244	408408681	9
	70022244	97649038	9
	70022244	70568760	8

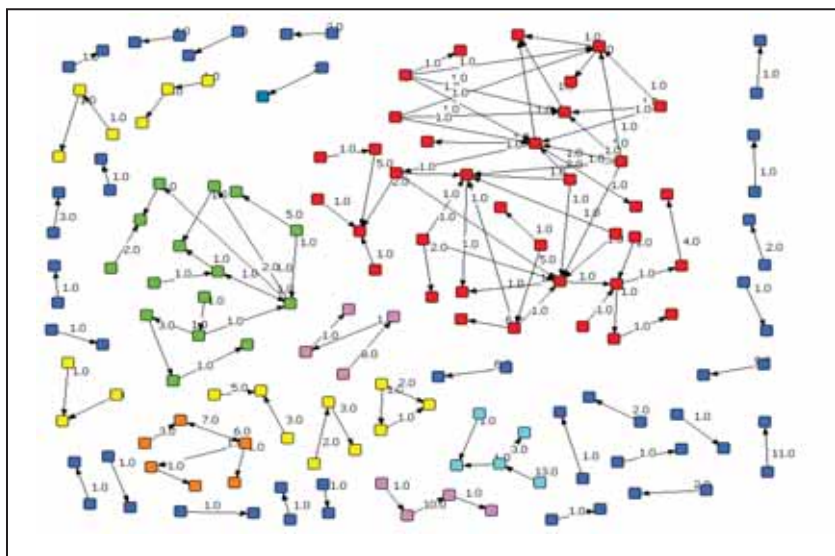
3.3 Build the directed and weighted network

After measuring the relationship between interpersonal influence of online shopping, we will draw on the relevant methods of graph theory , use the directed and weighted network to describe the relationship in the community, every community will has its own network. This method will help to describe the influential relationship more intuitively, and more convenient for further study the characterizations of the relationship. In this paper, the data model is defined as $G=(V,E,W)$, which describes the relationship between interpersonal influence of online shopping.

The nodes in the node set V represents the users from the same community.

"E" represents the edges of the network. "Edge $\langle V_i, V_j \rangle \in E$ " represents the relationship between V_i and V_j , which means if V_i has an impact on V_j , there will be an edge or connection pointed from V_i to V_j .

"W" represent the weight of the edge. " W_{ij} " describes the intensity of the effective relationship. In the network, the value of W represents the frequency of the influence. For example, if V_j has 5 shopping actions which are influenced by V_i , the weight W_{ij} between V_i and V_j will be 5. The directed and weighted network not only describes the direction of influence between users, but also describes the intensity of the influential relationship.



Picture 3.1 the network of the relationship

4. THE RECOGNITION OF ONLINE SHOPPING OPINION LEADERS

4.1 The Construction of the Community Members online shopping impact index

In this study, the evaluation of the influence from the community members will be described through two aspects, the breadth and the depth of the influence. In the directed network, the out-degree of nodes means the numbers of users which directly influenced by it, it measures the breadth of the direct influence. The in-degree means the number of the users who has influenced this node, measuring the acceptance of this node. High out-degree means wider breadth of the node. High in-degree means this node is much easier to be influenced, which shows that the trend following is more obvious.

In the network we have built before, the relationship of influence between users not only has direction, but also has weight “W”, the weight “W” represents the intensity of the relationship. We defined the number of impact as the weight of the relationship. If user i influenced user j , there will be a connection pointed from node i to node j , $E_{ij}=1$. If the number of impact from node i to node j is n , the weight of influence relationship between i and j will be n . The weight reflects the depth of the impact between member i and j , depicts the loyalty or stickiness.

Therefore, we construct an evaluation index of the influence on online shopping of the members from the community. This index shows the members’ influence in the whole network.

4.2 The reorganization result of high influential members

In this study, we measured the influence index of 50 thousand members from 2 hundred communities. We have found that the exponents also distribute as long-tailed. The minority members as head has stronger influence, and the tail represents most of ordinary customers.

Table4. 1 the influence index of community members

Out-degree (Influence Breadth)	max	79
	min	0
	mean	4.32

From the table, we can see that the max out-degree of the virtual community reached at 79, which means that some member in this virtual community can influence 79 other members, the breadth of influence is wide. Fwhich means there might be some members who have much stronger influence.

4.3 High impact users identification test based on user interview

In order to test the validity and accuracy of the identification methods, we use a method that make user survey in the small scale to verify the high impact users. This method is used because the test work has some difficulties itself. First, the high impact users belong to minority groups. For the identification rate is 3 percent, which means there may be 3 users in 100 of Taobao users. It is very difficult to find a large number of effective samples in reality. If we choose to deliver the questionnaires to random users, most of them will be meaningless for the test. This method is inefficient. Second, the test work shown in this paper need to identify the users’ identities, which means we have to get the users’ Taobao ID. Then we can compare the survey result with the identification result, which also has difficulties in operation. Therefore, the author finally choose to use the first

method, interviewing in the small scale, to simply test the identification result with the help from Taobao's user survey department.

4.3.1 The interviewee

The interviewees are the users whose node degrees are more than 4. These users influenced others or have been influenced in the measuring result. They belong to the "characteristic users". We finally got 19 users to interview.

4.3.2 Interview outline

The interviews are conducted around the following questions, and the interview time is about 10 to 20 minutes.

- 1) *How many times per month do you shop on Taobao?*
- 2) *Do you ask for advice from colleagues or friends while shopping on Taobao?*
- 3) *How do you get advice from others?*
- 4) *Do you shop because of colleagues or friends' recommendation in this two month?*
- 5) *Do you think it is more satisfied to shop following the colleagues' recommendation?*
- 6) *Do you still pay attention to colleagues or friends' advice and experience sharing after shopping?*
- 7) *Do you share your shopping experiences during or after shopping?*
- 8) *How do you share and communicate your shopping experience with others?*
- 9) *Did someone shop because of your recommendation in this two months?*

we compared the users' answers with the measuring result. If the users answer "don't know" or "don't remember", the interview will be invalid. If the answers are "no" or something like this, we determine that the degree of the node is 0 in this paper. If the answers are "occasionally", "sometimes", "no many" or something, the degrees of nodes will be 1 to 3. If the answers are "absolutely", "always" and so on, the degree of the nodes will be 4 and above. Finally the investigation result will be compared with the measuring result to test.

4.3.3 Analysis of the Test Result

In these 19 users' interviews, one is invalid, and the following is the other 18 analysis results. the test of accuracy, 15 users' interview results are the same as the measuring result, the other 3 are different. The accuracy rate is 83.3 percent, which means that the high impact users' identification in this paper has high accuracy.

For the ways to spread the word-of-mouth information, the most frequency method is "oral communication among colleagues and friends around". The interviewees said, "they asked for advice from the colleagues and friends" and "they always recommend the satisfactory goods to colleagues and friends". Another common way is spreading among friends on some instant messengers like QQ, MSN, Wangwang and so on. The interviewees said, "everybody has different hobbies and interests, colleagues and myself maybe not interested in the same things, so they like to share the shopping information with friends who interested in the same things on QQ and MSN."

4.4 Category attribute characters of members with high influence

Many researches about opinion leaders indicate that generally they have some professional knowledge and high influence in some professional field. So, when shopping online in the virtual community, are those members with high influence possess the same characters, which means that dose their influence only work in some specific field? In chapter one and two of this research, we only consider the whole influence of members, but don't consider the professional field their influence work in. In this character, we will focus on the specialty of high influential members which we call the category attribute character. It subdivides influence of members

into attribute dimension, which shows the characters of members with high influence accurately.

Research method: further analyze those members with high influence whose out-degree is not null, refine the number of total influence according to category, calculate the number of total influence of this member and the number of influence under all categories, and find the category with the maximum number of influence. If that maximum number owns equal or greater than a half proportion of the number of total influence, this member has higher influence in that category and is the highest influential member during this study. Or, the influence of this member is a little dispersive or average, and category attribute is not obvious.

On the whole, in 11583 members whose out-degree is bigger than zero, there are 10988 members whose category with the maximum number of influence is equal or greater than a half proportion, which means about 95 percentage of members have category attribute on influence. To exclude the disturbance of those members whose out-degree is 1 or 2, we only consider those members whose out-degree is equal or greater than 3. The outcome shows that in 5034 members, there are 4439 members whose category with the maximum number of influence is equal or greater than a half proportion, which means about 88 percentage of members have category attribute on influence.

In a word, statistics shows that most (about 88%) high influential members have obvious category attribute, which means the specialty always reflects on one kind of category, and they are more familiar or professional with one kind of category, instead of with many kinds of categories. We classify and collect those high influential members. From the diagram, we can see that high influential members from these fields such as clothes for female, cosmetic, food and maternal and child supplies have almost a half proportion of all fields. Those categories are goods for use and need higher degree of trust.

Table 4.2 top10 categories forming high influential members

Category ID	Category name	Quantity of high influential members	Percentage
16	clothes for female/ladies boutique	657	0.1487
1801	beauty and skin care/bodybuilding/essential oil	446	0.1005
50002766	snacks/nut/tea/specialty	430	0.0968
50014812	diaper/care/feeding/cart bed	394	0.0887
50008165	clothes for child/shoes for child/parent-child	323	0.0727
29	pet/food and supplies for pets	261	0.0588
50010788	cosmetic/perfume/tools for make-up	157	0.0353
25	toy/model/comic/early education/puzzle	156	0.0351
50008090	3C digital accessory market	152	0.03424
50016422	grain and oil/vegetable and fruit/aquatic/fast food	132	0.02974
50022517	clothes for pregnant/supplies for pregnant/nutrition	118	0.02658
35	milk powder/solid food/nutrition	113	0.02546

5. CONCLUSION

Though the view of social network, this study is based on the real transaction data of Taobao user. It judges the relationship between people in the community by the behavior of buy and follow-buy. It structures the directional weighted network which displays the influential relationship between people in the community and also structures the judgment model of influence of members in the community, which identifies high influential

members in the community. By verifying those high influential members in the community, the outcome shows that the identification of high influential members has high accuracy which is more than 83%. The further study of high influential members shows that those members' influence works on one specific classify of goods, which means that they have the ability to influence others in some professional fields.

ACKNOWLEDGEMENT

This research was Supported by Science Foundation of Beijing Language and Cultural University (supported by “the Fundamental Research Funds for the Central Universities”) (17YJ0304)”

REFERENCES

- [1] Momtaz N J, et al. Identifying opinion leaders for marketing by analyzing online social networks[J]. *International Journal of Virtual Communities and Social Networking*, 2011, 3(1): 43-59.
- [2] Hazeldine M F. An exploratory role analysis of opinion leaders, adopters, and communicative adopters with a dynamically continuous innovation[J]. *Journal of Applied Business Research*, 2010, 26(4): 117-130.
- [3] Forbes L P. Does Social media influence consumer buying behavior? An investigation of recommendations and purchases[J]. *Journal of Business & Economics Research*, 2013, 11(2): 107-112.
- [4] Rogers E M. *Diffusion of innovation*[M]. New York: Free Press, 2003.
- [5] Shoham A and Ruvio A. Opinion leaders and followers: A replication and extension[J]. *Psychology & Marketing*, 2008, 25(3): 280-297.
- [6] Huffaker D. Dimensions of leadership and social influence in online communities[J]. *Human Communication Research*, 2010, 36(4): 593-617.
- [7] Doumit G, et al. Opinion leaders and changes over time: A survey[J]. *Implementation Science*, 2011, 6(1): 117-123.
- [8] Ghosh R, Leman K. Predicting influential users in online social networks// *Proceedings of the 4th KDD workshop on Social Network Analysis*. Washington, DC, USA, 2010
- [9] Malmgren R D, Stouffer D B, Motter A E, Amaral L A N. A Poissonian explanation for heavy tails in e-mail communication. *Proceedings of the National Academy of Sciences*. 2008, 105(47): 18153-18158.

The Evaluation of E-commerce Efficiency in China using DEA-Tobit model: evidence from Taobao data

Zongwei Li^{1*}, Dongqing Luan¹, Yanhui Zhang², Peng Guo²

¹School of Management, Shanghai University of Science and Engineering, Shanghai, 201620, China

²School of business, East China University of Science and Technology, Shanghai, 200237, China

Abstract: Using the analytical framework of DEA-Tobit, this paper investigates the efficiency of e-commerce in China's provinces based on the cross-section data of 31 provinces in China and the data of e-commerce service providers from Taobao's open platform. The data envelopment analysis (DEA) is used to calculate the technical efficiency and scale efficiency. Furthermore the paper gives an empirical test on the relationship between the scale efficiency and influencing factors by using the censored Tobit model. The results show there are significant regional differences in the efficiency of e-commerce services in provinces of China, and the Real GDP per capita, the seller number on e-commerce platform, the retail sales and wholesale are important reasons for the different efficiency in each province of China. This study provides a domain-specific, integrative approach in evaluating the E-commerce development combining macro data from National Bureau of Statistics of China and micro data from taobao.com.

Keywords: E-commerce, DEA, Tobit, Efficiency Evaluation, Measurement

1. INTRODUCTION

For the past years, the rapid development of e-commerce in China has become the new engine of China's economic growth. Data from China e-Business Research Centre shows Chinese e-commerce transactions amounted to 13.35 trillion Yuan, an increase of 27.1% on year-over-year basis by the mid of 2017. Nevertheless, the e-commerce development of China is also facing some problems, such as development level of e-commerce in China's provinces is unbalanced, and the regional differences are huge^{[1][2]}. It is valuable to measure the development of e-commerce through the lens of integrative approach with an emphasis on empirical data. How should the regional efficiency of e-commerce development be measured? What factors determine the efficiency of e-commerce services? Measuring and assessing the efficiency of e-commerce efficiency is a difficult process and often involves methodological issues.

The main purpose of this paper is to investigate the efficiency of e-commerce services in China by applying the DEA-Tobit model. In particular, the paper concentrates on investigating: e-commerce efficiency measurement of 31 provinces in Chinese mainland combined with macroeconomic indicators and e-commerce platform indicators; Tobit regression analysis the determinants of e-commerce services efficiency.

The rest of the paper is arranged as follows. Section 2 shows the most relevant results in the literature and briefly outlines the related research using this method. Section 3 presents the DEA-Tobit structure design methodology and explains the data and variables. We provide the demonstration results for DEA-Tobit applied to 31 provinces in China. The paper offers our experimental results of DEA-Tobit model applied in China in section 4. Section 5 concludes. The results may serve as a reference for measuring the regional efficiency of e-commerce and e-commerce development promotion.

2. LITERATURE REVIEW

E-commerce can be regarded as the use of the Internet to conduct business transactions. In the context of

* Corresponding author. Email: lzw0118@163.com

E-commerce efficiency, the literatures focus on measurement framework and index system design, measurement methods and index acquisition and so on.

2.1 E-commerce development measurement

The NSF Workshop presented e-commerce development measurement index system firstly, including the E-Readiness, E-Intensity and E-Impact. OECD pointed out the index system of e-commerce development measurement including three categories: the readiness, the application and the impact of e-commerce. APEC put forward nine kinds of index system for the measurement of e-commerce level. But the measuring method involved too many indexes to collect data. The China Internet Research Center e-commerce issued index system with 9 major categories of 32 indicators, including trading, efficiency, security, IT infrastructure, network economy, human capital, development potential, policy Environmental, user satisfaction. Similarly it was not easy to demonstrate the application of the model in reality. The calculation method of e-commerce development is mainly the use of information index model (RITE model) now, including four elements of information to conduct a comprehensive evaluation.

2.2 E-commerce development index system

According to prior research in this stream, most of the index systems include economic and financial resources, trading volume, e-commerce system user, IT infrastructure and logistic & delivery system. We further discuss some of these studies to apply specific indicators that will help us to measure the level of development of e-commerce efficiency.

2.2.1 Economic and Financial Resources

It is generally assumed that Gross National Product (GNP) is an indicator of national wealth or drivers^[3]. Wealth has been shown to be the strongest individual factor driving e-commerce sales. Economic and financial resources appear to play a direct role as drivers of e-commerce. Thus, GDP per capita as a key factor is mentioned strong correlation with the measures of e-commerce development^[4].

2.2.2 Trading volume

Trading volume is significantly relative to e-commerce transaction, which depends on the nature of the good and offline trading volume^[5]. And E-commerce will also have a notable impact on trade in services. Liu and Wang(2009) utilize the transaction volume(wholesale and retail) and proportion of e-commerce as key index to demonstrate the e-commerce level^[6]. As a result of the rapid development of e-commerce, a lot of manufacturers have opted to choose the online and offline distribution channels at same time. Although the multi-channel system enables manufacturers to offer their customers products through synchronized channels, it is still a common phenomenon that the online channel is controlled by the traditional retailer or the supplier.

2.2.3 E-commerce system user

The number of Internet users in the world reaches 3.88 billion worldwide by June 2017 and is growing. In terms of e-commerce, the primary users are customers(buyers) or suppliers(sellers).Buyers and sellers could establish a genuine market price in e-commerce system^[7]. From a business ecology perspective, e-commerce business ecosystem includes five populations as bellows: Leader Population(e-commerce platform), key species, supportive species, related Population and parasitic species^[8]. Besides customers or suppliers, ISV(independent software vendor) as user in e-commerce system may develop some e-commerce relative system based on the e-commerce giant open platform^[9].

2.2.4 IT infrastructure

Kraemer et al.(2003) and Lawrence et al.(2010) point out that information infrastructure is an important facilitator of e-commerce development^{[10][11]}.In the diffusion of B2C e-commerce, the information infrastructure

is also considered as an important booster. It is not easy to shop online for people without reasonably quick Internet access. Thus, available IT infrastructure, such as ICT underlying framework, electronic data interchange (EDI) and open systems, is a prerequisite for e-commerce growth^[12].

2.2.5 Logistic & delivery system

Logistics capability is critical for e-commerce progress^[13]. Logistics service quality can exert an influence on online shopping experiences, loyalty, and satisfaction^[14]. It is critical to support the complexity capabilities and diversity of supply chains for e-commerce collaboration technologies^[15]. By using SEM model, Y Nurdani et al.(2016) shows that express delivery's service quality has positive influence on both customer satisfaction and trust of online shopping^[16]. In the Brazilian scenario, de Oliveira et al.(2017) evaluates the applicability of different solutions to urban e-commerce deliveries^[17]. Although automatic delivery stations have the potential to reduce delivery problems, it still relies on the transportation system to complete delivery.

2.3 E-commerce development evaluation approach

The index system of network economic measurement established by U.S. government and the Ministry of Commerce can help to understand the classification of information technology measurement. OECD, APEC, Harvard University and the Nordic countries basically adopt qualitative method. The OECD's research on knowledge economy brings the concept of "knowledge-based economy", which not only opens the way for the theory of knowledge economy but also sets up framework on the index system of knowledge economy. Various studies are conducted from different aspects of the network economy, the information, e-commerce, and digital issues in Japan respectively. Based on e-commerce firms, H. Joseph Wen et al.(2003) utilize a DEA model to evaluate the relative efficiency^[18]. In terms of the empirical research, Liu and Wang(2009) use AHP model to determine the weight of indicators and give the method of this index of e-commerce on city level^[19].

3. METHODOLOGY AND DATA

3.1 Research Design

In this paper, we combine the DEA and Tobit regressing model to appraise the e-commerce efficiency of 31 provinces in China. In the first stage, the indicators of e-commerce efficiency are defined. And then five group factors are selected to construct input indicators, respectively Economy level, Trading volume, User, IT infrastructure and Transportation. The relative e-commerce efficiency is then assessed by DEA model. Finally, we use Tobit regression model to examine the effect of inputs on the e-commerce efficiency score calculated.

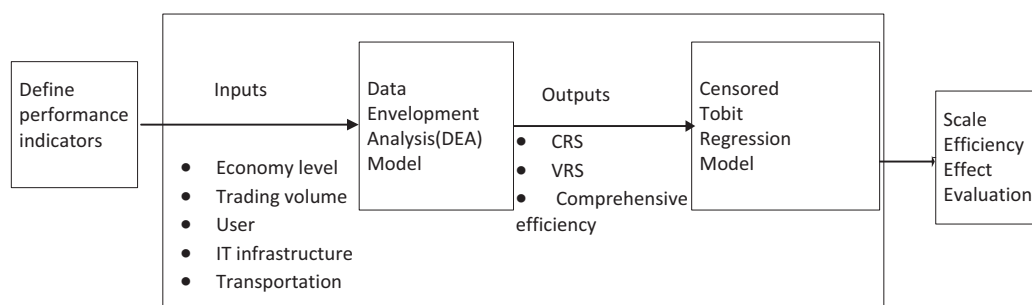


Figure 1. Diagram of methodological steps

3.2 Methodology

By using data from National Bureau of Statistics of China online database and Taobao's open platform, We apply DEA-Tobit analysis framework to evaluate the e-commerce development relative efficiency in China. The

relative efficiencies can be estimated by using DEA model ,which is a non-parametric approach combining various input/output. In the particular case of our research, there are 6 inputs and 2 outputs for 31 DMUs, which represent provinces in China. The model is expressed mathematically as follows:

$$\begin{cases} \text{Min } \theta \\ \theta, \lambda \\ -y_i + Y\lambda \geq 0 \\ \theta x_i - X\lambda \geq 0 \\ N_1 \lambda \leq 1 \\ \lambda \geq 0 \end{cases} \quad (1)$$

In order to further understand the impact factors of input-output efficiency, this paper carries on the Tobit regression analysis to the comprehensive efficiency. However, the efficiency index obtained by the DEA model is between 0 and 1, and if the ordinary least squares (OLS) method is adopted, the estimated values of the parameters will be biased and inconsistent. In this paper, the censored Tobit regression model limited dependent variable is used in the regression analysis, which was proposed by Tobin^[20]:

$$Z_i = \begin{cases} \beta^T X_i + \varepsilon_i, & \beta^T X^m + \varepsilon_i > 0 \\ 0, & \beta^T X^m + \varepsilon_i \leq 0 \end{cases} \quad (2)$$

3.3 Data and variables

Statistical data for 31 provinces in China are extracted from the National Bureau of Statistics of China online database and Taobao's open platform. Taobao is the world's largest C2C e-commerce platform based in China, started in May 2003. According to Alibaba, its online trading volume exceeded the milestone of ¥3 trillion in 2016, accounting for 10 percent of the overall retail volume in China. We selected two output variables: Number of Sellers and number of ISV (Independent Software Vendors) on Taobao's open platform and six input variables: Real GDP per capita, Total Retail Sales of Consumer Goods, Total Sales Value of Wholesale and Retail Trades, Number of Internet Users, Number of Port of Internet, and Freight Traffic.

According to Taobao's open platform, the number of sellers represents the number of active e-commerce shops providing transaction on taobao.com. The number of sellers is considered one of the most direct indicators of the e-commerce systems. Data from Alibaba group show that there are more than 8.5 million active sellers distributed in different provinces in mainland China in Taobao platform.

Number of ISV measures the number of Independent Software Vendors providing third-party services on Taobao's open platform, such as e-shop decoration services, marketing services, and data services etc. Number of ISV is treated as output variable because it represents E-commerce services capabilities. At present, the service market has gathered tens of thousands of providers to supply services for Taobao and Tmall's sellers, the transaction size has reached billions in 2016.

As input variables, Real GDP per capita, Total Retail Sales of Consumer Goods, Total Sales Value of Wholesale and Retail Trades represent Chinese provincial economic level and trading volume. The degree of development of e-commerce is positively correlated with the level of regional economic development.

Number of Internet Users and Number of Port of Internet are included in the model as input variables. Number of Internet Users represents the degree of access to e-commerce applications, and Number of Port of Internet is handled as an important indicator of information infrastructure. Both of the variables represent information technology facilities and Applications.

In terms of Logistics & delivery capability, Freight Traffic is also included in the model as input variable as

transportation infrastructure. The rapid growth of e-commerce in the United States benefits from the development of the logistics infrastructure and the rapid delivery system. Similar system is boosting in China and supports application of e-commerce, for example cainiao use big data and intelligence to promote the smart logistics upgrade. The variable description see table1.

Table 1. Variable description

New Variable	Variable
lnGDP	the Real GDP per capita
lnretail	Total Retail Sales of Consumer Goods
lnwholesale	Total Sales Value of Wholesale and Retail Trades
lnnetuser	Number of Internet Users
lnfreight	Freight Traffic
lnnetport	Number of Port of Internet
lnISV	Number of ISV
lnseller	Number of Sellers

4. RESULT AND DISCUSSION

We first present the descriptive statistics for the input and output variables and then discuss the results of the DEA model (see table 2).

Table 2. Descriptive statistics for input and output variables (the natural logarithm adopted)

statistics	lnGDP	lnretail	lnwholesale	lnnetuser	lnfreight	lnnetport	lnISV	lnseller
Max	11.51	9.155	11.04	8.853	12.89	8.109	10.31	14.44
Min	10.04	4.539	5.195	4.745	7.523	3.786	3.951	7.997
Average	10.68	7.606	9.056	7.311	11.40	6.722	7.438	11.62
SD	0.402	1.134	1.275	0.873	1.090	0.974	1.502	1.526
Coeff.of Variation	0.0377	0.149	0.141	0.119	0.0956	0.145	0.202	0.131

On the whole, the coefficient of variation is between 3.77% and 20.2%, and the coefficient of variation of lnISV is the highest(20.2%), thus the sample for the variables is homogeneous. DEA Model result shows as figure 2. The model CRS is obtained with average performance score of 0.961, VRS score of 0.975, then the scale efficiency is calculated as the ratio between efficiency scores in CRS and VRS models. Ten provinces out of 31 are scale efficient: Beijing, Tianjin, Ningxia, Tibet, Jiangxi, Fujian, Zhejiang, Shanghai, Guangdong, and Guangxi. Interestingly Ningxia and Tibet are usually regarded as underdeveloped provinces in western China, Therefore, the development level of E-commerce is also relatively backward. Seven provinces out of 31 are decreasing return to scale: Neimenggu, Xinjiang, Jilin, Qinghai, Shandong, Liaoning, and Jiangsu. This means that an increase in input will result in less output increase. Therefore, for these provinces, to enhance the service efficiency of e-commerce is critical, rather than blindly expand the scale of e-commerce. The remaining 14 provinces are increasing return to scale. In conclusion, these provinces need to increase investment in e-commerce and improve service efficiency.

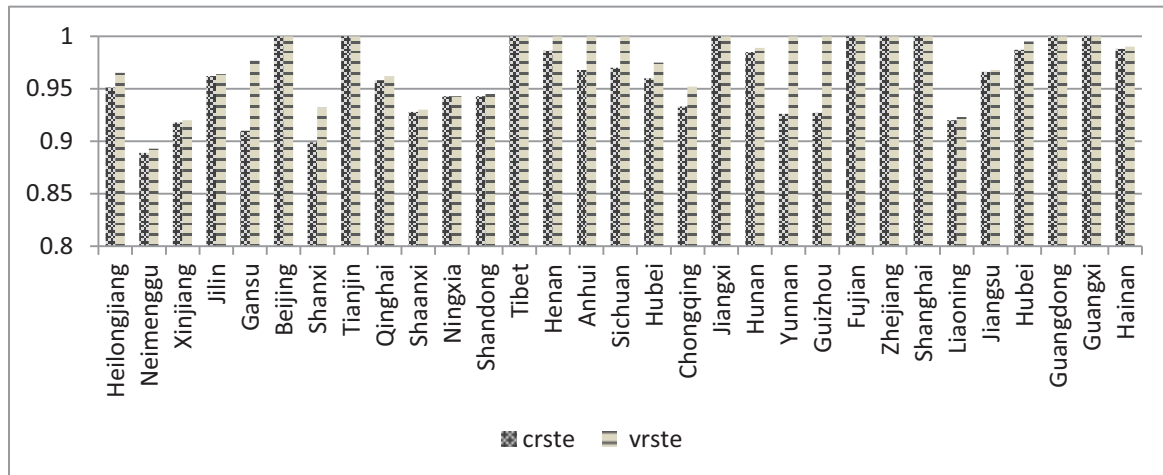


Figure 2. Efficiency scores for DEA model

Table 3 reports the Tobit regression results for scale efficiency. According to Table 3, four variables besides constant—lnseller, lnGDP, lnretail and lnwholesale—are statistically significant at the 5% level, while others show non-significance, both at the 5% and 1% levels. (1) The Real GDP per capita and Number of Sellers have a significant impact on the scale efficiency, and the influence coefficient is positive, indicating that the greater the Real GDP per capita and the greater the number of sellers can lead to the greater the scale efficiency value. This shows that increasing province's economic level and improving the number of e-commerce sellers will promote the efficiency of e-commerce development and enhance the competitiveness of e-commerce services. (2) There is a negative correlation Total Retail Sales of Consumer Goods, Total Sales Value of Wholesale and Retail Trades with scale efficiency, which indicates that the development of e-commerce would be damaged if a province overdevelop the offline trade, because channel conflict may occur when the supplier enters the online channel. (3) The Number of Internet Users and Number of Port of Interne have correlation with scale efficiency, but they are not significant, which indicate that the network infrastructure is no longer a bottleneck for the development of electronic commerce; there is a negative correlation of freight volume and comprehensive efficiency, but not significantly either.

Table 3. Tobit regression on efficiency related to inputs

Variable	Coef.	Std. Err.	t	P>t	[95%Conf. Interval]
lnseller	0.013874	0.005614	2.47	0.021**	0.002287 0.025461
lnGDP	0.068679	0.010328	6.65	0**	0.047364 0.089994
lnretail	-0.02463	0.008206	-3	0.006**	-0.04156 -0.00769
lnwholesale	-0.01847	0.005422	-3.41	0.002**	-0.02966 -0.00728
lnnetuser	0.020161	0.019262	1.05	0.306	-0.01959 0.059917
lnfreight	-0.00131	0.004768	-0.27	0.787	-0.01115 0.008536
lnnetport	0.000974	0.015165	0.06	0.949	-0.03033 0.032273
constant	0.307809	0.108742	2.83	0.009	0.083377 0.53224

** indicates 5% level of significance

Thus, based on the Tobit regression model, the DEA efficiency score of e-commerce development could be

improved through two main variables: $\ln \text{Seller}$ (Number of Sellers), $\ln \text{GDP}$ (the Real GDP per capita) and declined by offline trade volume. Hence, economic development level, e-commerce participation have a direct effect on increasing the efficiency score of e-commerce for regional development.

5. CONCLUSIONS

Due to the data from Taobao Open Platform and the National Bureau of Statistics of China, this paper presents a DEA model consisting of 6 inputs and 2 outputs to measure the efficiency of e-commerce services across provinces in China.

The contributions of this research are that we have proposed a "broader picture" of the efficiency of e-commerce initiatives by combining macroeconomic statistics with the micro-data from e-commerce platforms. The results of our study found some interesting phenomena, the efficiency level of e-commerce development in economically developing provinces is not necessarily lower than developed provinces, such as Tibet and Ningxia. In addition, it empirically demonstrated using censored Tobit model that some factors which affect the scale efficiency, such as the number of sellers on the platform and the level of economic development in the region.

Although this study provides meaningful implications for efficiency of e-commerce evaluation, it has some limitations and thus has further research issues. On one hand, if the representative indexes should be discussed and the long time sequence selection should be chosen in the future research, the results may be more accurate. Although the results show that the factors of economic development level, e-commerce participation have a direct effect on increasing the regional e-commerce efficiency e-commerce,. Recognizing that other factors may also play an important role, such as economic, political environment and tax policies. On the other hand, dynamic methods method can be used, for example Malmquist method. The research results of China's economically developed and developing provinces need to find more ways to allocate resources for their e-commerce development.

ACKNOWLEDGEMENT

This research was supported by Humanity and Social Science Youth Foundation of Ministry of Education of China under Grant 13YJCZH122 and Shanghai soft science project Funds under Grant 17692109000. Zongwei Li is grateful for Alibaba Huoshui Research Plan(2015).

REFERENCES

- [1] Yu Jinguo, Wang Li - hua&Li Na. (2010).E- Shops Spatial Distribution Rule:A Case Study of Taobao Net, Economic Geography, 30,1248-1253(in Chinese).
- [2] Tang Yinghan. (2015).A Spatial Difference Study on China's E-Commerce Development Level, Economic Geography, 35, 9-14 (in Chinese).
- [3] Dewan, S., & Kraemer, K. L. (2000). Information technology and productivity: evidence from country-level data. *Management Science*, 46(4), 548-562.
- [4] Oxley, J. E., & Yeung, B. (2001). E-commerce readiness: Institutional environment and international competitiveness. *Journal of International Business Studies*, 32(4), 705-723.
- [5] Kaplan, S., & Sawhney, M. (2000). E-hubs: the new B2B marketplaces. *Harvard business review*, 78(3), 97-106.
- [6] Liu, Y., & Wang, W. (2009). The reconstruction and empirical analysis of regional e-commerce development index. *Science of Science and Management of S. & T*, 30(7), 144-147.
- [7] Cai, G. G., Zhang, Z. G., & Zhang, M. (2009). Game theoretical perspectives on dual-channel supply chain competition with price discounts and pricing schemes. *International Journal of Production Economics*, 117(1), 80-96.

- [8] Huang, L., Hu, G., & Lu, X. (2009). E-business ecosystem and its evolutionary path: the case of the Alibaba group in china. *Pacific Asia Journal of the Association for Information Systems*, 1(4).
- [9] Chen, J., Wang, C., Liu, F., & Wang, Y. (2017). Research and Implementation of a Software Online Testing Platform Model Based on Cloud Computing. In *Advanced Cloud and Big Data (CBD)*, 2017 Fifth Internataional Conference, 87-93.
- [10] Gibbs, J., Kraemer, K. L., & Dedrick, J. (2003). Environment and policy factors shaping global e-commerce diffusion: A cross-country comparison. *The information society*, 19(1), 5-18.
- [11] Lawrence, J. E., & Tar, U. A. (2010). Barriers to e-commerce in developing countries. *Information, society and justice journal*, 3(1), 23-35.
- [12] Iacovou, C. L., Benbasat, I., & Dexter, A. S. (1995). Electronic data interchange and small organizations: Adoption and impact of technology. *MIS quarterly*, 465-485.
- [13] Joong-Kun Cho, J., Ozment, J., & Sink, H. (2008). Logistics capability, logistics outsourcing and firm performance in an e-commerce market. *International journal of physical distribution & logistics management*, 38(5), 336-359.
- [14] Ramanathan, R. (2010). The moderating roles of risk and efficiency on the relationship between logistics performance and customer loyalty in e-commerce. *Transportation Research Part E: Logistics and Transportation Review*, 46(6), 950-962.
- [15] Tan, P. S., Lee, S. S. G., & Goh, A. E. S. (2012). Multi-criteria decision techniques for context-aware B2B collaboration in supply chains. *Decision Support Systems*, 52(4), 779-789.
- [16] Nurdani, Y., & Sandhyaduhita, P. I. (2016). Impact of express delivery service quality towards repurchase intention by B2C and C2C: A case of Indonesia. In *Advanced Computer Science and Information Systems (ICACSIS)*, 2016 International Conference , 221-227.
- [17] de Oliveira, L. K., Morganti, E., Dablanc, L., & de Oliveira, R. L. M. (2017). Analysis of the potential demand of automated delivery stations for e-commerce deliveries in Belo Horizonte, Brazil. *Research in Transportation Economics*.
- [18] Wen, H. J., Lim, B., & Lisa Huang, H. (2003). Measuring e-commerce efficiency: a data envelopment analysis (DEA) approach. *Industrial Management & Data Systems*, 103(9), 703-710.
- [19] Liu Yue, Wang Wenqing. (2009). The Reconstruction and Empirical Analysis of Regional E—Commerce Development Index. *Science of Science and Management of S.& T*, 30.7:144-147(in Chinese).
- [20] Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, 24-36.

The Differences of Online Review Textual Content:

A Cross-Cultural Empirical Study

Liyi Zhang^{1}, Li Li²*

¹School of Information Management, Wuhan University, China

²School of Information Management, Wuhan University, China

Abstract: This research do a cross-cultural study by examine the differences of online review textual content between China, America and Australian. Through the online review text segmentation, classifying words by coding schema, calculating word proportion of each category, the research analyzes the differences of online review textual content from the aspects of textual type, content preference and textual emotion. The research finds that, cultural differences have significant effect on the online review textual type, Chinese customers prefer to describe objective facts while American & Australian customers prefer to describe subjective feelings; For textual emotion, Chinese customers prefer to express negative emotions while American & Australian customers prefer to express positive emotions. But cultural differences show no significant effect on the online review content preference.

Keywords: culture, online review textual content, textual type, content preference, textual emotion

1. INTRODUCTION

Online reviews are the important information resource for both merchants and consumers in the Internet era. With the accelerating process of globalization and the rapid development of cross-border E-commerce, many online reviews from online platforms are not limited to a single country. And especially for the online travel platforms, because tourism and hotel industry are more accessible to customers from different countries, online travel reviews are from more diverse sources. Different countries have different cultural characteristics, thus, cultural factors have become a noticeable difference in online reviews.

Online platforms present both the ratings and the review texts of products or services, although ratings can provide the overall evaluation, the textual content of review texts can provide more details, such as descriptions of specific function of products and the consumers' emotional attitude, these detailed descriptions help people get a more thorough understanding of products or services when they read online reviews. Compared to the online rating, therefore, online review textual content can provide better decision support, it's particularly important to consider textual content when studying the impact of culture on online reviews. In this research, the online reviews of China, America and Australia are selected as the research object, a comparative analysis of differences of online review textual content are conducted, in order to explore the impact of culture on online review textual content. The results will help both merchants and consumers better understand the online reviews and guide them to make decisions.

2. LITERATURE REVIEW

Some researchers have already put forward specific cultural theories during the study of culture, one of the most influential theories is the cultural dimension theory proposed by Hofstede. The theory points out that cultural dimensions include power distance, uncertainty avoidance, individualism / collectivism, masculinity / femininity, long-term orientation / short-term orientation, self-indulgence / restraint ^[1]. Besides, considering the accuracy and clarity of the transmission and reception of information in the process of communication, Hall has

* Corresponding author. Email: lyzhang@whu.edu.cn (Liyi Zhang).

put forward the high context / low context theory, the theory points out that the expression of information is direct and unambiguous in the culture preferring low context communication, whereas in the culture preferring high context communication the expression of is usually vague and implicit ^[2].

Hofstede's culture dimension theory and Hall's high context / low context theory have been used to study the impact of culture on online behaviors and information systems, such as the impact of culture on online search process ^[3], the impact of culture on e-WOM ^[4], and the impact of culture on online promotion effect ^[5]. Meanwhile, there are also studies just select different countries to compare the differences of online behavior. For example, Sakarya and Soyer compare the differences of consumption values (utilitarian / hedonic) and online shopping behavior (online purchase frequency, goods purchased, online site type used and payment method) between Turks and Britons ^[6].

At present, among the studies about the culture impact on online users and information systems, the studies about the culture impact on review behaviors are still scarce. Lai et al. from the perspective of individualism / collectivism study the online review textual content between China and America, and point out that American reviews are more likely to express their own opinions on products and contained more recommendations ^[7]. Fang et al. based on the Hofstede's cultural dimension theory, compare the online review differences between China and America in the average number of reviews, average number of votes per review, average word counts per review, percentage of 1-star reviews and so on ^[8]. Hong et al. study the impact of culture (individualism / collectivism) on the tendency of consumers to conform to prior opinion and the emotionality of the review text, find that ratings from collectivists are closer to the average prior rating and collectivists express less emotions in reviews ^[9]. Zhou, Xia and Zhang exploit developed opinion mining techniques to study the online review textual content between China and America and point out that Americans express their opinions more directly and pay more attention to product details ^[10].

3. RESEARCH METHOD

3.1 Data

With the prevalence of cross-border tourism and the maturity of online tourism platform, the impact of culture on online tourism reviews has become increasingly prominent. Bangkok is one of the most famous tourist destinations among many tourist cities. Well-developed tourism and moderate consumption price attract tourists from all over the world to favor Bangkok. This research selects AGODA which is the well-known online hotel booking website in the Asia Pacific region, and crawls all the reviews of the Bangkok Hotel by the end of June 2015. For each review, the data includes the user name, the travel type, the user nationality, the review time, the review theme, the review of advantages, the whole review content, and the rating.

3.2 Measurement

3.2.1 Culture

The cultural differences between China and America have always been the focus of many researchers, and they find that there are obvious cultural differences between the two countries, for example, China is a typical collectivist country, and attaches great importance to the status of rights and prefer high-context communication. However, America is just the opposite. Countries which have the similar cultural characteristics with America include Australia, Canada and so on.

Nowadays, China and America are the world's most populous country with advanced information technology, which has the large number of netizens and extensive influence. This research still chooses China and America to represent culture. Considering the demographic factors that the number of Chinese reviews will

be much larger than the number of America reviews, Australia which has the similar cultural characteristics with America also be selected as the research object to balance sample sizes.

3.2.2 Textual content

In the existing research on textual content and semantic information, Ghose and Ipeiritis point out that textual content can be divided into subjective information and objective information from a stylistic point of view, and they have different effects on product sales and review usefulness^[11]. Berger and Milkman focus on the subjective information, studied the impact of emotional tendencies and specific emotions (anger, sadness, etc.) on social transmission, find that positive emotions, as well as specific high-arousal emotions, are more likely to enhance transmission^[12]. Ludwig, Ruyter, and Friedman study the semantic content and linguistic style properties of online reviews, find that the increase of positive affective content in the review text and the similar linguistic style between product interest group will enhance the online retail sites' conversion rates^[13].

In the field of online tourism, by analyzing textual content of online reviews, researchers can explore the most frequently used words in online reviews, and further combine with the rating data to find topics which affect customer satisfaction most^{[14][15]}.

Based on the above research, this research analyses online review textual content from three aspects: textual type, content preference, textual emotion. Textual type includes: the description focusing on objective facts or the description focusing on subjective feelings. Content preference includes: the description preferring tangible objects or the description preferring intangible services. Textual emotion includes: the description with positive emotions or the description with negative emotions.

3.3 Research model

$$\text{Online review textual type} = \alpha_1 * C + \alpha_2 * R + \sum a_i * T_i + \sum b_j * H_j + \sum c_k * M_k + \varepsilon \quad (1)$$

$$\text{Online review content preference} = \beta_1 * C + \beta_2 * R + \sum a_i * T_i + \sum b_j * H_j + \sum c_k * M_k + \varepsilon \quad (2)$$

$$\text{Online review textual emotion} = \gamma_1 * C + \gamma_2 * R + \sum a_i * T_i + \sum b_j * H_j + \sum c_k * M_k + \varepsilon \quad (3)$$

C represents culture, it's a dummy variable, 1 indicates china, and 0 indicates America or Australia.

R represents the rating per review, we control it for different ratings reflect different degree of customer satisfaction and will affect the emotional tendencies of online reviews.

T_i represents the tourism type, we control it for tourists with different tourism type will have different expectations for the quality of the hotel and services.

H represents the different hotel, we control it to reduce the impact of different hotel facilities and service levels on customers' perception.

M represents the time effect, the use of month as a time variable because of the possible of unobservable impact and trend on online reviews at different time periods (for example, the decline in hotel quality).

3.4 Textual content measurement index

The difficulty of this research is how to measure textual type, content preference and textual emotion of online review textual content, and how to avoid deviations caused by grammatical differences between Chinese and English. Based on the existing research method^{[9][10]}, and combined with research purpose, this research solves the problem from the lexical level, using the content analysis method to construct a word coding schema about textual type, content preference and textual emotion. Then through the online review text segmentation, classifying words by coding schema, calculating word proportion of each category, the data of online review textual content can be quantified.

Online review textual type: online review textual type includes the description focusing on objective facts or the description focusing on subjective feelings. This research calculates the number of words that describe the

products and services accounts for the proportion of review length (total number of words), and the number of words that describe emotion accounts for the proportion of review length (total number of words) to quantify textual type.

Online review content preference: online review content preference includes the description prefer tangible objects or the description prefer intangible services. Tangible objects are related to physical facilities, while intangible services are related to all services and quality of services. This research calculates the ratio of words related to tangible objects and the ratio of words related to intangible services to quantify textual type.

Online review textual emotion: on emotion measurement, positive and negative words that express subjective feelings, as well as positive and negative words describing products or services both can reflect the emotional tendency of customers. This research calculates the ratio of positive words (negative words) to review length (total number of words) to quantify textual emotion.

3.5 Coding schema

The words related to research purpose need to be filtered manually and simultaneously through an iterative process. Based on the existing research and research purpose, this research constructs a coding schema to ensure the rationality and effectiveness of the filter work (Figure 1).

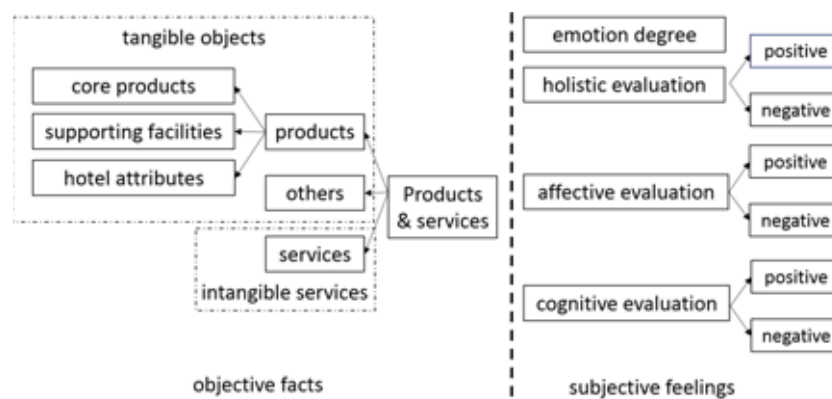


Figure 1. Coding schema

When describing products and services, core products refer to room, bed, bathroom and other core products of room. Supporting facilities refer to environment or facilities such as pool, breakfast, lobby. Hotel attributes include location, view, price, etc. And others refer to things that have nothing to do with the hotel, such as taxi, airport. Services refer to words such as service, check in, staff. When describing emotion, cognitive evaluation expresses the feeling of specific products and services, such as clean, delicious and friendly. Affective evaluation reflects the customer's mood, such as happy, like, enjoy and fun. Holistic evaluation includes good, perfect, excellent, etc. And the degree of emotion refers to the words such as very, too, extremely.

4. DATA PREPROCESSING & EMOTION CLASSIFICATION

In order to ensure data balance that each hotel has the same amount of data in China, America and Australia, this research finally chooses 7 hotels among all the hotels of Bangkok, and selects all Chinese, American, Australian reviews, then eliminates empty reviews and short reviews (such as: good!), finally gets 1073 reviews. There are 549 Chinese reviews, 254 American reviews and 270 Australian reviews respectively. After data acquisition, Chinese and English word segmentation software are utilized to extract words of all reviews and count word frequency, 3440 Chinese words and 3077 English words are got.

Any word that related to research purposes should be included, and some exceptions should be considered, such as stop words; and generic nouns such as “size”, “people”, “effort”, “fault”, etc. due to the lack of specificity, these words need to be filtered out.

Coding is primarily conducted by two researchers respectively to classify words according to the coding schema. When two researchers are in disagreement of any word classification, the third researchers joins and discusses together. Misspelling and English singular-plural pair should be correct during coding.

Coding begins from the high frequency word, and stops at the point when words appeared generic and irrelevant to the coding schema, resulting in 193 Chinese words and 195 English words. Then, coded words are marked in the original online reviews, words and word frequency involved in each review can be obtained. Through calculating word proportion of each category in coding schema, textual content can be quantified..

To judge the positive or negative emotional tendency of emotion words, researchers bring words into the context of the original online reviews to judge it artificially.

5. DATA ANALYSIS

The research calculates word proportion of each measurement index to quantify textual content, and table 1 shows descriptive analysis of measurement indexes. The proportion of words describing objective facts to review length (objective facts / review length), the proportion of words describing objective facts to total number of words (objective facts / total number of words), the proportion of words describing subjective feelings to review length (subjective feelings / review length), the proportion of words describing subjective feelings to total number of words (subjective feelings / total number of words) measure the online review textual type. And, “tangible objects / review length”, “tangible objects / total number of words”, “intangible services / review length”, “intangible services / total number of words” measure the online review content preference. And, “positive emotion / review length”, “positive emotion / total number of words”, “negative emotion / review length”, “negative emotion / total number of words” measure the online review textual emotion.

Table 1. Descriptive analysis

	China					America & Australian				
	N	Min.	Max.	Avg.	Std. error	N	Min.	Max.	Avg.	Std. error
objective facts / review length	549	0.00%	57.14%	19.62%	9.17%	524	0.00%	42.86%	13.69%	6.77%
objective facts / total number of words	549	0.00%	100.00%	58.39%	17.31%	524	0.00%	100.00%	51.15%	16.69%
subjective feelings / review length	549	0.00%	55.56%	14.79%	9.51%	524	0.00%	57.14%	13.52%	8.73%
subjective feelings / total number of words	549	0.00%	100.00%	41.43%	17.22%	524	0.00%	100.00%	47.37%	16.98%
tangible objects / review length	549	0.00%	54.55%	16.53%	8.61%	524	0.00%	42.86%	11.71%	6.54%
tangible objects / total number of words	549	0.00%	100.00%	49.43%	18.34%	524	0.00%	100.00%	43.91%	17.99%
intangible services / review length	549	0.00%	57.14%	3.09%	4.87%	524	0.00%	23.08%	1.98%	2.96%

5.1 Culture effect on textual type

When studying the culture effect by textual type measurement indexes, the result indicates that culture has a significant effect on textual types. Besides, the coefficient of “objective facts / review length” and “objective facts / total number of words” is positive, indicating that compared with American and Australian culture, Chinese culture is more descriptive of objective facts. And, the coefficient of “subjective feelings/ total number of words” is negative, indicating that Chinese culture has less description of emotion. But, the culture effect on “subjective feelings / review length” is not much significant, the reason may be that, coding stops at the point when words appear generic and irrelevant to the coding schema, but due to the diversity of emotion words, many low frequency emotion words are eliminated, so the analysis result has been disturbed.

Table 2. Culture effect on textual type

	objective facts /review length	objective facts/total number of words	subjective feelings /review length	subjective feelings /total number of words
constant	24.051**	85.835***	3.972	13.989
culture	6.113***	7.812***	1.042 ⁺	-6.473***
rating	-0.111	-3.039***	1.497***	3.071***
travel type	YES	YES	YES	YES
hotel	YES	YES	YES	YES
sample size	1073	1073	1073	1073
R square	0.206	0.188	0.145	0.17

Note: +<0.1, *p<0.05, **p<0.01, ***p<0.001

5.2 Culture effect on content preference

When using “intangible services / review length” to measure the content preference, the culture effect on content preference is significant, the positive coefficient indicates that compared with American and Australian culture, Chinese culture describes more intangible services in online reviews. But, when using “intangible services / total number of words” to measure the content preference, the culture effect is not significant. Besides, when using “tangible objects / review length” or “tangible objects / total number of words” to measure the content preference, although culture effect is significant, the positive coefficient indicates Chinese culture describes more tangible objects in online reviews. Thus, Chinese culture describes more of both tangible objects and intangible services. Cultural differences show no significant effect on the online review.

Table 3. Culture effect on content preference

	tangible objects / review length	tangible objects /total number of words	intangible services /review length	intangible services /total number of words
constant	23.042**	78.555***	1.008	7.281
culture	4.998***	6.39***	1.114***	1.422 ⁺
rating	0.03	-2.165***	-0.141	-0.874**
travel type	YES	YES	YES	YES
hotel	YES	YES	YES	YES
sample size	1073	1073	1073	1073
R square	0.176	0.141	0.131	0.11

Note: +<0.1, *p<0.05, **p<0.01, ***p<0.001

5.3 Culture effect on textual emotion

When using “positive emotion / review length” or “positive emotion / total number of words” to measure the positive textual emotion, the culture effect on positive textual emotion is significant, and the positive coefficient indicates that compared with American and Australian culture, Chinese culture expresses less positive emotion. When using “negative emotion / review length”, the culture effect on negative textual emotion is significant, and the negative coefficient indicates that Chinese culture expresses more negative emotions. Thus, for textual emotion, Chinese customers prefer to express negative emotions while American & Australian customers prefer to express positive emotions.

Table 4. Culture effect on textual emotion

	positive emotion / review length	positive emotion / total number of words	negative emotion / review length	negative emotion / total number of words
constant	-4.376	-14.673	11.862***	36.875***
culture	-1.697**	-11.189***	0.431**	0.417
rating	1.669***	4.703***	-0.515***	-2.205***
travel type	YES	YES	YES	YES
hotel	YES	YES	YES	YES
sample size	1073	1073	1073	1073
R square	0.175	0.291	0.228	0.237

Note: +<0.1, *p<0.05, **p<0.01, ***p<0.001

6. CONCLUSION & DISCUSSION

This research examines differences of online review textual content between different culture and select China, America and Australian as research object. The research finds that, cultural differences have significant effect on the online review textual type, Chinese customers prefer to describe objective facts while American & Australian customers prefer to describe subjective feelings; For textual emotion, Chinese customers prefer to express negative emotions while American & Australian customers prefer to express positive emotions. But cultural differences show no significant effect on the online review content preference.

Different from the existing research, this research obtains data from the highly diverse online review text, and especially studies the different characteristics of textual content. The method that calculating word proportion of each measurement index to quantify textual content overcomes the impact of grammatical differences to some extent.

The results show some guiding significance. On the one hand, Chinese customers focus on describing objective facts and can provide more online reviews on specific hotel products, whereas American and Australian customers are more concerned about personal feelings. Thus, for hotel managers, more attention should be paid to Chinese online reviews when taking steps to improve hotel facilities and management, the details will help to understand deficiency. Meanwhile, hotel staff should be required to concern more about the emotional state of American and Australian customers in the service contact and improve their emotional experience. For hotel customers, when referring to the online reviews, more information about specific conditions of hotel products will be revealed from Chinese customers' online reviews, and their reviews are more objective. On the other hand, because Chinese customers are more inclined to express negative emotions, hotel managers need to be more careful in dealing with Chinese's feedback and complaints, managers should respond promptly and properly so as to avoid further complaining in online reviews. Therefore, the impact of

cultural factors should be taken into consideration when referring to online reviews while making business marketing strategy or consumer consumption decision. This research selects China, America, Australia as research objects, moreover, countries similar to Chinese culture also include: Japan, Russia, Thailand, etc. Most of them are in Asia; and countries similar to American and Australia culture include: Canada, New Zealand, and other European countries, the results can be further applied to these countries.

There are also some limitations in this research. First, considering the sample balance of each hotel, only 7 hotels of 1073 online reviews are selected, under the background of extensive big data research and application, if the sample size can be further expanded, the results will be more accurate. Second, although most culture comparative researches have selected China and America as the research object, if more representative countries can be choosed, the results will be more objective. Finally, if natural language processing method can be used to identify emotions, the error will be reduced more effectively.

ACKNOWLEDGEMENT

This work is supported by the MOE Project of Key Research Institute of Humanities and Social Science of China under Grant 14JJD870002.

REFERENCES

- [1] Hofstede G. (1980). *Culture's Consequence*. Beverly Hills: SAGE Publications Inc, 25.
- [2] Hall E T. (1976), *Beyond Culture*. Garden City, New York: Anchor Press, 34.
- [3] Vuylsteke A, Wen Z, Baesens B, et al. (2010). Consumers' search for information on the internet: how and why china differs from western europe. *Journal of Interactive Marketing*, 24(4): 309-331.
- [4] Luo C, Wu J, Shi Y, et al. (2014). The effects of individualism–collectivism cultural orientation on eWOM information. *International Journal of Information Management*, 34(4): 446-456.
- [5] Phang C W, Tan C H, Sutanto J, et al. (2014). Leveraging O2O commerce for product promotion: an empirical investigation in mainland china. *IEEE Transactions on Engineering Management*, 61(4): 623-632.
- [6] Sakarya S, Soyer N. (2013). Cultural differences in online shopping behavior: Turkey and the United Kingdom. *International Journal of Electronic Commerce Studies*, 4 (2): 213-237.
- [7] Lai J, He P, Chou H M, et al. (2013). Impact of national culture on online consumer review behavior. *Global Journal of Business Research*, 7(1): 109-115
- [8] Fang H, Zhang J, Bao Y, et al. (2013). Towards effective online review systems in the Chinese context: A cross-cultural empirical study. *Electronic Commerce Research & Applications*, 12(3): 208-220.
- [9] Hong Y, Huang N, Burtch G, et al. (2016). Culture, conformity and emotional suppression in online reviews. *Journal of the Association for Information Systems*, Forthcoming(11).
- [10] Zhou Q, Xia R, Zhang C. (2016). Online shopping behavior study based on multi-granularity opinion mining: China versus America. *Cognitive Computation*, 8(4): 587-602.
- [11] Ghose A, Ipeirotis P G. (2011). Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics. *IEEE Transactions on Knowledge & Data Engineering*, 23(10):1498-1512.
- [12] Berger J, Milkman K L. (2012). What makes online content viral? *Journal of Marketing Research*, 49 (2): 192-205.
- [13] Ludwig S, Ruyter K D, Friedman M, et al. (2013). More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing*, 77(1): 87-103.
- [14] Berezina K, Bilgihan A, Cobanoglu C, et al. (2015). Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *Journal of Hospitality Marketing & Management*, 25(1): 1-24.
- [15] Xiang Z, Schwartz Z, Gerdes J H, et al. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44(44): 120–130.

e-Control of Online Customers: When Companies Make Customers Compliant ... Lessons for 21st Century China

Edward Kasabov

(Huddersfield Business School, Advanced Institute of Management Research, University of
Exeter Business School, United Kingdom)

Alexander James Warlow

(Institute of Direct and Digital Marketing, United Kingdom)

Abstract: With the advent of new technologies, an increasing number of companies have been redefining the manner in which they interact with their customers. Some of the fundamental precepts of marketing thinking, including what we read in our textbooks - such as customer centricity and customer orientation - have been questioned by practitioners. Companies treat customer demands as illegitimate until proven otherwise. Literature advice of individualised problem solving, and the adoption of an organic approach to complaint management, has been overturned and replaced with control and disciplining through the use of scripted rules and standard, cost-effective complaint processes. As a result, customers defined by the company as difficult and expensive are subordinated, whereas customers who comply with company expectations benefit from the low cost and better service levels, and appear willing to accept these new rules of engagement between customer and provider.

1. INTRODUCTION

Over the last decade or so, a new type of company has emerged, being enabled by deregulation, privatisation, liberalisation of markets, but most significantly by technological advances including the Internet, intranets and the growth in call-centres. The software systems, providing real-time direct communication with the customer has allowed these companies to employ a strategy of 'Customer Compliance' in place of the traditionally practiced 'Customer Centricity' which suggests that each complaint and complainer should be dealt with individually.

The press, bloggers, and even some academics argue that 'Customer Compliance' companies treat their customers as if they are NOT right and so the action of such companies is unfair and only to the benefit of the company, typified by an article in *Harvard Business Review* where McGovern and Moon (2007) suggest that 'many companies' allegedly 'infuriate their customers', 'bleeding them' with the commissions and fees, 'confounding them' with the small print regulations, and 'penalizing' them in whatever way they could^[1]. To McGovern and Moon, customers make substandard purchase decisions because they are 'confused'.

Our research fundamentally disagrees with such an assessment. To us, it was marketers who were wrong by incorrectly identifying what they thought customers did or did not want. We argue, and our research supports the conclusion, that a large majority of the customers of businesses practising 'Customer Compliance' are satisfied because the goods and services are supplied at a price and a service level which the customer could not previously afford. Importantly while a small minority of complaining customers may be excluded, the majority receive a better service. We question the ideas that these companies practising 'Customer Compliance':

1. 'accidentally' 'end up' with such strategies,

2. detract value, and ultimately risk failure through customer dissatisfaction,
3. should reform their ways.

‘Customer Compliance’ businesses have appeared across sectors: low cost air travel, banking, insurance and financial services (Barclays Banks, Egg, Directline and E-sure), retail sector (Tesco and IKEA), telecommunications, TV and broadband supply (BT, Virgin and Sky), tourism (Hertz and Holiday Autos, Lastminute.com, Expedia, and Holiday Extras). Central to the strategy of these businesses is that the customer complies with the company systems. In return, the customer benefits from lower cost of product or service provision and a good level service which may incorporate timeliness, quick delivery, no quibble return systems and 24/7 access to the service.

By studying this new breed of businesses, we show how, by ‘disciplining’ customers, making them compliant and even excluding some, these businesses:

1. meticulously and consciously plan and implement ‘compliance’ and ‘exclusion’ strategies,
2. far from adopting an ‘antagonistic strategy’, add value by offering low price, good service by identifying customer needs,
3. provide examples of best practice for other businesses to emulate, particularly those that they have displaced.

We define ‘Compliance Marketing’ in terms of ‘disciplining’ customers and exclusion, with exclusion marketing implying companies being selective about the customers with whom they choose to work. Such companies tend to select and maintain relations with customers who cost the least to serve. It could be argued that this is a more democratic relationship between customers and businesses than was the case in traditional businesses which spend disproportionate resources to communicate with and placate few vociferous complainers who absorb company resources and time which can be better directed towards improving the service to the majority of the company’s customer base. If the company deems that a customer is a complainer and is ‘not always right’ then they would be excluded to the benefit of the majority.

2. DEFINING CUSTOMER COMPLIANCE AND EXCLUSION

Businesses have always excluded customers through various means, usually based on segmentation. Most traditional practices designed at generating additional profits were hidden and customers were not aware of the strategies used to generate profits and commissions let alone understand them. The new ‘Customer Compliance’ businesses often make their charges and commissions very transparent and thus have opened themselves up to analysis and criticism. However, the picture of this new model of customer-provider relations is considerably more complex and nuanced. The use of language such as ‘companies extracting value’, ‘preying on customers’, and of customers developing ‘pent-up hostility’, as found in McGovern and Moon and similar writings, demonstrates a misunderstanding of what is actually happening in practice. The ‘compliance’, ‘disciplining’ and ‘exclusion’ that we describe provide as much if not more advantage to the customer than to the company. It is only through offering good customer value and service that these companies have been able to grow and expand, often at the expense of traditional so-called ‘Customer Centric’ businesses.

The companies that we describe adopt a ‘take it or leave it’ approach, usually led by entrepreneurs who have little respect for what the press, academics or the establishment think. They focus their company’s resources and efforts on providing novel, lowest cost services and products to customers and involve their services at a phenomenal rate because it is relatively simple to change back-office software. Low-cost airlines for instance openly ‘train’ their customers and the general public to adhere to their rules in return for which the customer benefits from low price, good and sometimes novel service. Similar applications of regimented procedures are found in IKEA’s store layout which forces the customers to follow a specific path through the store. Online banking and financial service providers require that customers input the correct (and complete) information before they are able to proceed with a transaction.

Such regimented procedures are enabled partly by the de-personalisation of the interactions between providers and customers. Exchanges and ‘relations’ these days are with computers and databases, and less so with the staff of providers. Direct lines have computerised telephone systems which take the calling customer to a computerised answering machine where much of the information they need can be obtained automatically. Only in extreme cases is the customer ‘allowed’ to get through to a call centre operative. Call centres are one common device that such companies use in order to apply regimented interactions, especially in the area of managing customer complaints. Unwanted calls are screened out, as typically practiced by online insurance and financial service providers. Most call centres operate a policy of not allowing a complaining customer to escalate a complaint by asking to speak to a manager.

Another factor which observers and academics appear to have failed to notice is that these new businesses generate considerable value to customers by providing customers with multiple opportunities to actively engage in ‘co-production’, ‘customisation’ and do-it-yourself marketing. However, this involvement of the customer is constricted within parameters pre-defined by the company which, nonetheless, provides good service and may generate loyalty. Examples include easyJet ‘allowing’ their customers to design their own service by optionally booking priority boarding. The take-up of this option has turned into an important revenue generator, suggesting that it is clearly of value to many customers – an example of generating additional profit while addressing the idiosyncratic needs of individual clients. Similarly, IKEA encourages the customer to carry out part of the service themselves, by selecting, loading and transporting self-assembly furniture themselves – activities that traditional stores would have carried out by employing warehouse staff. The savings are passed on to the company’s customers in the form of lower costs.

There is also a certain degree of ‘dis-intermediation’ practised by ‘Customer Compliance’ companies which encourage customers to talk to each other and thus create virtual communities and chat forums. Customer recommendations are utilised by Amazon in online retail and by eBay in e-auctions. It is particularly important on holiday, hotel and travel booking sites.

‘Compliance Companies’ appear ready not only to pay little attention to the demands of serial complainers but also abandon customers whom they perceive as troublesome and of little value to the company, both with respect to direct (payment, volume, quality, safeguarding) and indirect value-creation (innovation, information, access, motivation). They have abandoned the ‘customer is all was right’ ethos. Such relationships are not

wanted as they are of little value. Alternatively, they may develop into unwanted relationships, over time, as the initial expectation of potential may never be realised.

How this works to the benefit of the majority of customers and to the company is illustrated in the way that Ryanair ‘shows no mercy if you’re late’ (McGovern 2008)^[2]. We also agree with McGovern’s comment that ‘That’s terrible for you. But is it so terrible for the 200 people on the plane who were on time? If Ryanair waited for you, they’d make you very happy. But there’d be 200 people who’d be somewhat unhappy’.

3. INGREDIENTS OF COMPLIANCE AND EXCLUSION MARKETING

The first major ingredient of compliance and exclusion marketing is the unwillingness of such companies to follow the ethos of ‘the customer is always right’ by listening to complaining customers and especially to serial complainers, typically by getting rid of ‘bad’ customers or by ignoring their complaints. This is often achieved by the customer eliminating themselves because they will not comply with the systems, rather than the company having to take positive action to remove them. Such successful companies have considerably redesigned service recovery and customer complaint management so that to exclude unwanted customers.

Companies practicing compliance and exclusion marketing do not view the feedback of complainers as an invaluable form of market and marketing research, thus clearly breaking with standard assumptions in the marketing literature. Instead, most seem to use online systems which provide research and intelligence gathering in real time, for instance through the response obtained from the pattern of the sales of their products and services. While marketing texts habitually remind their readers that complainers are a valuable source of information and help companies with intelligence gathering, contacting and talking to complainers is actually very costly and is therefore a poor ROI. What these companies have realised is that in order to placate one complainer they have to disadvantage the majority of good customers or to add to the cost of their service provision.

The companies studied here do not necessarily view complaints as legitimate. This may be explained by the effect that new technologies have had on modes, intensity and types of complaints. For instance, it appears that people are more likely to complain online than by other means, with research showing that 43 percent of U.S.-based Web users feel less inhibited online. ‘Customer Compliance’ companies may thus expect a greater number of complaints and also complaints of higher intensity because most do their business on-line. There is also the question of serial complainers setting up incidents to claim compensation, largely facilitated by advice available on the internet. This has prompted companies practicing compliance and exclusion marketing such as Thompson TUI travel to prevent serial or fraudulent complainers from rebooking.

The effect to regular, ‘good’ customers of pandering to complainers would be negative. Listening to and satisfying complaints and especially serial complaints which customers expect to get escalated may increase the cost of products or of service provision to all customers. This may result in substandard, expensive or otherwise poorer service across the board and for every customer. The example of holding a plane for one late passenger while 180 other passengers sit in the plane, with the plane possibly losing its take-off slot, is one example. In order to avoid such costs of dealing with ‘poor’, ‘unreasonable’, ‘difficult’ and ‘demanding’ customers, low-cost airlines maximise aircraft operational time and arduously force passengers to be on time by closing the checking

desk 40 minutes before takeoff time. This benefits the vast majority of passengers and further reduces costs. Not surprisingly, such companies are often assessed positively in terms of their service provision. In 2007, low cost airline Ryanair was voted ‘best airline in Europe’ across indicators such as flight cancellation, punctuality and lost luggage. It appears that full cost airlines cannot compete with such companies either on price or on key aspects of customer service expected from such carriers.

Compliance and exclusion marketing is also marked by a certain transparency and openness of its nature, aims and the ramifications (benefits) both for the companies practicing it and their customers. The principles of such marketing are meant to be highly visible and are deliberately publicised by the companies practicing it. Companies take advantage of every opportunity to teach existing and potential customers as well as the general public about their procedures and processes, their requirements of customers when partnering with the company, in order to achieve high levels of service at minimal cost. The ‘Airport’ series which appeared on UK TV in the mid 2000s is an example of such clear, transparent communication.

Such openness is important. The strategic credibility of such companies is being positively affected by transparency of rules, including the rules of exclusion. This also impacts on the image and brand of the company in the eyes of its various stakeholders and audiences. We suggest that transparency is key to relating to one’s customers. Therefore, customers (and staff who become empowered by the application of rigid procedures) are expected to know and understand the objectives of such practices as well as the strategy and business philosophy informing them. The companies also clearly communicate the ‘outputs’ of compliance and exclusion marketing and how they benefit them as customers. Such company strategies reliant upon transparency have had a profound effect on their corporate image, partly by opening them to criticism that their rules are unfair even though they are probably considerably more transparent than the rules of traditional businesses which are hidden and arbitrary.

4. DRIVERS OF COMPLIANCE AND EXCLUSION MARKETING

Customers are becoming more cynical about the practices of traditional businesses and appear to be less willing to enter into relations with companies based on loyalty to the business. For instance, in a weblog, Gerry McGovern maintains that ‘some customers are not worth caring about’ and goes on to say that success is linked as much to ‘figuring out who is *not* your customer ... as anything else’^[2].

An additional, equally potent driver is the advent, growing sophistication and application of new technologies which have facilitated the development of this new set of companies which have dis-intermediated marketing channel relations and use database technologies to offer a new type of relationship with customers which is advantageous for both and which is based on the clear understanding that the company offer is superior but if ‘things go wrong’ the company is unwilling to engage in costly recovery procedures.

5. THE E-CUSTOMERS OF E-COMPLIANCE BUSINESSES

The practices discussed above are based on ‘minimalism’, but their model of operation has its attractiveness to a wide range of customers. We define the characteristics of such customers and identify the attitudes of these customers. In the process of doing, an answer to a question with important management

implications is sought: What drives these customers to consume the products and services of companies practicing compliance and exclusion marketing?

Customer comments posted on blogs show that the customers' expectations are actually managed very effectively by such companies. Bloggers understand that 'people ... don't expect much and they really shouldn't'. Companies as the ones described by us here have turned previously inaccessible or unavailable products and services widely accessible and available, with 'customers [becoming] more forgiving (or less demanding)'.

Empirical findings from an empirical research carried out by the authors, first, in 2009 by interviewing 235 respondents in the UK, and second, in 2010 by surveying a panel of 1,243 consumers representative of the general UK population confirm such attitudes of customers. For instance, the semi-structured interviews asked respondents to comment on the service provision, satisfaction and dissatisfaction with, complaint behaviour towards, and purchase intentions towards one low-cost airline which is probably the most representative in the UK of the practices outlined by us here. The empirical findings revealed that the customers' perceptions and attitudes towards businesses practicing compliance and exclusion marketing were positive. Far from McGovern and Moon's argument, customers did not hate these providers, irrespective of negative stories of others' poor experiences^[1]. Even those customers who knew someone with a poor experience did not seem discouraged by such stories. The customers also gave mostly positive reviews of the airline in question and seemed to practice positive word-of-mouth and word-of-mouth when communicating with others on the Internet and when relating to colleagues, friends and family members. Respondents either enjoyed using such services or at least 'did not mind', in the words of one interviewee. Media coverage did not have a significant impact on perceptions. Rather, the interviewees focused on their own experiences which were rather unproblematic and largely satisfying. Importantly, these customers also knew the 'compliance' and 'exclusion' rules of such companies.

Similar are the findings from the 2010 survey research. In spite of the high level of compliance, disciplining and exclusion incidents practiced by companies, customers appear these practices, with a high percentage of respondents reporting company constricting and controlling service provision but who have not complained (Table 1) and with a high percentage of respondents reporting company constricting and controlling but positive future behavioural intentions towards the service provider (Table 1).

Table 1. Customers' reaction to compliance and exclusion efforts of businesses

Issue Studied	Customer response measured	Sector				
		Airlines	Comms and telecomms	Finance & banking	Electronic retail	Travel
Extent of customers' acceptance of disciplining (1)	% of reporting respondents who have not complained	63%	35.5%	45%	39%	44.5%
	% of reporting respondents who have not complained but who will (or may) use the company's services in the future	75.5%	75%	79%	79%	66.5%
	% of reporting respondents, who have complained but who will (or may) use the company's services in the future	51.5%	55%	72.5%	77%	43.5%

6. IMPLICATIONS FOR E-MARKETING, E-STRATEGY AND 21ST CENTURY CHINA

In writing this paper, we wanted to draw attention to concepts which we believe have supplemented or even replaced earlier marketing and marketing strategy concepts such as ‘Customer Centricity’. The notion of the customer ‘NOT always being right’, excluding unwanted, unprofitable, unpromising and problematic customers such as complainers and especially serial complainers is backed by empirical examples of the practices of a set of highly successful companies. We also drew attention to the emphasis placed by exclusion-practicing-companies on removing bad payers, small order contracts, customers who cherry pick from the company’s range, customers outside the geographic area serviced or too far from the delivery routes, and even customers who do not use the automated ordering systems properly and as expected by the company. This is a very different approach to that based on traditional exclusion through segmentation of homogenous sets of customer groups. Here compliance and exclusion can be practiced with respect to individual customers and from the very start of communicating with them.

The literature which analyses terminating relations between companies and their customers is not voluminous. Commentators have noted the propensity of marketing scholars to investigate the start, building up and development of relations, but such issues teach us little if anything about the sources, drivers, nature, processes and outcomes of relationship termination. The little that has been written on these issues is largely limited to the analysis of relations terminated by the customer, whereas we are interested in the growing phenomenon of successful companies practicing compliance and exclusion by terminating or not starting at all a relation with individual customers and less so with whole customer segments.

Our research indicates that the picture is much more complex than simply a question of ‘the customer always being right’, or ‘the customer NOT being right’. It does appear that though ‘Customer Compliance’ businesses force customers to comply with company systems, they also enjoy customer satisfaction and a certain level of behavioural loyalty. What marketers originally thought the customer wanted appear to be somewhat misguided. In the airline business, for example, it seems that customers prefer low price, on time flights with minimal lost bags and simple checking procedures to executive lounges, little trays of packaged food, along with eyeshades and footrests, and the special treatment of difficult and complaining customers.

As noted earlier, excluding customers and making them compliant is not a new phenomenon and is part of traditional business operations. However, it has traditionally been achieved through segmentation. For example, customers who did not have the characteristics ‘desired’ by the marketer, such as social status or sufficient income, were excluded from many retail premises, banks, transport systems, hotels and restaurants.. Traditionally companies have also put in place various conditions barring certain customers from interacting with their organisation. However, this model of marketing and management also presupposed the existence of approved agents, dealers and distributors each adding on a margin and preventing customers from buying directly and at a low price. For example, insurance companies, banks and other financial institutions would have traditionally excluded customers by using very complex and detailed application forms and, on the basis of these, rejected unwanted customers. Another example is that of education establishments which would have excluded pupils and students of certain backgrounds, thus practicing demographic and psychographic segmentation. The type of compliance and exclusion currently practiced that we analysed here has come about

following the breakdown of the traditional supply chain distribution systems through use of the Internet which allows organisations to make sales directly to the end customer, to cut out the middlemen and gatekeepers, and to remove existing and potentially troublesome customers.

REFERENCES

- [1] McGovern, G. & Moon, Y. (2007). Companies and the customers who hate them. *Harvard Business Review*, June, 78-84.
- [2] McGovern, G. (2008). Some customers are not worth caring about. <http://www.gerrymcgovern.com/nt/2008/nt-2008-02-18-focus.htm>.

A Research about Consumer's Usage Intention to Green Finance Products

---- Taking the Alipay's Ant forest as the Example

Hong Peng¹, Jun Mi^{2}*

^{1,2}Institute of business administration, Zhongnan University of Economics and Law, China

¹Abstract: This paper taking the Alipay's ant forest as an example builds the consumer's usage intention model of green financial products from the perspective of perceived value, collecting data by the questionnaire. The model is verified by SPSS19.0. The research shows that function value, social value and emotional value have significant influence on the consumer's usage intention to green financial products. Green value and perceived cost have no significant influence on the consumer's usage intention to green financial products.

Keywords: green financial products; perceived value; usage intention; ant forest.

1.INTRODUCTION

With the increasingly serious destruction of ecological environment and the increasing consumption of resources, how to achieve sustainable development of economy and society has become an important issue faced by human society. The financial industry can effectively transform the mode of economic growth and the consumption habits of residents through the allocation of financial resources, the guidance of market investment and the consumption of residents in the process of economic and social development. Therefore, taking the financial way to improve the predicament has become a universal choice for all countries in the world. The public have put forward an important idea of the green finance in the research and practice.

In recent years, with the rapid development of Internet and the increasing awareness of citizens' environmental protection, the participants of green finance have expanded from traditional finance industries such as the banks, insurance companies to enterprises and citizens. All kinds of green financial products are also becoming richer with the expansion of the participants. In January 2017, the Green Digital Finance digital Alliance was launched by the "ant gold service" and the United Nations Environment Programme in Davos. The alliance would build a world's leading financial technology companies, searching a new way to promote global sustainable development. This has also opened the prelude to the participation of enterprises and consumers in the development of green economy. In early August 2016, the ant payment service under the Alipay platform build personal carbon account named as "ant forest" to consumers. At the end of April 2017, the ant forest had over 220 million users, contributing carbon emission reduction of 5,000 tons per day. It has planted a total of 8.45 million trees, reducing more than 2,500 tons of carbon emissions every day ^[1]. The green financial product of ant forest can attract such a huge number of users to use the product in more than half a year. The reasons behind it are worth discussing. The current researches on green finance are focused on green finance theory, the classification of the green financial products, national policy. There are little researches on green financial products from the consumer's perspective. Taking the ant forest as an example, this paper researches the consumer's usage intention based on the consumer perceived value theory to provide a useful reference for the marketing of green financial products.

2. REVIEW

2.1 Green Finance

The green finance is also known as the environmental finance and the sustainable finance, which has not yet reached a consensus on the definition. The American English Dictionary defines the green finance as dealing with the environmental

* Hong Peng, Email:hpeng520@126.com Jun Mi,Email:hub20160907@163.com

crisis, and studying how to use diverse financial tools to protect the environment. The official's definition of green finance at home is an economic activity to support environmental improvement, coping with climate changing, resource conservation and efficient utilization. In academia, Salazar believes that green finance is a great innovation of financial industry, which can achieve a balance between the economic development and environmental protection ^[2]. Labatt and White believe that green finance is a financial tool for improving the environmental quality and transferring environmental risks on the basis of market research ^[3]. Li points out that green finance is a financial activity aimed at promoting the coordinated development of economy, resources and environment for credit, insurance, securities, industrial funds and other financial activities ^[4]. Although different views have been proposed on the concept of green finance, its core concept does not deviate from the concept of environmental protection and sustainable development. At present, the main body of green finance has expanded from traditional finance industries such as the banks, insurance companies to enterprises and citizens with the development of practice. Referring to the research results and the actual development situation of green finance both at home and abroad, this paper thinks that the late definition of green finance in China's official department is more realistic.

The scholars' mainly researches about the green finance focus on the structure system and the impact on the sustainable development so far. Weng and Ge sum up China's green financial products based on a large number of cases and data research, including environmental protection industry index products, environmental protection and energy saving financing products, and carbon financial products. The foreign green financial products are divided into four categories: retail banks, enterprises and investment banks, asset management and insurance ^[5]. Jiang sums up the green financial system, including green credit, green securities, government funds and so on, through a lot of research on the development of green finance in developed countries ^[6]. Marcel has discussed the relationship between the banking industry and the sustainable development of the economy ^[7]. Panayotou thinks that carbon emissions and economic growth are in a "U" shape in the long run ^[8]. Li systematically studies the existing green financial activities in China, and believes that green finance is an important way to raise the sustainable development of economy. China's financial industry should establish the concept of developing green economy and support the development of green industry. The Southwestern University of Finance & Economics and the Ministry of environmental protection analyzes the relationship between green finance and economic sustainable development from the level of policy and practice, and put forward some policy recommendations to improve China's green financial system. Wang analyzes the contribution of green finance to the economic development from the aspects of green finance optimizing the macroscopic and microcosmic economy ^[9].

To sum up, the scholars at home and abroad have more researches on green finance from its concept, structure and the impact on sustainable development. There are few researches on green finance and products from the perspective of consumers.

2.2 The theory of perceived value

The perceived value is considered to be the ultimate motivation to the user's behavior. It plays a decisive role in consumer's adoption, acceptance or use and purchase. The scholars don't define the perceived value in the early. Zenithal (1988) is the first scholar who proposes the customer perceived value theory from the customer perspective. She defines the customer perceived value as the customers' overall evaluation to the products or service after they weigh the benefits and payments. From the view of relations, Gronroos stresses on the effect of relationship on customer value. He argues that the trade-off between gains and losses cannot be confined to single scenario but should be extended to the value of the process of the entire relationship episode. Woodruff puts forward that customer value is the consumer's preference or evaluation when the products or service helps (impedes) them to achieve their goals under specific usage scenarios by the empirical research ^[10].

The scholars have studied the composition of perceived value from different aspects. Zeithaml systematically analyzes the composition of perceived value from two aspects of gains and losses. She divided the gains into product attributes, perceived quality and other related high-level attributes, and the profits and losses included two levels which are monetary and non-monetary ^[11]. Sheth divides the customer value into five elements including functional value, social value,

emotional value, cognitive value, situational value. Wang holds that the price reflects the value function of consumers to purchase goods or services but not fully reflect the cost to consumers when he studies the Sweeney's perceived value model of construction. He takes the financial industry as the research background and establishes a consumer perceived value model which is composed of functional value, perceived loss, social value and emotional value^[12]. Zhou is the first scholar who finds green value is one part of the customer perceived value when they buy green cosmetics and divides the customer perceived value into five factors which are functional value, emotional value, social value and green value and perceived sacrifice^[13]. Her finding enriches the composition of the customer perceived value.

2.3 Usage intention

The usage intention refers to the relatively permanent cognitive and emotional orientation of consumers for a product or service, that is to say the consumers' psychological activities whether they will use or purchase the products. Jillian & Geoffrey make a conclusion that emotional value, social value, functional value and cognitive value from the perspective of perceived value theory have different effects on consumption intention under different consumer behavior situations^[14]. Kim & Chan establish the consumer perceived value acceptance model (VAN) based on the technology acceptance model (TAM) and the perceived value theory proposed by predecessors^[15]. The model divides consumer perceived value into two parts: perceived benefit and perceived cost. Perceived benefit includes perceived usefulness and perceived entertainment. Perceived cost includes perceived risk and specificity. Perceived risk refers to the risk of the cost of money and the specificity refers to the non-monetary costs including the time, effort, psychology which can't be evaluated by money. Zeng finds that perceived risk has a significant negative impact on the user's willingness to use the online bank in the study of the perceived risk of users' online bank^[16]. Yu researches the consumer's willingness to buy green food on the basis of the theory of customer perceived value. His research concludes that functional value, social value and green value had a significant positive impact on consumers' purchase of green food, and emotional value and perceived cost had little effect on consumers' purchase of green food^[17].

3. HYPOTHESES AND MODEL

The independent variable of this study is the perceived value of the consumer to the green financial products, and the dependent variable is the consumer's usage intention. This paper quotes the research results from the domestic and foreign scholars and determines functional value, social value, emotional value, green value and perceived cost as the five dimensions of the customer perceived value.

3.1 Functional value

Functional value is the value to meet the needs of the consumer's functional properties. It is the key factor for users to choose and purchase products. Consumers can experience functional benefits, social benefits, personal benefits and experience gains when they buy or use some kind of products or service. Wang points out that the functional value of financial products has a positive impact on the purchase of financial products. Therefore, we believe that the functional value of ant forest has a positive impact on the consumer's usage intention.

H1: the functional value of ant forest affects the consumer's usage intention positively.

3.2 Social value

Social value refers to the social identity and social self-concept produced by consumers in the use of the product. The users will be affected by friends around their Alipaies or in real life when they use the ant forest. The ant forest has the function of collecting friend's energy and sending energy to friends which integrates the interaction way of social network and increases the connection between friends virtually. Owing to being benefit to break the barriers that users encounter in traditional interpersonal communication, this function has an important impact on attracting consumers to use ant forest. In addition, the consumers can also improve their social images by using ant forest. Therefore, this paper believes that the social value of ant forest will affect the consumer's usage intention.

H2: the social value of the ant forest has a positive effect on the consumer's usage intention.

3.3 Emotional value

Emotional value is the level of emotional identity that consumers can obtain after using (buying) products or services. They will get great psychological satisfaction and joy through their efforts to transform virtual trees into a real tree after they use ants' forests. Therefore, this paper puts forward the hypothesis.

H3: the emotional value of the ant forest positively affects the consumer's usage intention.

3.4 Green value

Green value is the ecological utility of the consumer from the product or service. Consumers collect the energy by online payments, green trips, green parcels and so on. On the one hand, it can reduce environmental pollution and carbon emissions. On the other hand, these ways can collect energy to turn virtual trees into reality trees, so as to achieve the purpose of environmental protection. Xie finds that the green value has a positive impact on the consumer premium purchase of agricultural products^[18]. Therefore, this paper believes that the green value of ant forest will affect the consumer's usage intention.

H4: the green value of ant forest has a positive impact on the consumer's usage intention.

3.5 Perceived cost

Perceived cost refers to the monetary or non-monetary costs of a consumer when he wants to obtain a product or service, such as money, time etc. Consumers need to pay money online or offline by Alipay in the process of using ant forest trees need to spend time collecting their own or their friends' energy at a particular time every day. Will these help consumers to increase their consumption and time? Yu (2012) makes a conclusion that perceived cost l has a negative impact on the consumer's purchase of green products. Therefore, this paper puts forward the hypothesis.

H5: the perceived cost of the ant forest has a negative impact on the consumer's usage intention.

Based on the above hypothesis, we obtain a hypothesis model for the consumer's usage intention on the ant forest. As shown in Figure 1.

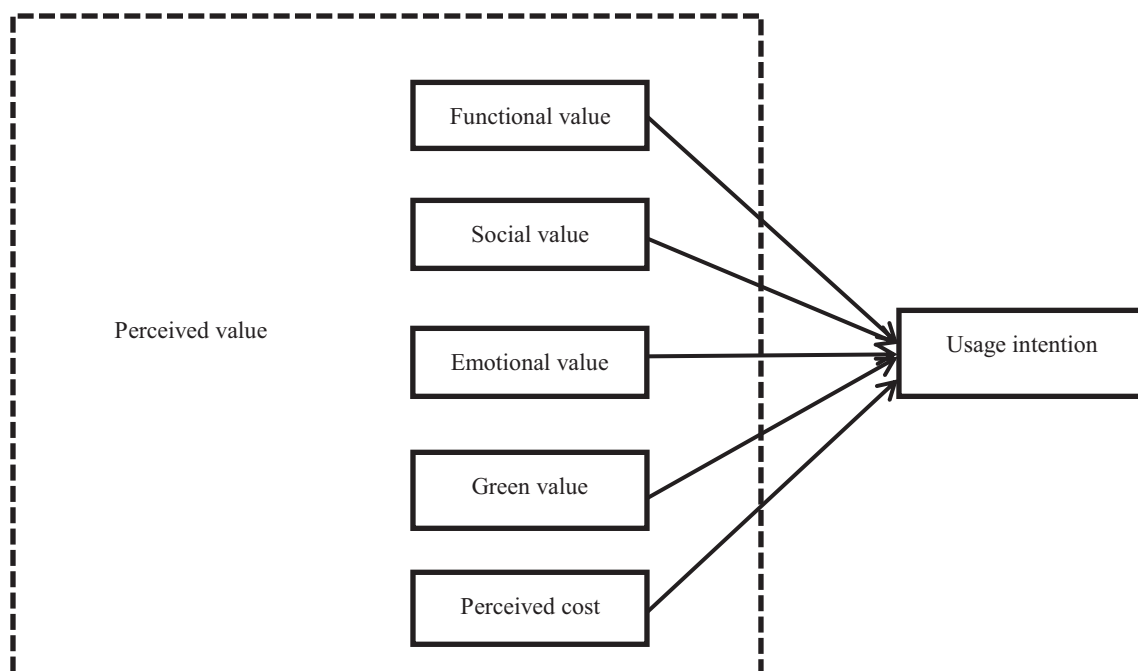


Figure 1. the research model

4. RESEARCH DESIGN

4.1 The questionnaire design and variable measurement

The measurement items of the independent variables and the dependent variables in the questionnaire are directly from the existing literature. In order to meet the present situation, the measurement items of each variable have been modified on the basis of the existing literature. The questionnaire is designed with the Likert 5 component scale. The reliability and validity of the questionnaire are tested by pre-investigation before the formal investigation. According to the results of the test, some items of the questionnaire are revised, and the final formal questionnaire is formed.

4.2 Data collection

There are 328 valid questionnaires are collected by the network. The proportion of males and females are 32.32% and 67.68% respectively in the survey and the females have a relatively high proportion. The responds' age from 18 to 35 years old has a proportion of 98.47% and their degree is mainly undergraduate and master which occupies the proportion of 87.42%. Most of them are students(66.47%) and 20.12% of the total number of samples are employees. monthly disposable income in 1000 yuan and 1000-3000 yuan, which accounted for the total sample size of 71.65%, the main reason for this result is the sample proportion of college students is very high, and most of the students have no direct economic source, the main source of income in the family, so monthly disposable income is low. To sum up, the main features of the sample can be summarized as: the majority of women, the young, the higher education, and the less disposable income. This shows that the use of Alipay's ant forest users are mainly concentrated in the students and young enterprise staff.

According to the survey , 250 responders open the ant forest, accounting for 76.22% of the total numbers. It shows that the ant forest has a high degree of audience. Based on whether the opening of ant forest, investigators are divided into two categories, for the opening of ant in forest, the monthly electricity consumption amount, frequency of use of Alipay and ant forest energy source survey. In the sample, by the payment of Po consumption accounted for the largest amount, up to 72% yuan in 500-2000, and the main research object is students, the income level is not high correlation; weekly use of more than 1 forest ant consumers accounted for 77.6%, that consumers use ant forest frequency is relatively high; most of the three ways the forest energy comes from the payment of the next line (87.6%), green travel (72%), (67.6%) energy collection of friends. This reflects the use of ant forest to a certain extent promoted the online interaction between the development of Alipay and the consumer.

In this survey, there are 78 responders who have the Alipay but do not open the ant forest. The main reason for the users do not open the ant forest (62.82%) is that they do not comprehend it. It can be seen that the awareness of the users need to be improved although the introduction of ant forest has been recognized by most users and we need to strengthen marketing to improve the audience of the ant forest.

5. DATA ANALYSIS AND RESULTS

5.1 The reliability and validity analysis

The reliability and validity of the data are analyzed by SPSS19.0. The results show that the alpha coefficients of all variables are above 0.8, and the overall reliability is 0.902, indicating that the reliability of the research scale is perfect, and the reliability and stability of the questionnaire are reliable. This paper adopts factor analysis in KMO and Bartlett sphericity test for statistical analysis, the results show that the KMO value is 0.948. The probability of significant chi square Bartlett sphericity test statistic value of 0, less than 0.001, indicating the validity of the inventory can be further analyzed

5.2 Regression analysis

In order to verify the causal relationship between the independent variables and the dependent variable, we make a

regression analysis of the data and obtain the results as follows.

model	R	R ²	The adjustment of R ²	Standard estimation error
1	.792 ^a	.627	.625	.5421910
2	.826 ^b	.682	.679	.5018525
3	.831 ^c	.691	.687	.4956774

1. Predictive variable: Emotional value. 2. Predictive variables: Emotional value, functional value. 3. Predictive variables: Emotional value, functional value, social value.

Fig 2. The summary of the model

model	Non standardized coefficient		Standard coefficient	t	Sig.
	B	Standard error	Trial Edition		
1	.865	.146		5.938	.000
EV	.771	.039	.792	19.533	.000
2	.280	.164		1.702	.090
EV	.561	.050	.577	11.317	.000
FV	.358	.057	.318	6.242	.000
3	.244	.163		1.498	.135
EV	.500	.054	.514	9.201	.000
FV	.321	.058	.285	5.486	.000
SV	.122	.047	.130	2.582	.010

a. the dependent variable: usage intention

Fig 3. The coefficient

We can find that the third regression model is better than the first and second regression model from the results. From the third models, the R is 0.831 and R² is 0.691 in the third model and the third model is obviously better than the two models before from the fitting degree. Therefore, emotional value, functional value and social value have a significant positive impact on consumer's usage intention. The degree of the influence is emotional value, functional value and social value and we can verify the hypothesis 1, 2, and 3 are supported.

The T value of green value is 1.439, and the F value of the model is 0.152 which is greater than 0.05. Therefore, green value has no significant impact on the consumer's usage intention and we can draw that the hypothesis 4 is not supported. The T value of perceived cost is -1.673, and the F value of the model is 0.096 which is greater than 0.05. Therefore, perceived cost has no significant negative impact on the consumer's usage intention and we can assume that the hypothesis 5 is wrong.

6. DISCUSSION

6.1 The conclusions

This paper puts forward the model of consumer's usage intention to ant forest, and verifies the validity of the model from the view of empirical. We draw some conclusions as follows. Firstly, emotional value, functional value and social value have a significant positive impact on consumer's usage intention to ant forest. The degree of the influence is emotional value, functional value and social value. This is consistent with the influence effect of previous scholars on the emotional value,

functional value and social value of other products. Secondly, green value and perceived cost have no significant influence on consumer's usage intention to the ant forest. Green value has no significant influence on consumers' willingness to use, which is inconsistent with previous research hypotheses. The possible reason is that the lack of authenticity causes users to not fully feel the actual effect of their actions on environmental protection. The ant forest is a virtual product so that the forest users only see a virtual tree after collecting energy. In addition, owing to lack of timely and effective information feedback makes the users unable to know their actions whether making practical improvements to the environment. Another important reason is that the green value of products cannot significantly affect consumers. As a large domestic mobile payment platform, the Alipay can provide a guarantee for the safety of consumer accounts and consumers need not have too much to consider the safety cost in the use of ant forest. The energy acquisition way of ant forest is quite diverse. Consumers can not only get energy through walking, collecting friends' energy, but also getting energy through online payment. They do not need to pay more time cost and capital cost in this product, so the perceived cost has no significant impact on the consumer's willingness to use.

6.2 The significance of the research

From the level of theoretical, the articles on the study of green financial products from the consumer perspective are relatively absence. This paper studies the consumer's usage intention to green financial products from the perspective of customer perceived value theory, and expands the scope of the research on green financial products. At the same time, this paper adds green value to customer perceived value theory based on the specific research subjects and previous studies, enriching the content of customer perceived value theory.

From the level of practical, this paper studies the factors of consumers' usage intention to green financial products, which has a certain guiding significance for the financial industry to develop and spread green financial products. At present, more and more financial companies begin to develop and spread green financial products, but they don't know what factors they should consider when they develop green financial products. This paper based on the theory of customer perceived value provides directions for enterprises to focus on product's functional value, social value and emotional value when develop green financial products.

6.3 limitations and Prospects

There are some limitations although the author has made a lot of efforts. Firstly, the object of the responders in this paper is mainly about the university student so that the conclusions should be tested for other occupation groups. Therefore, the future researches can expand the range of sample selection to improve the credibility of the study conclusion. Secondly, this paper does not test the individual characteristics of the consumer, such as sex, age, degree of education and income whether have an impact on the usage intention. In the future, we can introduce control variables such as age, sex and occupation to test whether they have an impact on the usage intention to the ant forest. Thirdly, this paper limited to the scope of the study does not consider the others factors whether affect the consumer's usage intention to green financial products. Some responders indicate that the ant forest can attract them in the early stage, but they cannot produce sustained use effect during the process of the investigation. It is worth exploring in future research.

REFERENCES

- [1] Ant Financial (2017). Ant Financial 2016 Sustainability Report, 51-55.
- [2] Salazar, J. (1998). Environmental Finance: Linking Two World. Presented at a Workshop on Financial Innovations for Biodiversity Bratislava,(1)2-18
- [3] Labatt, S., White, R.(2002). Environmental Finance: A Guide to Environmental Risk Assessment and Financial Products. Canada : John Wiley & Sons.
- [4]Li Xiaoxi, Xia Guang, Cai Ning.(2015). Green finance and sustainable development. The Financial

- Forum,238(10):30-40.(in Chinese)
- [5]Ge Chazhong, Weng Zhixiong, Duan Xianming. (2015). Green Finance Policies and Products: Status and Recommendation. *Environmental Protection*,43 (2):32 —37.(in Chinese)
- [6]Jiang Xianling, Zhang Qingbo.(2017). A Review of The Theory and Practice of Green Finance in Developed Countries. *China's population, resources and environment*, 27(5): 323-326.(in Chinese)
- [7] Marcel Jeucken.(2001). *Sustainable Finance and Banking: The Financial Sector and the Future of the Planet*. London: Earths Can Publications Ltd.
- [8] Panayotou T D.(2003). *Economic Growth and the Environment*. Boston: Harvard University and Cyprus International Institute of Management, 30-34.
- [9] Wang Yao, Pan Dongyang, Zhang Xiao.(2016). A Research on the contribution of green finance to China's economic development. *Comparison of Economic and Social System*, (11): 33-42. (in Chinese)
- [10] Woodruff.(1997). Customer Value: The Next Source for Competitive Advantage. *Journal for Academy of Marketing Science*, 25(2): 139-153
- [11] Zeithaml, V.A.(1988). Consumer Perceptions of Price, Quality and Value: A Means-end Model and Synthesis of Evidence. *Journal of Marketing*, 52(3): 2-22.
- [12] Wang Yonggui.(2003). Customer Value and Customer Relationship Management: Theoretical Framework and Empirical Analysis. *The Seventh National Symposium and System Science for Youth Management*. (in Chinese)
- [13] Zhou yijin.(2006). The Research on Dimensions of Customer Perceived Value of Green Products—Taking Green Cosmetic for Example. *Guangdong University of Foreign Studies*. (in Chinese)
- [14] Jillian, C.S., Geoffery, N.S.(2001) Consumer Perceived Value: The Development of a Multiple Item Scale. *Journal of retailing*,77(20:203-220.
- [15] Kim, Hee—Woong, Chan,Hock Chuan.(2007). Value — Based on The Adoption of Mobile Internet: An empirical investigation.*Decision Support Systems*, 43(1):43-52.
- [16] Zeng Yurong, Ma Yan.(2011). Perceived Risk and Perceived Benefit and Consumer's Use of the Internet banking business. *The Financial Forum*,1 (181) :70-75. (in Chinese)
- [17] Yu jia. (2012). The Study on the Impact of Customer Perceived Value on Purchasing Behavior of Green Food. *South China University of Technology*. (in Chinese)
- [18] Xie Qiang.(2016). A Study on the Premium Payment of Green Agricultural Products from the Perspective of Perceived Value. *Northwest A&F University*. (in Chinese)

Can the Adjustment of Consumer Finance Credit Line Cause Impulse Purchase?---An Analysis from the Perspective of E-commerce Promotion

Hong Peng¹, Yushan Yusufujiang^{2*}

^{1,2}Institute of business administration, Zhongnan University of Economics and Law, China

Abstract: Consumer finance has penetrated into people's daily life, especially in the field of e-commerce online shopping. In order to stimulate consumers' desire to buy, adjusting credit line has become the marketing strategy of many merchants. However, little research has been done on whether this method can really cause consumers to buy impulsively. This paper discusses the relationship between the credit line adjustment of e-commerce consumer finance and the impulse purchasing desire of consumers in the context of online sales promotion in the "double eleven shopping festival". In this research, data is collected by adopting simulated methods, and the data is analyzed through the SPSS. The research result shows that the increase of the consumption finance credit line positively affects the consumer impulse purchasing desire, and positive emotion has mediating effect; the results of this research can help e-commerce platforms to manage the credit line reasonably and formulate effective marketing strategies to enhance its competitiveness in the online market.

Keywords: Consumer Finance; Credit Line; Impulse Purchase; Positive Emotion

1.INTRODUCTION

In recent years, internet consumer finance in China has been developing rapidly. For instance, JDcom, Ali, Tencent and so on have successively launched their consumer financial products. In order to stimulate consumption, e-commerce platforms have launched a variety of marketing strategies on consumer finance, among which credit line adjustment has been widely used in all kinds of e-commerce shopping festivals. For example, prior to the 2016 "double eleven shopping festival", Ant Check Later announced that it could offer a temporary credit line increase for its users with a total amount of more than 10 billion yuan, with about 60% of its users eligible for receiving a temporary credit line increase, up to 55,000 yuan. This undoubtedly greatly stimulates the consumer's purchasing desire, especially for "shopaholics". When consumers use consumer finance, credit line and customer consumption are interdependent and influence each other. Therefore, this paper discusses how the credit line adjustment of e-commerce consumer finance will affect the impulse purchase of consumers and how the e-commerce platform and merchants should deal with this change and so on under the promotion environment of online shopping. It is of great significance for the platforms to effectively manage consumer financial credit lines so as to enhance customer satisfaction and increase sales, as well as promote the long-term development of e-commerce enterprises in the fierce market competition.

Impulse purchase has attracted the attention of scholars in the traditional retail market. Some investigations have found that the category of products, the display mode of store shelves and the arrangement of goods will affect the frequency of impulse purchase of consumers. Luoji concluded that the behavior of other people on the purchase site will affect the impulse purchase of consumers^[1]; some scholars also think that the impulse purchase of consumers is closely related to their psychological mood, especially when making shopping decisions, emotion is a great reference factor^[2]. Davis and Li argue that in the internet age, impulse purchase is more likely to occur because of the convenience and variety of products available online^[3]. In addition, many of the latest research findings of psychology are increasingly applied to consumer behavior. For example, some scholars have proved that the self-construction concept in the field of psychology can significantly affect

* Hong Peng, Email:hpeng520@126.com YushanYusufujiang,Email:391823426@qq.com

consumer's impulse purchasing behavior. However, in the process of using e-commerce consumer finance, less attention has been paid on how the credit line adjustment causes consumers to buy impulsively. Therefore, based on the previous researches, this paper explores the influence of the credit line adjustment of e-commerce consumer finance on impulsive purchase, and whether the influence is regulated by product category and self-construction under the online promotion environment.

2. REVIEW

2.1 Consumer Finance

Merton defines consumer finance as the finance of consumers. More specifically, it means that citizens can satisfy their own consumer demand by using their existing assets and their own accumulated credit under the existing financial environment. Domestic scholars Liao Lian define consumer finance as the financial products and services provided by financial institutions to consumers^[4].

The credit line of consumer finance generally refers to the maximum amount available to consumers offered by financial institutions, i.e., the maximum amount that can be overdrawn by consumer finance; temporary adjustment means that financial institutions temporarily adjust the credit line of consumers under certain circumstances. In most cases, the credit line is increased on the basis of the original credit line, and there will be a certain time limit for the increased credit line (it will return to the original credit line of the consumer when it expires).

2.2 Impulse Purchase

Song Tao holds that impulse purchase is a purchase decision made by consumers in an instant^[5]. The decision is completely determined by the aesthetic unconscious behavior, and it is a perceptual behavior. Verplanken & Herabadi define impulse purchase as the behavior of unplanned purchase caused by a series of uncontrollable positive emotional reactions that consumers do not plan and think over^[6].

In this paper, impulse purchasing desire is defined as consumer's desire to buy a certain unplanned product immediately due to positive emotional reactions stimulated by the external environment. Such a desire is what this paper calls impulsive desire to buy.

Generally, there are two factors leading to impulse purchase: one is the external environment, such as store design style, promotional or marketing stimulation and so on. Its influence on impulse purchase is mainly reflected in triggering psychological or emotional fluctuations of consumers through environment to cause impulse purchase^[7]; The other factor is the individual characteristics of consumers. For instance, impulsive traits will have a significant impact on impulse purchase^[8].

2.3 Credit Line Adjustment and Impulse Purchase

The researches on impulsive purchase in the online environment started relatively late, most of the researches focus on applying the results of the traditional offline environment to the online environment, and at the same time, constantly enriching and adding new factors such as web design, online reviews, transaction records and so on according to the characteristics of the online shopping environment. As an e-commerce platform, e-commerce consumer finance provides a new way of payment for consumers. So far, no scholars have discussed and verified its impact on consumer buying behavior. Based on the actual shopping scenarios, this paper takes the adjustment of consumer finance credit line as the stimulus variable of environmental psychological model under the background of promotion during the "double eleven shopping festival" to study its influence on consumers' impulse purchase intention.

2.4 Emotional and Impulse Purchase

Scholars' researches can be divided into three categories based on the relationship between emotional and impulse purchase: First, impulse purchase is accompanied by emotional experience. Gardern & Rook's study in 1998 proved that impulse purchase can counteract consumers, bring a sense of joy to them, and often maintain and prolong that joy^[9]. Second, impulse purchase is distinguished from planned purchase based on emotional characteristics. Impulsive consumptions took place when consumers purchase impulsively show excitement and happiness after being stimulated by external environment (especially marketing activities) are mostly recreational consumptions. Third, studies confirmed that emotion can directly affect impulse purchase, such as impulse purchase behavior induced by intense emotional fluctuations caused by external stimuli^[10].

3. RESEARCH ASSUMPTIONS

3.1 The Effect of Credit Line Adjustment on Impulse Purchasing Intention of E-commerce Consumer Finance

Marketing stimuli and situational factors are important influencing factors of consumers' impulse purchase. Consumers often experience strong emotional activities when they make impulsive consumption. The credit line adjustment of e-commerce consumer finance is a change in payment mode for consumers, and payment mode belongs to situational factors. Therefore, it can be envisaged that consumers' impulse purchase intention will be affected when using financial products to make consumption decisions. Huizinghi's study also confirms that the use of consumer finance can affect consumers' impulse purchase^[11]. For example, the increase of credit card line will result in more impulse purchases. This paper presents the following assumptions:

H1: Compared with the credit line maintenance, the increase of the credit line of e-commerce consumer finance can positively affect consumers' impulse purchase intention.

3.2 Mediating Effect of Positive Emotion

Based on empirical research, Zhang Di concluded that the stimulation of external environment has a significant impact on consumer online shopping emotional response^[12]. By reviewing the research results of domestic and foreign scholars, she found that consumers' emotion plays a very important role in the decision-making process of impulse purchase. The increase of e-commerce consumer financial credit line is beneficial for consumers in the promotion environment such as during the "double eleven shopping festival". Moreover, compared with the negative emotional response, the positive emotional response has a greater impact on the consumer's impulse purchase intention, so the positive emotion will be the intermediate variable in this paper. Based on this, the following assumptions are presented:

H2: Positive emotion has significant intermediary effect in the influence of the credit line adjustment on impulsive purchase.

3.3 The Regulatory Role of Product Category and Self-construction

Many scholars in marketing field have defined and categorized hedonic and utilitarian goods based on psychology. Hoch & Loewenstein and Keinan & Kivetz measure the level of impulsive buying in terms of consumer choices between hedonic and utilitarian goods^[13]. Compared with will utilitarian products, hedonic goods can better stimulate consumers' feelings of pursuing happiness and in turn arouse their impulse to buy; Lee Shin & Kim think online shopping with hedonic browsing increases impulse purchase^[14], but did not verify the difference between hedonic and utilitarian products. Based on the differences in the performance of impulse purchase between the different product categories in the above research results, this paper presents the following

assumptions:

H3: Product category significantly regulates the impact of e-commerce consumer financial credit lines on positive emotions.

H3a: For utilitarian goods, increasing e-commerce consumer financial credit line can have a more positive impact on consumer emotion than maintaining a credit line.

H3b: For hedonic goods, increasing e-commerce consumer financial credit line can have a more positive impact on consumer emotion than maintaining a credit line.

Lee Angela & Aaker believes that self-construction plays a significant role in consumer purchasing decisions^[15]: consumers with independent self-construction are less likely to take risks and buy impulsively, while the dependent self-construction consumers consider more about the group, especially the family. They usually take the family as a whole to bear the consequences of impulse purchase behavior. As a result, consumers with self-constructed characteristics of dependence are more likely to buy impulsively. Independent self-constructed consumers pay more attention to the characteristics of the product itself and the marketing environment at that time, such as the quality of products, preferential sales discounts, and convenient payment methods and so on. When stimulated, consumers usually focus on their own feelings, and will experience emotional changes when they think there are good prospects of gain, which in turn induce impulse purchase; dependent self-construction consumers pay more attention to the added value and symbolic significance of the product. When facing marketing incentives such as sales promotion, they will consider more about how people around feel about themselves and benefits to the group. They tend to give up buying on second thoughts. Based on the researches of Hamilton and Biehal, this paper presents the following assumptions:

H4: Self-construction significantly regulates the impact of e-commerce consumer financial credit line on consumer positive emotions.

H4a: For independent self-constructed consumers, increasing e-commerce consumer financial credit line can have a more positive impact on consumer emotion than maintaining a credit line.

H4b: For dependent self-constructed consumers, increasing e-commerce consumer financial credit line can have a more positive impact on consumer emotion than maintaining a credit line.

4. RESEARCH DESIGN

Based on the literature on impulsive purchase and consumer finance, and in combination with the environmental characteristics of real online shopping, this paper adopts the scenario experimental method to simulate the situation of e-commerce consumer finance shopping under the promotion environment during “double eleven shopping festival”. There are four kinds of scenario: the maintenance of the credit line of hedonic goods, the increase of the credit line of hedonic goods, the maintenance of the credit line of utilitarian goods and the increase of the credit line of utilitarian products. In order to ensure the authenticity of the experimental situation, a number of pre-tests were carried out in advance, and different subjects were selected to collect opinions in order to avoid errors in the experiment.

4.1 Variable Measurement and Questionnaire Design

4.1.1 Variable Measurement

The variables in this study include the adjustment of consumer financial credit line, positive emotion, impulse purchase intention, product category and self-construction. Among them, the adjustment of consumer financial credit line and product category are set by the situation, so the variables that need to be measured are positive emotion, impulse purchase intention and self-construction. On the basis of the relevant literature, in combination with the experimental situation characteristics of this study, this paper defines the operation of each

variable and forms the corresponding measurement scale.

4.1.2 Questionnaire Design

The questionnaire in this study consists of four parts: the first part is an experimental scenario in which consumers shop online using Ant Check Later during promotions in the “double eleven shopping festival”; the second part consists of eight questions. Subjects are asked to evaluate the authenticity of the shopping situation during the “double eleven shopping festival”, and then the positive emotions in the shopping decision-making process are evaluated, and finally the impulse purchase intention is scored; the third part consists of 11 questions on the measurement and test of self-construction; the fourth part is to investigate the basic information of the subjects. All the scales were measured with the Likert7 scale to measure all the questions in the questionnaire. “1” represents for “strongly disagree” and “7” represents “strongly agree”.

4.2 Experimental Design

In this study, 2 (credit line: maintenance vs increase) *2 (product category: hedonic goods vs utilitarian goods) intergroup factor design method based on scenario simulation is used to verify the hypothesis of the study. There are four different experimental scenarios: the maintenance of hedonic credit line and the increase of hedonic credit line, and the maintenance of utilitarian goods credit line and the increase of utilitarian goods credit line. Before starting the scenario design of using Ant Check Later to shop online during the “double eleven shopping festival”, we learned that the average credit line of Ant Check Later for students around ranged from 1,000 to 1,500 Yuan. On the basis of this, the available credit line of Ant Check Later was set as 1500 Yuan, and the temporary credit line increase during the “double eleven shopping festival” was set as 1,000 Yuan, which can roughly reflect the situation in the promotion environment during the “double eleven shopping festival”. After that, several students of different grades were invited to evaluate the experimental situation anonymously, which included authenticity, objectivity, readability and feedback. The content of the shopping scenario was revised before it was used as an experimental material. The details are as follows:

“Suppose you have a credit line of 1,500 Yuan on Ant Check Later. The “double eleven shopping festival” on Taobao is approaching, and you have recently added goods with a total price of nearly 1,500 Yuan to your shopping cart, ready to pay using Ant Check Later on the “double eleven shopping festival”.

The second part describes 2 (e-commerce consumer finance credit line adjustment: maintenance vs increase) *2 (product category: hedonic goods vs utilitarian goods). There is a total of four different shopping scenarios. The details are as follows:

(1) Credit line maintenance of utilitarian goods: The day before the shopping festival, you see a very beautiful and thick sweater that cost 150 yuan, but the credit line of Ant Check Later is less than 100 yuan. You can't use Ant Check Later to buy this sweater.

(2) Credit line increase of utilitarian goods: The day before shopping festival, you see a very beautiful thick sweater that cost 150 yuan while browsing Taobao. At this time, you have a temporary credit line of 1,000 Yuan, and the total credit line available rises to 2,500 Yuan. You can use the Ant Check Later to buy this sweater.

(3) Credit line maintenance of hedonic goods: The day before the shopping festival, you see a box of beautifully packaged, long-cherished imported chocolate that cost 150 yuan, but the credit line on the Ant Check Later is less than 100 yuan. You can't use the Ant Check Later to buy the chocolate.

(4) Credit line increase of hedonic goods: The day before the shopping festival, you see a box of beautifully packed, long-cherished imported chocolate that cost 150 yuan. At this time, you have a temporary credit line of 1,000 Yuan, and the total credit line available rises to 2,500 Yuan. You can use the Ant Check Later to buy the chocolate.

5. DATA ANALYSIS

The main experimental objects of this paper are college students. Data is collected from both online and offline sources. Questionnaires began to be distributed on December 5th 2016 and were collected by December 10th 2016. There were 94.21% valid questionnaires and 94.21% valid questionnaires after excluding the non-qualified questionnaires.

5.1 Descriptive Statistical Analysis

A descriptive statistical analysis of the samples was carried out using SPSS19.0 software. The results showed that the number of men and women were 115 and 211 respectively in the 326 valid subjects who participated in the experiment, accounting for 35.3 and 64.7 percent of the total respectively, which is in line with the overall trend that women hold a majority in online shopping. In terms of age, people in the age range of 19-22 splurge the most on online shopping platforms, accounting for 72.39 percent. This age group mainly consists of undergraduate students. In general, the age of the subjects is mostly 19-26 years. This age group is relatively young, and is the main force of online shopping, so it shows that the selected sample is representative. In terms of educational level, college and university students account for 97 percent of the sample, including 196 undergraduate students and 123 graduate students, accounting for 60% and 37% respectively. This indicates that the subjects have a higher level of education, so they have a higher level of acceptance for emerging e-commerce consumer finance. This proves that the experimental samples are highly reliable.

5.2 Reliability Analysis

This paper uses Cronbach α coefficient to test the reliability of each variable scale in this study. The test results of impulse purchase intention, positive emotion and self-construction (including dependent type and independent type) were 0.923, 0.903, 0.915 and 0.915, all above 0.7. These meet the requirements of reliability test, and show that the scale used in this experiment questionnaire is effective and reliable.

5.3 Hypothesis Test

5.3.1 Main effect test

SPSS19.0 is used to carry out one-factor ANOVA on the experimental sample, and the variance homogeneity test is first performed before one-factor ANOVA, and the test result is that the average value of consumers' impulse purchase intention is 4.5347 when the credit line is invariable. The average value of consumers' impulse purchase intention is 4.9158 when the credit line is increased. This means the increase of the credit line of e-commerce consumer finance can make consumers more impulsive to buy than when the credit line remains unchanged. Further analysis of variance (ANOVA) show $F(1,324) = 5.031$ 5.031, and the significant level $p = 0.026 < 0.05$, which indicates that the significant level is reached, so the hypothesis H1 is verified.

5.3.2 Intermediate Variable Test

SPSS19.0 is used to carry out regression analysis of the samples, and the three-step intermediary regression analysis method is used. This study involves three variables: independent variable e-commerce consumer finance credit line adjustment (X), dependent variables consumer impulse purchase intention (Y), intermediary variables positive emotion (M). First, the effect of independent variables on dependent variables is tested. The equation $Y=cX+e1$ was used to test. The test results are the regression coefficient is 0.381 and the significance is 0.026, which is below the significant level of 0.05. This indicates that the adjustment of independent variable credit line can explain impulse purchase intention significantly, which is consistent with the results of one-factor ANOVA.

At this point, it is not clear that there is an intermediary effect of positive emotion, and then the test between independent variable credit line adjustment (X) and intermediary variable positive emotion (M) is carried out using the regression equation $M=aX+e_2$. The results show that the regression coefficient $a=.880$, $p=0$ less, which is far below 0.05. This indicates that the independent variable credit line adjustment (X) could predict the positive emotion of intermediary variable (M). Then the relationship between intermediary variables (positive emotions) and dependent variables consumers' impulse purchase intentions (Y3) can be tested using the equation $Y=c_1X+bM+e_3$. The results show that the regression coefficient b exists and $p=0$ is extremely significant, and the regression coefficient c_1 and $p=.475>0.05$, so c_1 is not significant. At this point, the intermediary variable positive emotion (M) complete intermediary credit line adjustment (X) on the consumer impulse purchase intention (Y) can be concluded, so H2 is verified.

5.3.3 Adjustment Variable Test

Two-factor ANOVA is used to test the adjustment effect. For the product category and credit line adjustment, the result is that the $F(3,304)=15.788$, $p=0.000<0.05$, which reaches the significant level. Therefore, at least one of these two factors has an impact on consumers' impulse purchase intention. At the same time, the interactive effect of the product category of credit line adjustment * is $F(1,304)=10.133$, $p=.011<0.05$, which reaches a significant level. That is, the product category has a significant role in regulating the relationship between the credit line of consumer finance and the positive emotion of consumers, so H3 is verified. The positive emotion of hedonic goods when the credit line is increased (mean value 5.176) is greater than that when the credit line is maintained (mean value 3.907). Likewise, the positive emotion of utilitarian goods when the credit line is increased (mean value 4.767) is greater than that when the credit line is maintained (mean value 4.211). The position emotion for both hedonic goods and utilitarian goods when the credit line is increased is higher than that when the credit line is maintained, so H3aH3b is verified.

The results of ANOVA of self-construction and credit line adjustment are $F(3,304)=6.703$, $p=0.001$, which is far below 0.05, reaching and significant level. Therefore, at least one of the two factors has an impact on consumers' impulse purchase intention. The main effect of credit line adjustment is $F(1,304)=9.023$, $p<0.05$, which reaches a significant level. The self-construction * the interactive effect of credit line adjustment is $F(1,304)=13.783$, $p=0$, which is far below 0.05. This means self-construction significantly regulates the influence of e-commerce consumer finance credit line on consumers' positive emotion, so H4 is verified. For independent self-constructed consumers, the positive emotion when credit line is increased (mean value 5.206) is higher than that when the credit line is maintained (mean value 4.043), so H4a is verified; for dependent self-construction consumers, the positive emotion (mean value 4.713) when the credit line is increased (mean value 4.835) is lower than that the credit line is maintained (mean value 4.835), so H4b is verified.

6. CONCLUSION AND PROSPECTS

6.1 Research Conclusion

This paper examines the relationship between the adjustment of the credit line of e-commerce consumer finance and the impulse purchase intention of consumers, and on this basis, examines whether positive emotions have significant intermediary effects on impulse purchase when the credit line of e-commerce consumer finance is adjusted, as well as whether product category and self-construction have regulating effect on the relationship between e-commerce consumer finance credit line adjustment and consumer impulse purchase intention. The following conclusions are drawn:

6.1.1 Increased Credit Lines Can Have Significantly Positive Effect on Consumers' Impulse Purchase Intentions

Many scholars have confirmed that various marketing incentives can stimulate consumers to buy impulsively, such as discount coupons, time-limited buying, gift or complimentary voucher for consumption above a certain amount, the use of credit cards and so on^[16]. A study conducted by Wei Xin also confirmed that a temporary increase in credit card lines increased consumer spending with credit cards^[17]. Under the promotion environment of the “double eleven shopping festival”, e-commerce platform temporarily increase consumers’ financial credit line. This allows consumers to purchase their favorite discount products without having to spend cash balance, and the repayment period is relatively long, which makes consumers make unconscious impulsive consumptions.

6.1.2 The Intermediary Role of Positive Emotions

The increase of the credit line of e-commerce consumer finance has a significant positive impact on the positive emotions of consumers, and then makes consumers more impulsive to purchase. Under the promotion environment of the “double eleven shopping festival”, consumers are exposed to the information of product discount promotion and are offered increased credit line of consumer finance. They are likely to have more stimulated positive emotions, more active thinking and more rapid reactions, and their consumer consumption needs are better met, resulting in more impulse purchases.

6.1.3 The Regulatory Role of Product Categories And Self-Construction

Shiver Band Fedorikhin research shows that the higher the hedonic nature of the goods, the stronger the desire to buy. This study also shows that hedonic goods can stimulate consumers’ positive emotions more than utilitarian goods, which can lead to impulse purchase. Independent self-constructed consumers usually pay more attention to their own personality and their own attitude and emotion when making consumer decisions. Therefore, under the promotion environment of the “double eleven shopping festival”, consumers acquire a positive emotion of pleasurable excitement when facing the external marketing stimulations such as the increase of credit line of e-commerce consumer finance, thereby having strong impulse purchase intention; on the contrary, dependent self-construction consumers usually value more about whether their own behavior conforms to the social norms, and they are more self-disciplinary and concern more about group harmony.

6.2 Management Enlightenment

E-commerce platforms can increase consumer finance credit line when appropriate. In addition to temporarily increasing consumer credit line under the promotion environment in the “double eleven shopping festival”, e-commerce platforms can also adequately increase the credit line when appropriate so as to improve consumer’s consumer experience of using consumer financial products. For example, increase the credit line for consumers on consumers’ birthday. In addition, the channels for consumers to apply for credit line increase can also be opened up to meet the consumer needs for using consumer financial products to the greatest extent as well as improve consumer satisfaction and loyalty under the premise of ensuring risk control.

Online merchants should know how to stimulate positive consumer emotions. When making shop adjustments or designing product page layout, internet merchants should consider using elements that can make consumers produce more positive emotions, such as consistent overall style of the store and reasonable product layout. Promotion information should be clear at a glance and be highlighted on the page, especially for products that can be purchased using consumer finance products.

Different credit line management methods should be adopted for consumers with different personality characteristics. While managing the consumer credit line, E-commerce platforms can have an overall grasp on consumers’ personality characteristics and their attitude towards consumer finance based on their previous purchase records, and provide more credit line management measures. For example, dependent self-construction consumers put more value on the added significance of products, and e-commerce platforms can introduce the

social impact and value of consumer finance and so on when describing the consumer finance of e-commerce. While for independent self-construction consumers, they put more value on their own emotional feelings in the decision-making process, and they are not easily affected by the external environment, so e-commerce platforms can point out the conveniences brought by consumer finance credit line increase etc.

REFERENCES

- [1] Bellenger.D.N,Robertson.D.H,Hirschman.E.C(1978).Impulse buying varies by product. *Journal of Advertising Research*, pp15-18.
- [2] Sevgin A. Eroglu;Karen A. Machleitb and Lenita M. Davisb(2001).Atmospheric qualities of online retailing A conceptual model and implications[J].*Journal of Business Research*.Vol.54(NO.2): 177-184.
- [3] Hyunjoung Lee, Sohyoun Shin, Sanguk Kim(2012) . Surrogate Internet Shopping Malls: The Effects of Consumers' Perceived Risk and Product Evaluations on Country-of-Buying-Origin Image[J]. *Journal of Global Academy of Marketing Science*, 20(2):208-218.
- [4] Madhavaram S R, Laverie D A(2004).Exploring Impulse Purchasing on the Internet[J]. *Advances in Consumer Research*.
- [5] West J(1951).Results of two years of study into impulse buying.*Journal of Marketing*,pp362–363.
- [6] Stern H(1962).The significance of impulse buying today. *Journal of Marketing*,pp59-63.
- [7] Rook.D. W, Hoch.S.J, Consuming Impulses. *Journal of Advances in Consumer Research*,1985, pp23-27.
- [8] Rook.D.W, Fisher.R.J(1995).Normative influences on impulsive buying behavior, *Journal of Consumer Research*, pp305-313.
- [9] Hausman.A(2000).A multi-method investigation of consumer motivations in impulse buying behavior. *Journal of Consumer Marketing*., pp403-419.
- [10] Beatty S E, Ferrell M E (1998) . Impulse buying: Modeling its precursors.[J]. *Journal of Retailing*,74(2):161-167.
- [11] Singelis T M(1994). The measurement of independent and interdependent self-construals.[J]. *Personality & Social Psychology Bulletin*, 20(5):580--591.
- [12] Lee A Y, Keller P A, Sternthal B(2010). Value from Regulatory Construal Fit: The Persuasive Impact of Fit between Consumer Goals and Message Concreteness[J]. *Journal of Consumer Research*, 36(5):735-747.
- [13] Mandel N (2003). Shifting Selves and Decision Making: The Effects of Self-Construal Priming on Consumer Risk-Taking, in: *Journal of Consumer Research*, Vol. 30, No. 1, pp. 30-40[J]. *Journal of Consumer Research*,30(1):30-40.
- [14] YinlongZhang, Shrum L J(2009). The Influence of Self-Construal on Impulsive Consumption[J]. *Journal of Consumer Research*,35(5):838-850. (in Chinese)
- [15] Kacen J J, Lee A J A(2002). The Influence of Culture on Consumer Impulsive Buying Behavior[J]. *Journal of Consumer Psychology*, 12(2):163-176.
- [16] Youn S(2000). Faber R J. Impulse buying: Its relation to personality traits and cues[J]. *Advances in Consumer Research*.
- [17] Ian Phau(2004), Chang-Chin Lo,Profiling fashion innovators: A study of self-concept,impulsebuying and Internet purchase intent. *Journal of Fashion Marketing & Management*, pp399-411.

Research on Innovation Ecosystem of the Crowd Innovation

Space Based on System Dynamics

Bin Chen¹, Xin Jin^{2}, Jing Wu³, Hu Yang⁴*

¹School of Information, Central University of Finance and Economics, Beijing, 100081, China

²School of Information, Central University of Finance and Economics, Beijing, 100081, China

³School of Information, Central University of Finance and Economics, Beijing, 100081, China

⁴School of Information, Central University of Finance and Economics, Beijing, 100081, China

Abstract: The crowd innovation space is a new innovation business service platform, developed from the traditional incubator, including makerspace, entrepreneurship café and so on. It is essential for mass entrepreneurship and innovation to get full use of the creativity of the crowd innovation space and explore new impetus in New Economic Norm. Based on the traditional enterprise innovation ecosystem theory, this paper explores the connotation of innovation ecosystem of the crowd innovation space. And then, we use the knowledge to model the innovation ecosystem of the crowd innovation space with system dynamics. Based on the model, we explore the operational mechanism of innovation ecosystem of the crowd innovation space. Through the analysis of the operational mechanism and connotation, we put forward to some proposal to improve the effectiveness of the service ability in innovation of the crowd innovation space.

Keywords: crowd innovation space, innovation ecosystem, system dynamics, knowledge

1. INTRODUCTION

In the article entitled "Guiding Opinions of the General Office of the State Council on Developing Crowd Innovation Space to Promote Mass Entrepreneurship and Innovation" issued by General Office of the State Council of the People's Republic of China, it emphasizes that the crowd innovation space is a new service platform under the new trend of "mass entrepreneurship and innovation"^[1]. It establishes four basic principles: adhering to market orientation, strengthening policy integration, expanding the scope of openness and innovating service mode. The creative role of the crowd innovation space is taken as one of the four basic principles. It is significant to develop a new situation of "mass entrepreneurship and innovation" and explore a new acceleration under the New Economic Norm. Through the research at home and abroad, this article summarizes the connotation and characteristics of the crowd innovation space, modeling it with knowledge using system dynamics and exploring its operational mechanism so as to provide theoretical proposal.

2. RESEARCH STATUS

The crowd innovation space is an all-inclusive ecosystem, which takes innovation as the core, entrepreneurship as the way and low-cost, facilitation, total-factor, open and integrated business service platform as the carrier to integrate innovation and entrepreneurship, online and offline, incubation and investment. As the core of the ecosystem of the crowd innovation space, innovation plays an important role in impelling makers to start up business. In recent years, innovation ecosystem of the crowd innovation space research raises great concern on academic and industrial circles. It is an in-depth study of this new model of the crowd innovation space, based on innovative ecosystem research. The current status of related research will be described from three aspects, including innovation ecosystem, innovation ecosystem of the crowd innovation space and system dynamics.

* Corresponding author. Email: xin_jin@163.com (Jin Xin)

Previous research on innovation ecosystem begins with the study of innovation. The earliest concept of innovation was proposed by the Austrian economist Joseph Alois Schumpeter in his classic book "Economic Development Theories"^[2]. Schumpeter's concept of innovation involves five aspects, but it only restricted innovators to entrepreneurs, neglecting the influence of innovators such as colleges, universities and the crowd innovation space. After Freeman drew on Schumpeter's core idea of technological innovation economics, he began to study more about the impact of institutional environment on innovation and economy^[3]. Cooker (1992; 2004) and Malerba (1996; 2002) began to focus their research on regional innovation. They divided regional innovation into knowledge-based application and utilization subsystems as well as generation and diffusion subsystems^[4]. Knowledge is taken as the core element of innovation. The term "ecosystem" was first proposed by Moore. At this time, the ecosystem mainly refers to the enterprise ecosystem. It is a dynamic structural system with certain interests constituted by the stakeholders such as customers, suppliers and intermediaries^[5]. At this moment, Moore neither noticed the innovation ecosystem, nor noticed the important role played by the crowd innovation space, which represented by makerspaces and incubators. The core concept of innovation ecosystem comes from the metaphor of ecology concept and the expansion of its extension. Moore (1996) studied the attributes of self-organization evolution and diverse symbiosis of innovation system itself^[6]. Adner (2006) emphasized that the essence of innovation ecosystem is the coordination of heterogeneous stakeholders to realize value creation^[7]. At this stage, there are mainly two paradigms of innovation ecosystem: closed innovation ecosystem and open innovation ecosystem. Closed innovation system refers that enterprises increase investment in research and development costs within the enterprise, in order to obtain a breakthrough in key basic technologies, thereby improving product quality and service levels. Representatives of the closed innovation ecosystem include IBM's Watson Labs, DuPont's DuPont Labs and Lucent's Bell Labs. The open innovation ecosystem refers to the idea that the enterprise's idea is not only obtained from internal enterprise, but also obtained from the external enterprise. The commercialization path of the idea is obtained simultaneously inside and outside the enterprise. Innovation ecosystem of the crowd innovation space is a typical open innovation ecosystem. Vanhaverbeke (2006) proposed that an open innovation ecosystem is a new path for developing countries to shorten their stock of knowledge and technology with developed countries rapidly^[8]. Chesbrough (2014) emphasized that enterprises in open innovation have to internalize external knowledge (internal-opening) and externalize internal knowledge (external-opening), ultimately realizing their own innovation performance and improving their own innovation capability^[9]. Innovation Ecosystem of the crowd innovation space is a typical open innovation ecosystem.

Most previous research call "the crowd innovation space" as "hackerspaces" or "fab labs". Hansen (2000) and other specialists focused on the networking capabilities and networking services of the crowd innovation space^[10]. Benkler (2006) defined the crowd innovation space as an innovation model based on "mass innovation", emphasizing its innovative functions and innovative models^[11]. Hwang (2012) and others started to study innovation ecosystem and gave full attention to the role of the crowd innovation space. Hwang mentioned in the paper that there are only $(n-1) / 2$ communication nodes between the traditional innovation network and the innovator. However, among the creative eco-networks, there are $n(n-1) / 2$ communication nodes due to the existence of communication nodes represented by the crowd innovation space^[12].

Through the previous research, we can find that the crowd innovation space is an important part of innovation ecosystem, while knowledge is the most important measure of innovation. At the same time, it can be found that innovation ecosystem of the crowd innovation space is a typical open innovation ecosystem, which has the characteristics of internalization of external knowledge (internal-opening) and externalization of internal knowledge (external-opening). We can find that the crowd innovation space is a core node in it, which has strong function of network construction.

Now the domestic academic community is only in infancy for the research on innovation ecosystem of the crowd innovation space. Zeng Guoping and others (2013) focused on the dynamic process of innovation ecosystem among elements, system and environment^[13]. Chen Jin (2013) found that a good innovation ecosystem needs the courage to break the conventional culture for optimal deployment from the perspective of culture innovation^[14]. However, they focused mainly on the theoretical aspects of innovation ecosystem without combining with modeling. At this stage, Chen (2015) and others conducted a case study on the entrepreneurship ecosystem of the crowd innovation space operation model, but mainly concentrated on entrepreneurship ecosystem^[15]. Based on the theory of innovation ecosystem, Huang Shifang (2016) analyzed the connotation of the crowd innovation space and regional innovation system, and took the crowd innovation space as the micro-ecology of the regional innovation system, demonstrating the later advantages of the feasibility of the crowd innovation space in the underdeveloped areas^[16].

System Dynamics is a new discipline founded by professor J.W.Forrester, who worked at the Massachusetts Institute of Technology. It based on the feedback control theory and quantificationally studied complex system by computer simulation technology. The advantages of dynamics analysis are that it can give the trend of the system through quantitative analysis. Li Zhenhua (2009) and others applied this viewpoint to technology business incubators and conducted empirical simulations based on cases^[17]. Xu Xiaocang (2017) and others applied it to the efficiency evaluation of innovation ecosystem, building a simulation model of regional ecosystem, simulating and predicting four aspects of enterprise behavior, including government, innovation talents and research institutions^[18].

Through literatures of innovation ecosystem of the crowd innovation space mentioned above, we find that domestic theoretical research is relatively scarce and lack of quantitative analysis. The few studies just focused on entrepreneurship ecosystem of the crowd innovation space, while just like its name, it emphasizes innovation. According to our previous summary, innovation ecosystem of the crowd innovation space has the characteristics of internalization of external knowledge (internal-opening) and externalization of internal knowledge (external-opening) as a typical open innovation ecosystem. Utilizing this feature of knowledge, we can construct a causal feedback structure, while it has a great fit with the connotation and characteristics of the system dynamics. We can model systems by using knowledge as a variable in creating innovation ecosystem from the perspective of system dynamics. Through modeling analysis, it explores the operating mechanism and development path of innovation ecosystem of the crowd innovation space, so as to provide theoretic guidance for effectively improving the innovation service capacity and efficiency of the crowd innovation space.

3. ANALYSIS OF INNOVATION ECOSYSTEM OF THE CROWD INNOVATION SPACE

3.1 The connotation of the crowd innovation space

The main participants of innovation ecosystem of the crowd innovation space include: crowd innovation space itself, makers, start-ups, customers, universities, research institutions, governments, intermediary agencies and so on. In order to fully understand the connotation of innovation ecosystem of the crowd innovation space, we must clarify the positioning of each participant in it. The crowd innovation space can realize one-stop service, from innovation to entrepreneurship. The reason why the makers need a crowd innovation space is that the knowledge and equipment is not easily accessible to individuals. Universities provide a constant source of innovation for the crowd innovation space. Institutes focus on funding and the landing of experimental results. The government, as the watchdog of the innovation ecosystem, is more of a referee. Intermediary agencies are enterprises that point to start-ups offering various intermediary services.

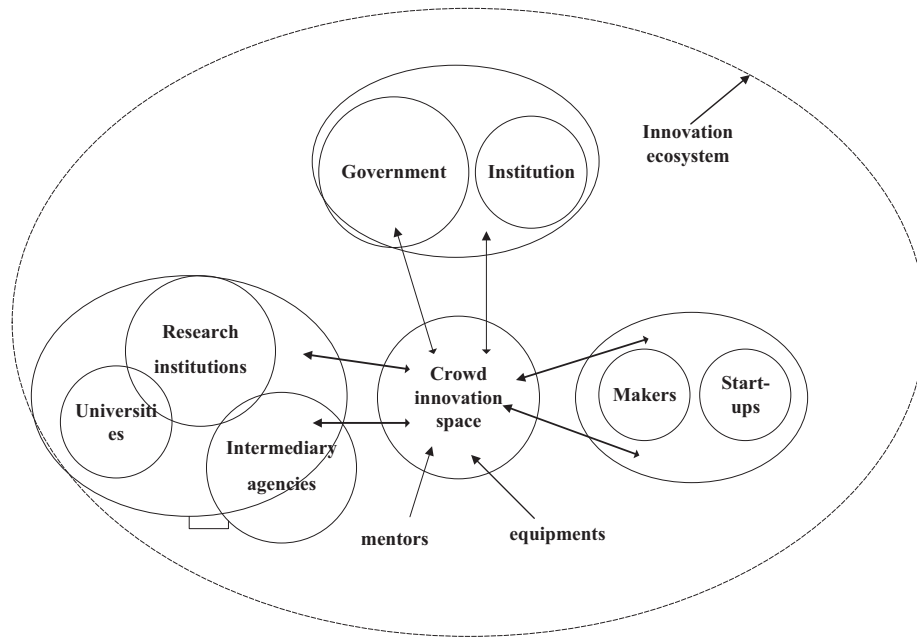


Figure 1. The ecosystem of crowd innovation space

3.2 Analysis of innovation ecosystem based on knowledge

Chesbrough proposed that valuable ideas could be obtained both internally and externally in open enterprise innovation ecosystem^[19]. Chesbrough divided innovation into positions, processes and paths. Positions represent the resources and knowledge that exists at this stage. Processes refer to the flow of knowledge and resources. Paths refer to the upgrading of knowledge and resources. In the innovation ecosystem, positions mean the existing knowledge distributed among the participating subjects. We can further divide the knowledge into internal knowledge and external knowledge. The knowledge that among the crowd innovation space which include the crowd innovation space itself, internal makers and start-ups, we can call it internal knowledge. The knowledge in universities, research institutions, intermediary agencies and external crowd innovation spaces, are called external knowledge. The growth of internal knowledge is mainly driven by the entrepreneurial mentors who generate creative knowledge by helping makers and start-ups solve problems. The entry of makers and start-ups can increase the internal knowledge and decrease the external knowledge. If it can be incubated successfully, the makers and start-ups will leave the crowd innovation space, which increase the external knowledge and decrease the internal knowledge.

With the characteristics of innovation ecosystem of the crowd innovation space, it can be modelled based on system dynamics using the causal feedback loop analysis. The internal knowledge and external knowledge are state variables, which fully describe the transformation of innovation ability in innovation ecosystem of the crowd innovation space. They satisfy the principle of the minimum set and independence. We constructed three feedback loops: two positive feedback loops and one negative feedback loop, based on the internal knowledge mentioned above.

Positive feedback loop 1: The mentors in the crowd innovation space, assist the makers and start-ups, which improve the performance of the crowd innovation space. The promotion of the crowd innovation space leads to higher rent, equity income and the government investment. The crowd innovation space will invest in intellectual capital with a certain percentage of the proceeds, such as inviting more entrepreneurial mentors, buying 3D printers and so on. All of this, can make the internal knowledge self-growth.

Positive feedback loop 2: the mentors in the crowd innovation space help the makers and start-ups solve the

problem of innovation, which will improve the reputation of the crowd innovation. The promotion of reputation will attract more start-ups and makers. It will bring their knowledge into the crowd innovation space, leading to an increase in internal knowledge and a decrease in external knowledge.

Negative feedback loop 1: Because of the service of the crowd innovation space, the start-ups will get a higher probability to get success. They will implement multiple rounds of financing, which will move them into a more mature enterprise stage and move away from the crowd innovation space. As the start-ups moving out, the knowledge which exists in it will augment the external knowledge and reduce the internal knowledge

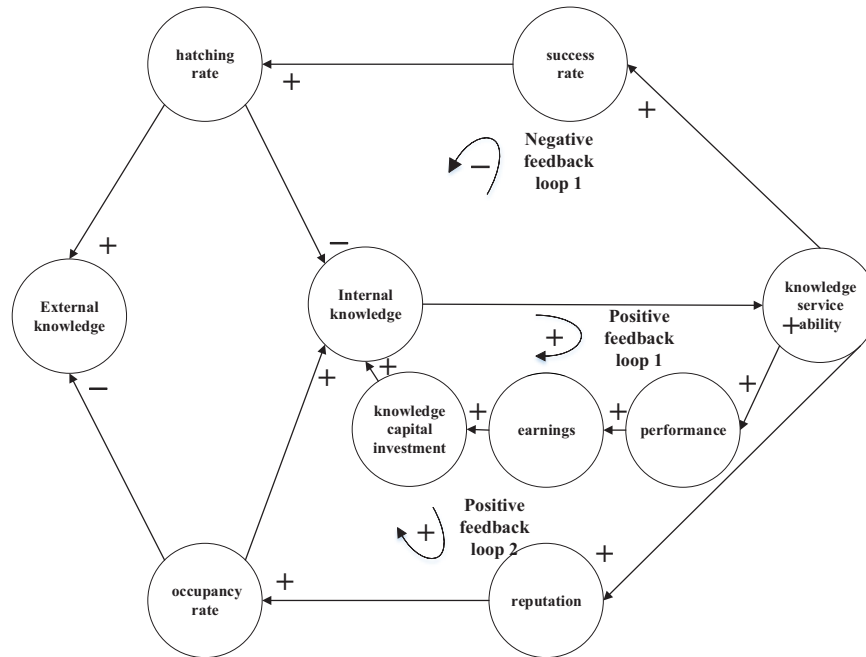


Figure 2 The feedback loop of innovation ecosystem of the crowd innovation space

4. MODELING AND SITUATIONAL ANALYSIS OF INNOVATION ECOSYSTEM OF THE CROWD INNOVATION SPACE BASED ON SYSTEM DYNAMICS

4.1 Modeling innovation ecosystem of the crowd innovation space using system dynamics

According to the feedback loops of innovation ecosystem of the crowd innovation space, we can determine the existence of two state variables, including internal knowledge and external knowledge. The rate variables are the change of the state variable in the unit time. we can determine five rate variables: internal knowledge endogenous growth rate, internal knowledge exogenous growth rate, internal knowledge reduction rate, external knowledge exogenous growth rate, external knowledge reduction rate. Another kind of variable is called auxiliary variables refer to the variables that follow the causal diagram, linking the state variables and the rate variables. We set more auxiliary variables in the model: the knowledge service ability, the performance, the space reputation, start-up success rate, the income of the crowd innovation space, the hatching rate and so on. One the basis of the establishment of causal analysis graph and model variables, the modelling of innovation ecosystem of the crowd innovation space, is constructed using system dynamics, as shown in figure 3.

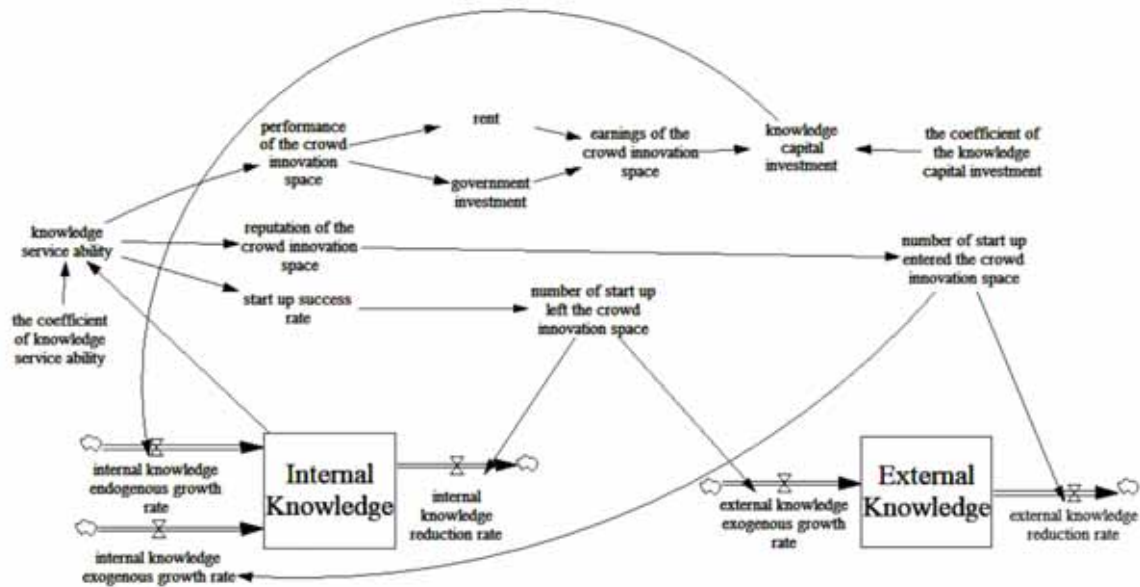


Figure 3. The modeling of innovation ecosystem of the crowd innovation space using system dynamics

4.2 Changing the coefficient of the knowledge capital investment

Keeping other variables unchanged, if we improve the coefficient of knowledge capital investment, the greater part of the earnings and will invest into the internal knowledge growth. This will make internal knowledge growth rate increased significantly. But because the growth of internal knowledge will also increase the number of incubators, which will reduce the external knowledge. The growth of internal knowledge increases the number of incubated enterprises, which increases the external knowledge. Therefore, the specific impact of external knowledge is still uncertain. After the expectation of the theory, the coefficients of knowledge capital investment are set as follow: condition1=0.1, condition2=0.11, condition3=0.13.

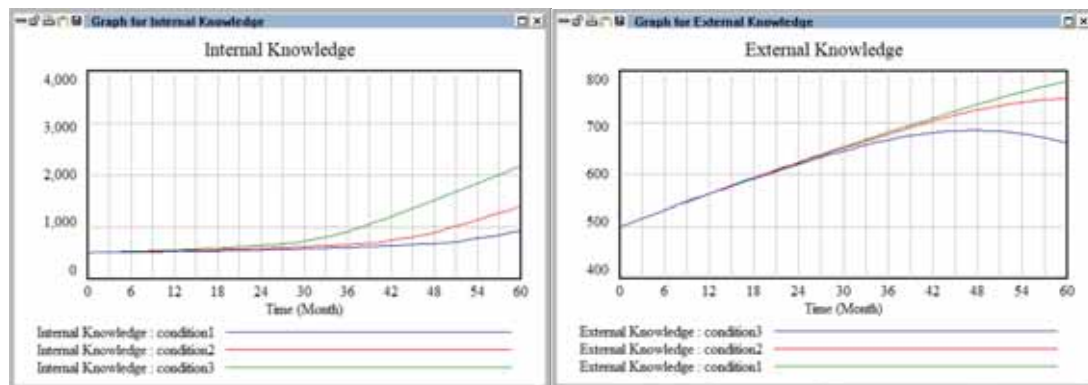


Figure 4. The evolution of knowledge under different coefficients of knowledge capital investment

The growth of internal knowledge is accompanied by diminishing marginal output, which conforms to the classical rice field conditions. By observing external knowledge, external knowledge also satisfies this condition. Although with the higher efficiency of knowledge capital investment, more external knowledge enters into the internal knowledge. The total knowledge still goes up steadily.

4.3 Changing government investment

Keeping other variables unchanged if we improve the government investment, it will increase the earnings

of the crowd innovation space. And then influence the growth rate of the internal knowledge under the condition that the coefficient of knowledge capital investment is constant. The effect of raising government investment would be similar to raising the coefficient of knowledge capital investment as to internal knowledge and external knowledge. So, we will discuss other variables such as performance of the crowd innovation space and internal knowledge endogenous growth rate. We set the maximum government investment as condition1=100, condition2=200, condition3=300.

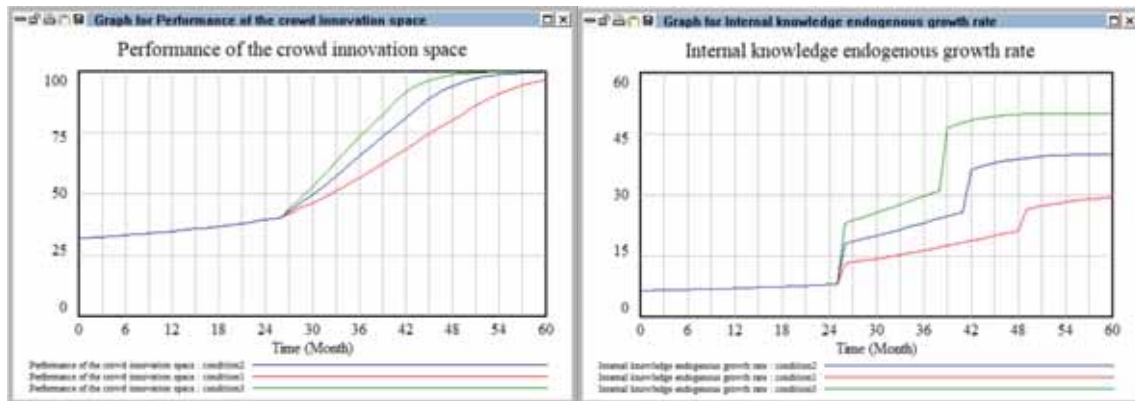


Figure 5. The evolution of innovation ecosystem under different government investment

By increasing the government investment, we can observe from the first chart that the greater government investment, the faster reaching the steady state value. But for the speed of steady state, there is diminishing marginal returns. The government investment can increase the endogenous growth rate of internal knowledge. And the increase of government investment can also shorten the time to steady state of the crowd innovation space.

5. CONCLUSIONS

In this paper, we first discuss the research status of innovation ecosystem of the crowd innovation space, finding that there is no systematic and theoretical research. Through the analysis of innovation ecosystem of the crowd innovation space, we model the system with knowledge based on the system dynamics. Finally, by adjusting the coefficient of the knowledge capital investment and government investment in the model, the following policy suggestions are proposed according to the scenario analysis of innovation ecosystem of the crowd innovation space using system dynamics.

First, the government should actively strengthen the investment in guiding funds for the crowd innovation space. It can be seen from the analysis of the model that government investment can effectively improve the innovation ability of the crowd innovation space and shorten mature time. Since there is diminishing marginal output effect, the government should make a balance between improving the innovation capacity of the crowd innovation space and avoiding waste of capital.

Second, the crowd innovation space should increase the knowledge capital investment. The knowledge capital investment here mainly refers to two aspects: on the one hand, it is the internal intellectual capital investment, such as hiring more mentors and expanding office space; on the other hand, it is the external knowledge capital investment, such as strengthening the crowd innovation space netting ability, associating the upstream and downstream enterprises and implementing one-stop service from knowledge to the innovation.

ACKNOWLEDGEMENT

This research was supported by the Beijing Social Science Foundation (No. 15ZDA50).

REFERENCES

- [1] General Office of the State Council of the People's Republic of China. (2005). Guiding Opinions of the General Office of the State Council on Developing Crowd Innovation Space to Promote Mass Entrepreneurship and Innovation. http://www.gov.cn/zhengce/content/2015-03/11/content_9519.htm in Chinese
- [2] Schumpeter J A. The Theory of Economic Development. (1934). Cambridge: Harvard University press,
- [3] Freeman C. The economics of industrial innovation. (1997). University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship
- [4] Cooke P, Gomez Uranga M, Etxebarria G. (1997). Regional in-novation systems: Institutional and organisational dimensions. *Research Policy*, 26 (4): 475-491.
- [5] Moore J F. (1993). Predators and prey: A new ecology of com-petition. *Harvard Business Review*, 71 (3) :75-86.
- [6] Moore J F. (1996). The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems. New York: Harper Business,
- [7] Adner R. (2006). Match your innovation strategy to your innovation ecosystem. *Harvard Business Review*, 84:98-107
- [8] Vanhaverbeke W, Cloudt M. (2006). Open innovation in value networks. *Oxford University Press*, 258-281.
- [9] Chesbrough H, Bogers M. (2014). Explicating open innovation//Chesbrough H, Vanhaverbeke W, West J. (2014). *New Frontier in Open Innovation* Oxford: Oxford University Press
- [10] Hansen Chesbrough, Nohria, Sull. (2000). Networked Incubators-Hothouses of the New Economy. *Harvard Business Review*, 78(5):74-84
- [11] Benkler Y. (2006). The wealth of networks: How social production transforms markets and freedom. New Haven: Yale University Press,
- [12] Hwang V W, Horowitz G. (2012). The Rainforest: The Secret to Building the Next Silicon Valley. Los Altos Hills: Regenwald
- [13] Zeng Guoping, Gou Youli, Liu lei. (2013). From innovation system to innovation ecosystem. *Studies in Science of Science*, 31(1): 4-12 in Chinese
- [14] Chen Jin. (2013). Innovation management and its future. *Technology Economics*, 32(6): 1-9 in Chinese
- [15] Chen Su, Xiang Liyao, Yu Rongjian. (2015). Co-maker Space Entrepreneurship Ecosystem: Features, Structure, Mechanism and Strategy—Case Study on Hangzhou Dream Town. *Journal of Business Economics*, 35(11): 35-43 in Chinese
- [16] Huang ShiFang. (2016). Public Innovation Space and the Construction of the Regional Innovation System: From the Perspective of Less Developed Areas. *Journal of GuangXi University for Nationalities*, 38 (1): 156-160 in Chinese
- [17] Li Zhenhua, Zhao Liming, Shi Quan. (2009). SD Modeling of the Development of High-Tech Business Incubators. *Science of Science and Management of S.& T*, 30(5): 31-35 in Chinese
- [18] Xu Xiaocang, Liu Junli. (2017). Simulation and Optimization of Chongqing Industrial Ecological Innovation System Based on System Dynamics. *Science and Technology Management Research*, 37(4): 12-18 in Chinese
- [19] Chesbrough H W. (2006) *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston: Harvard Business School Press

Inter-organizational Information Technology and Joint Competitive

Advantages: An Integrative Model of Co-creating IT value

Lu Sun^{1}, Chengyi Yue²*

¹ Business College, Jiangxi Normal University, China

² Department of Statistics, East China Jiaotong University, China

Abstract: The purpose of this paper is to facilitate knowledge accumulation and creation concerning IS research by summarizing an extension of the resource-based view and IT value literature. We define several constructs, including, inter-organizational IT resources & capabilities, and IT co-creating rents and illustrate a typology of inter-organizational IT resources and their attributes. Finally, we develop a conceptual model of co-creating IT value that integrates the above constructs. Our analysis provides a blueprint to examine the relational rents impacts of inter-organizational IT and motivates research incorporating the RBV and the extended RBV in the field of IS.

Keywords: Co-creating IT Value, Resource-based View, Rents, Inter-organizational IT Resources, Inter-organizational IT Capabilities

1. INTRODUCTION

Over the past twenty years, scholars always want to understand that the relationship between information technology (IT) and firm performance, and how to achieve competitive advantages. Most researches indicated that IT indeed creates value^[1-5]. Besides, more and more firms build and share inter-organizational IT resources, co-create value in collaboration. For example, both HP (Hewlett Packard) and UPS (United Parcel Service) co-create sustainable relational value through collectively developing inter-organizational logistics system. Thus, with internetworking technologies, there is a fundamental transformation taking place in the creation of business value. That is, Value Cocreation based on emerging IT business system became increasingly. Comparing IT resources and capabilities, inter-organizational IT resources and capabilities are more socially complex and difficult to imitate. A question is raised: how to leverage IT to co-create value in multi-firm environments? There is very little discussion on the co-creating IT Value. This paper provides inter-organizational IT resources and capabilities, and IT co-creating rents based on relational view, which is helpful to explain “co-creating IT Value” in multi-firm environments.

The paper is organized as following: (1) discuss “co-creating IT Value” based in IT value literature and define “co-creating IT Value”; (2) classify inter-organizational IT resources based on the resource-based view, analyze IT co-creating rents based on an extension of the resource-based view, and expound inter-organizational IT capabilities; and (3) develop a model of co-creating IT Value.

2. DEFINING CO-CREATING IT VALUE

Co-creating IT Value occur in co-opetition, IT-based environment, which is made by multiple companies in cooperative. Co-creating IT Value represents the idea that (1) the value is co-created by multiple parties involving a symbiotic relationship between a firm and its primary stakeholders^[6], (2) the value cannot be generated by either firm in isolation, which is more than sum of value generated by either firm in isolation^[7,8], (3) the main goal is to increase market demand and enlarge market space, rather than fighting with rivals over the market share, in other words, the aim is to promote overall industrial profits, including collaborators and competitors^[9], and (4) compared with IT value research examining the organizational performance impacts of

* Corresponding author. Email: lusun_jxnu@qq.com (Lu Sun)

information technology, co-creating IT value research examine the relational value impacts of inter-organizational IT resources and capabilities in multi-firms environments. The relational value is defined as a collaborative advantage made by multi-firms in cooperation that is a joint competitive advantage hold by partners. For example, customer satisfaction, share of wallet and loyalty^[10], abnormal return^[9], alliance innovation performance^[5], and so on.

3. INTER-ORGANIZATIONAL IT RESOURCES AND THE EXTENDED RESOURCE-BASED VIEW

The resource-based view (RBV) argues that competitive advantage derive from resources and capabilities of the firm that emphasizes heterogeneity and imperfectly mobile of resources^[11]. However, the value resources have spanned boundary of firm and have be embedded into inter-organization. This specific inter-organizational relationship may be sources of competitive advantages or the relational rents^[7]. Some researchers have broken the resource-based view of firm in isolation and extended to multi-firms environments. Dyer & Singh (1998) suggest that critical resources of a firm may span boundaries of firm and may be embedded into routines and processes of inter-organization. Therefore, the firm is to earn not only Ricardian rents and quasi-rents, but also relational rents. Moreover, the specific inter-organizational relationship has been sources of competitive advantages and relational rents. The common of all above studies are obviously: based on the RBV and spanning the boundaries of a firm, which are extensions of the RBV.

3.1 A Typology of Inter-organizational IT Resources

IT relation-specific assets IT infrastructure is invested by single firm that can be imitated. Hence, it is hard to acquire competitive advantage. However, when the firm invests in interfirm relation-specific assets with partners, it can gain competitive advantage^[12]. Firms gain advantages by creating specific assets in collaboration. Relation-specific assets are necessarily for “rent” and naturally are strategic assets^[13]. Early results show that the relational rents are generated by investing specific relation. The greater the alliance members’ investment is in relation-specific assets, the greater the potential will be for relational rents^[7]. Hence, firms in networked environments produce digital or physical products & services and gain competitive advantage through creating IT relation-specific assets in collaboration. IT can strengthen the safeguard by amplifying openness of use of assets and lowering transaction cost between partners. Besides, IT can enhance trust between partners through amplifying openness and increase the volume of transaction.

IT knowledge-sharing routines “IT knowledge-sharing routines” can be defined as standardization in inter-firms that can improve knowledge sharing and absorptive capacities, can improve partnership in collaboration, collectively create new product and service, and achieve co-create value. For example, Wal-Mart Stores, Inc. operates retail stores in various formats around the world. The use of information technology has been an essential part of Wal-Mart's growth. Wal-Mart have made achievement in saving cost taking advantage of IT and designing logistics systems that attain competitive advantage from information technology. Such as, both Wal-Mart and Procter & Gamble co-create supply chain collaboration mode that is CPFR (collaborative planning, forecasting, and replenishment, CPFR) based on IT. CPFR naturally is an IT knowledge-sharing routine. Furthermore, the data center of Wal-Mart has developed knowledge-sharing routines with thousands of suppliers and realized rapid response VMI (Vendor Managed Inventory, VMI). The supplies can directly access into the data center of Wal-Mart by VMI and timely know about dynamic processes of distribution, which are bases of producing and distributing. In summary, both CPRF and VMI based on IT take great benefits for Wal-Mart, suppliers, and manufacturers.

Complementary resources and capabilities by IT Complementary resources and capabilities are defined as specific resources of an alliance partner that co-create “rents” are more than add of “rent” created by each alliance member in isolation. The resources creating “rents” in collaboration cannot be purchased by

partners in market^[7]. Therefore, complementary resources and capabilities is a source of relational rents. Not all of resources and capabilities in alliances are complementary. To acquire complementary resources and capabilities, the firm should identify these firstly. IT is a valid instrument of recognizing and assimilating complementary resources and capabilities^[7,14]. For by IT/IS, it is easier to form a collaborative trust in relationship between suppliers, rivals, and customers^[12]. Based on these collaborative trusts, it is easier for firm to identify complementary resources and capabilities and to generate higher value of cooperative resources and competitive advantages.

Based on the above mentioned, we define “Complementary resources and capabilities by IT” as the firm identify and utilize complementary resources and capabilities of other partners depending on IT in collaboration. The value that is co-created by “Complementary resources and capabilities by IT” cannot be created by resources and capabilities of the firm in isolation. Hence, “Complementary resources and capabilities by IT” is also a source of relational value. For example, both General Mills Inc. and O’Lakes Inc. are suppliers of food retails in America. And, both of them are partners of Nistevo. Nistevo is the leading Collaborative Logistics Network for transportation management and is a hosted software service that enables manufacturers, retailers, distributors and logistics service providers to view, plan, execute, settle and analyze their inbound and outbound transportation. In 2000, General Mills and Land O’Lakes established an Innovative Supply Chain Alliance. The aim of the Alliance is to synergy purchasing process and improves customer service by the collaborative logistics network provided by Nistevo. Through sharing complementary ordering and distributing information of both General Mills and Land O’Lakes between distributors and retailers, the costs of inventory are reduced, customer services are improved, and synergy is generated. The collaborative logistics network provided by Nistevo between General Mills and Land O’Lakes is a “Complementary resources and capabilities by IT”. In the process, IT plays an important role in recognizing and utilizing complementary resources and capabilities of partners.

IT governance resources The effective governance can reduce transaction cost and derive value co-creation. Therefore, governance is important to relational rents^[7]. In this paper, “IT governance resources” is effective governance that can reduce transaction cost and promote information integration and information exchange based on IOS (Inter-organizational Information Systems, IOS) that is invested by partners in collaboration. In essence, the joint investment in information technology becomes the informal contract. The contract is neither in the form of text identifying responsibility and rights of both, is nor technological contract in technology innovation. The contract is similar to psychological contract in organizational behavior, which is the invisible contract between partners in collaboration. Comparing to the formal contract, the informal contract is characteristic of high sunk cost. Some research argued that the informal contract can protect investment in lowest cost and improve exchange between partners^[11]. For example, compared to the formal governance mechanism (e.g., financial hostages) based on contract, the informal governance (e.g., trust) can also play an important role in co-creating value. Furthermore, the lower the contract costs are, the greater the potential will be for relational rents. Actually, “IT governance resources” in collaboration is hard to imitate because the informal contract that jointly developed by partners investing in information technology is more complex in social relationship and more specific in communication. For example, Taobao.com which is the digital platform integrates thousands of retailers and opens in aspects of IT resources, operational services, financial services, logistic services, customer services. The digital platform provides the one-stop solution between retailers and customers and develops the e-commerce ecosystem with partners. In fact, Taobao.com is an “IT governance resources” that is invested by Alibaba. Taobao.com significantly reduces transaction cost and develops long-term trust with customer by providing commodity transaction records and customer evaluation records, and improving information integration and information exchange in this digital platform. The digital platform

achieves co-create value based on IT that other traditional assets cannot create.

Overall, “IT governance resources” is an integration of digital platform, involving IT relation-specific assets, IT knowledge-sharing routines, and complementary resources and capabilities by IT. Firms can improve information integration and information communication, and efficiently manage cooperation between partners through integrating IT governance resources and utilizing the digital platform.

3.2 Inter-organizational IT Resources Attributes

Value A resource has value when it enables a firm to implement strategies and improve firm performance in view of the RBV^[11]. If the resource has not value or a little value, the firm is hard to achieve competitive advantage. The extended RBV argued that the value of resource is important for firms to achieve competitive advantages in network environments. Some researchers have examined change of market value when firms entry into alliances and found that significant positive abnormal returns of the allying firms^[9]. So the value of resources is not only limited in internal but also in collaboration. Furthermore, outside-in and spanning resources seem to have potentially higher value than internal resources to firms. As noted earlier, IT resources are value for firms both in internal, in competitive environment including partners and competitors, and in macro environment^[3]. Inter-organizational IT resources that are derived from in complex organization and society tend to be socially complex. In network environments, inter-organizational IT resources have value and are more value than IT resources.

Rarity If resources that are valuable have been available to a large number of firms and are in abundant supply, the valuable resources cannot become sources of competitive advantages in an RBV context. According to Amit & Schoemaker (1993), the valuable resources have rarity when they are not available to many firms. Although inter-organizational IT resources are more open than IT resources, they have rarity. For example, Amazon opening platform that is an IT governance resource has significant value to Amazon.com and its partners. Although the platform co-create value for partners, it is rare for a large number of firms. That is, for many firms, if they want to achieve e-commerce service efficiently and rapidly, the opening platform is the only choice. Also, the opening platform is controlled and appropriated by only Amazon.com and it is most likely to confer a strategic benefit to Amazon.com. In network environments, inter-organizational IT resources have rarity.

Appropriability Resources and capabilities that lead to competitive advantage must be owned and controlled by a single firm in the proprietary assumption of the RBV. Furthermore, the advantage may not be competitive if the firm is unable to appropriate the returns accruing from the advantage in conventional RBV studies^[15]. According to Wade & Hulland (2004), the appropriability of the spanning IT resources tends to be lower than that of internal IT resources. Besides, the advantages created by inter-organizational IT resources and capabilities are shared with cooperators and competitors. Therefore, inter-organizational IT resources have not appropriability in an RBV context. However, in recent years, many studies show that a firm’s critical resources may span firm boundaries. Network resources can also lead to competitive advantages for firms in collaboration. The advantages of an individual firm are increasingly linked to the advantages of the network of relationship in which the firm is embedded^[7]. Therefore, Lavie (2006) argued that the proprietary assumption of the RBV hinders an exact evaluation and understand of a firm’s competitive advantage in network environments. According to Lavie (2006), the extended RBV in network environments relaxed appropriability assumption that enriches an exact understand of a joint competitive advantage. As noted earlier, inter-organizational IT resources that are owned by partners and embedded in interfirm routines and processes are typically network resources. In brief, relaxing the proprietary assumption of the RBV in network environments allow for the inter-organizational IT resources to create a joint competitive advantage in collaboration.

Above three attributes are belonged to ex ante limits to competition, which means that before any firm’s

developing a competitive advantage, these must be limited competition for that advantage^[15]. Then, another type of resource attributes, ex post limits to competition, will be analyzed. Ex post limits to competition suggest that after a firm's gaining a competitive advantage and earning rents, these must be limited competition for keeping rents^[15]. Attributes in this category include inimitability, non-substitutability, and imperfect mobility.

Inimitability If the advantage is imitated by competitors, there is typically short-lived and is unable to sustain. According to Barney (1991), there are three factors that can lead to inimitability: unique firm history, causal ambiguity, and social complexity. Inter-organizational IT resources are typically establishing between large firms and their partners. These firms have unique firm history that other firms are unable to imitate. For example, Google Inc. purchased Android OS in 2005 that subsequently becomes very valuable. Inter-organizational IT resources are more complex than IT resources and these are ambiguity that exists in how an inter-organizational IT resource leads to the sustained competitive advantage. So it is hard for competitors to imitate these resources or copy the way in which these are deployed^[15]. Inter-organizational IT resources are likely to be socially complex, which are derived from embedding of firms in networks. Thus, in network environments, firms and their partners defend their advantage against imitation by competitors outside network through segregation mechanism, such as causal ambiguity. Inter-organizational IT resources are likely to be more difficult to imitate than IT resource of a single firm. In addition, inter-organizational IT resources have opening. For example, Android OS have several third-party application service platforms, such as Google Play Store, Wandoujia, and Mi App. Therefore, the opening may reduce inimitability of inter-organizational IT resources. In a word, inter-organizational IT resources have inimitability, however, in some opening environments, which would be lowed.

Non-Substitutability A resource has non-substitutability if it has rarity and inimitability^[13]. Non-substitutability of a resource may depend on whether it has strategically alternative resources and it can lead to an equivalent performance. According to Wade & Hulland (2004), strategic substitutes for the outside-in and spanning resources are also likely to be rare. In the case of inter-organizational IT resources, it probably impossible that strategic equivalent resources exist that leads to the same competitive advantage. Therefore, inter-organizational IT resources have low substitutability. As noted in the above paragraph, it is likely to reduce inimitability in opening environments. Equally substitutability of these resources would be increased. Thus, inter-organizational IT resources have non-substitutability, however, in some opening environments, which would be lowed.

Imperfect Mobility In the context of the RBV, if firms are able to acquire and utilize the resources to imitate a rival's competitive advantage at zero cost, the rival's competitive advantage will be short-lived and the resources are mobile or tradable. On the contrary, if firms are difficult to obtain the resources to imitate a rival's competitive advantage, the advantage is sustained and the resources are imperfect mobile or non-tradable. For example, compared to technological assets of a single firm, inter-organizational IT resources in collaboration, such as IT knowledge-sharing routines, are hard to acquire and are imperfect mobile. Besides, as noted earlier, inter-organizational IT resources have opening. For example, an IT governance resource that is typically appropriated by a firm can be utilized by other partners in network environments at very low cost or at zero cost. However, Lavie (2006) have weakened the imperfect mobility condition in an extension of the RBV view: the network environments, such as alliances, are able to transfer of benefits associated with the resources that cannot be mobilized. Thus, even though the imperfect mobility condition may be weakened in some opening environments, inter-organizational IT resources have imperfect mobility.

As noted earlier, inter-organizational IT resources have interdependence, and path dependence. According to Grover & Kohli (2012), interdependence refers to an inter-organizational IT resource enable to advance investing another inter-organizational IT resource. For example, an IT governance resource can stimulate

knowledge sharing between partners, lead to greater investment in IT relation-specific assets, enhance synergy in collaboration, and provide incentives for value co-creation initiatives. Path dependence involved that value creation of one type of resource can create the option for further value^[14]. For example, these are typically great costs in the processes of knowledge sharing between partners^[7]. If firms enable to utilize inter-firms IT resource and capabilities, such as, an effective IT governance mechanism, these can stimulate alliance members to increase the degree of openness, preclude opportunistic bargaining, prevent against free-riding, and reduce transaction costs.

Table 1. Attributes of Inter-organizational IT Resources

Inter-organizational IT Resources	Advantage Creation			Advantage Sustainability		
	Value	Rarity	Appropriability	Imitability	Substitutability	Mobility
IT relation-specific assets	high	high	medium	medium	low-medium	medium-high
IT knowledge-sharing routines	high	high	medium	low-medium	low	low
complementary resources and capabilities by IT	high	high	medium	low-medium	low	low
IT governance resources	high	high	medium	medium	low-medium	medium-high

4. AN INTEGRATIVE MODEL OF CO-CREATING IT VALUE

4.1 IT Co-creating Rents

IT-Based Co-Creation of Value occurs in IT-based alliances or network environments. Lavie (2006) illustrated four different types of rents in network environments. However, the process of Co-creating IT Value is different from the commonly process of Value Co-creation in network environments. We define “IT Co-creating Rents” as the specific rents occur in the process of Co-creating IT Value that are derived from inter-organizational IT resources, which include IT relational rent and IT-based outbound spillover rent. Borrowing from framework and terminology used by Lavie (2006), composition of rents extracted by the focal firms in process of co-creating IT value is illustrated in figure 1: (1) internal rent is composed of Ricardian rents and quasi-rents of focal firm; (2) Regarding IT relational rent, alliance members jointly utilize inter-organizational IT resources and acquire a joint competitive advantage. IT relational rent that is part of appropriated relational rent cannot be created by a firm in isolation and can be co-created in alliance layer; (3) inbound spillover rent is derived from shared resources and non-shared resources of partners; (4) IT-based outbound spillover rent is derived from competitive advantage taken away by nonparticipating rivals. The first three rents have positive effects on competitive advantage. In contrast, IT-based outbound spillover rent has a negative effect on competitive advantage. However, it has a positive effect for nonparticipating firms. Besides, only IT relational rent involves distribution of the value. Both IT relational rent and IT-based outbound spillover rent compose IT co-creating rents. The former that is derived from inter-organizational IT resources is easy to understand, while the latter relatively complex and will be illustrated in the following.

Whether intended or unintended, Knowledge spillover typically occurs in the process of Co-creating IT Value in network environments. Some researchers have suggested that substantial knowledge spillover when firms or rivals outside alliances exploit the technological innovations created from the IT-based alliances at lower cost and without having to reinvent the wheel^[9]. Rent of the knowledge spillover that occurs in the IT-based alliances is different from outbound spillover rent of common network environments, which is defines as IT-based outbound spillover rent. IT-based outbound spillover rent is derived from nonparticipating firms utilizing and exploiting shared IT resources between partners. Nonparticipating firms and rivals outside alliances can replicate the innovation and improve performance without incurring significant additional costs or at zero cost through knowledge spillover. Knowledge spillover contributes to promote profits of an industry and

develop a business ecosystem. In turn, promotion of competitive advantage of an industry can lead to additional profits for alliance members. The profits are more than the rents that are taken away by nonparticipating rivals exploiting the shared inter-organizational IT resources.

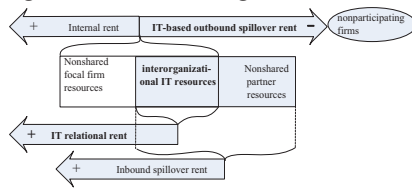


Figure 1. Composition of Rents Extracted by the Focal Firms in Process of Co-creating IT Value

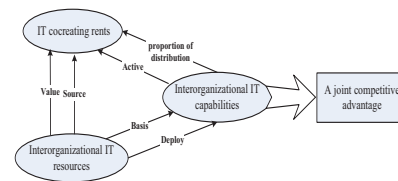


Figure 2. An Integrative Model of Co-creating IT Value

4.2 Definition of Inter-organizational IT Capabilities

As we noted earlier, we identify the relationship between inter-organizational IT resources and IT co-creating rents. However, these are questions in practice: why does the same IT investment lead to different performance for firms and alliances? high profits, low profits, and loss. The resource is a static factor of value in conventional RBV studies. Thus, IT resources are hard to have a direct impact on a sustained competitive advantage and are unable to explain proportion of distribution of rents^[12].

Recent years, inter-organizational IT capabilities are increasingly being studied^[5,10]. Compared with IT resource, it is the extension of IT resources and more difficult to imitate, substitute, and mobilize than IT resource. Borrowing from definition proposed by Bharadwaj (2000), inter-organizational IT capabilities refer to an organization's abilities to assemble, integrate, and deploy inter-organizational IT resources, and synergy with other complementary assets and capabilities in collaboration. Inter-organizational IT capabilities lead to increase cooperation performance, achieve relational value, and enhance the industry's total profitability.

4.3 An Integrative Model

In this section, we use inter-organizational IT resources, IT co-creating rents, and inter-organizational IT capabilities illustrated above to develop an integrative model of co-creating IT value (seeing in figure 2). The relationships between above three factors are summarized in the following.

Firstly, inter-organizational IT resources are basis of inter-organizational IT capabilities. Inter-organizational IT capabilities can achieve relational value and lead to competitive advantage, which are abilities embedded in organization process to integrate and deploy inter-organizational IT resources. Secondly, inter-organizational IT resources are basis of IT co-creating rents. IT co-creating rents is expression of value of inter-organizational IT resources. That is to say, IT co-creating rents is expression of relational value and competitive advantage in the sense of economics. Finally, inter-organizational IT capabilities utilize inter-organizational IT resources and other complementary assets to active value of these resources. Therefore, inter-organizational IT capabilities can generate IT co-creating rents and competitive advantage. In addition, inter-organizational IT capabilities are specific organization's abilities and every firm in network environments has different inter-organizational IT capabilities.

5. Conclusion

With internetworking technologies and the key resource of firms beyond the boundaries, co-creation of value in multiple-firm environments has been a typically way of value creation. In Particular, multiple firms collectively leverage IT to co-create IT-enabled products and services. Thus, there is a fundamental transformation arising in the conventional IT value research and co-creating IT value have been a critical theme in IS research. The extended RBV provide the way for IS researchers to understand the role of inter-organizational IT resources and capabilities in cooperative, platform-based, opening, and hypercompetitive

environments. Borrowing from definition and theory used by Dyer & Singh (1998) and Lavie (2006), we define inter-organizational IT resources and capabilities, IT co-creating rents, illustrate each inter-organizational IT resource with example of firms that co-created value by exploiting IT in this paper. Finally, we develop a conceptual model of co-creating IT value that integrates above three key constructs and their interrelationship.

Co-creating IT value research is a key new theme in IS discipline, our knowledge of which remains underdeveloped and unsystematic. In sum, Co-creating IT value offers a set of research questions that should be urgently addressed. It is our hope that the issues, definitions, and discussions illustrate in the paper will motivate interest and research incorporating the RBV and the extended RBV in the field of IS. Case and empirical studies are required to build a foundation for understanding the relational value and a joint competitive advantage impact of inter-organizational IT resource and capabilities in future.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71761012 and the Jiangxi Normal University doctorate fund "Research on the formation mechanism of information interaction capability of enterprise".

REFERENCES

- [1] Lim MK, Tseng ML, Tan KH. (2017). Knowledge management in sustainable supply chain management: Improving performance through an interpretive structural modelling approach. *Journal of Cleaner Production*, 162: 806-816
- [2] Lioukas CS, Reuer JJ, Zollo M. (2016). Effects of Information Technology Capabilities on Strategic Alliances: Implications for the Resource-Based View. *Journal of Management Studies*, 53(2): 161-183
- [3] Melville N, Kraemer K, Gurbaxani V. (2004). Review: Information Technology and Organizational Performance: An Integrative Model of IT Business. *MIS Quarterly*, 28(2): 283-322
- [4] Revilla E, Knoppen D. (2015). Building knowledge integration in buyer-supplier relationships. *International Journal of Operations & Production Management*, 35(10):1408-1436
- [5] Xue XF, Huo BF, Xu W. (2013). Research on Effect of IT Capability on Strategic Alliance Performance. *Science Research Management*, 34(12):326-333 (In Chinese)
- [6] Kohli R, Grover V. (2008). Business Value of IT: An Essay on Expanding Research Directions to Keep up with the Times. *Journal of the Association for Information Systems*, 9(1):23-39
- [7] Dyer JH, Singh H. (1998). The Relational View: Cooperative Strategy and Sources of Inter-organizational Competitive Advantage. *Academy of Management Review*, 23(4): 660-679
- [8] Lavie D. (2006). The Competitive Advantage of Interconnected Firms: An Extension of the Resource-Based View. *Academy of Management Review*, 31(3):638-658
- [9] Han K, Oh W, Kun SI, et al. (2012). Value Cocreation and Wealth Spillover in Open Innovation Alliances. *MIS Quarterly*, 36(1):291-325
- [10] Rai AA, Pavlou P, Im G, et al. (2012). Interfirm IT Capability Profiles and Communications for Cocreating Relational Value: Evidence from the Logistics Industry. *MIS Quarterly*, 36 (1):233-262
- [11] Barney JB. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17(1):99-120
- [12] Zhang J, Huang JH, Gao H. (2010). An Integrated Model for Evaluating IT Value in Firms from Resource-based View. *Science of Science and Management of S&T*, (2):130-136 (In Chinese)
- [13] Amit R, Schoemaker PJH. (1993). Strategic Assets and Organizational Rent. *Strategic Management Journal*, 14: 33-46
- [14] Grover V, Kohli R. (2012). Co-creating IT Value: New Capabilities and Metrics for Multifirm Environments. *MIS Quarterly*, 36(1):225-232
- [15] Wade M, Hulland J. (2004). Review: The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research. *MIS Quarterly*, 28(1): 107-142

Exploring Gender Effects on Peer Rating in Open Innovation and Crowdsourcing: A Case of Website Evaluation

Liang Chen

Department of Computer Information and Decision Management,
West Texas A&M University, USA

Abstract: Peer rating has been used by open innovation and crowdsourcing platforms to evaluate submissions and select winners because it not only represents a cheaper and more scalable way but also empowers and engages users. However, the literature on scholarly peer review suggests that peer rating may suffer from some biases. One of them is caused by gender. Therefore, this paper aims to examine gender effects on peer rating in open innovation and crowdsourcing. More specifically, we examine how judge gender and gender similarity between judge and designer affect peer rating score. This question has never been examined in the OI&C literature. Using a quasi-experimental design, we collect 1,585 evaluations and find that, overall, judge gender has no significant effect on peer rating score, but gender similarity has a negative effect. Further examinations reveal that rating mode (single-blind or double-blind) may moderate such gender effects: male judges are predicted to give a higher rating score than females when the designer's information is disclosed while in double-blind peer rating gender similarity reduces the peer rating score. This study has practical implications to the use and design of a peer rating system in OI&C platforms.

Keywords: crowdsourcing, gender effect, open innovation, peer rating, selection system

1. INTRODUCTION

As a selection system, peer rating or evaluation has been adopted by many open innovation and crowdsourcing (OI&C) platforms ^[1, 2]. For example, in Threadless, a crowdsourcing community of artists, all designs are created and then chosen by its community. Each week, hundreds of designed T-shirts are produced by artists and then rated by peer artists. Several other OI&C platforms such as Jovoto (for brand innovation), Zooppa (for video commercials), PimTim (for graphic design), and Lego Ideas (for Lego design) have similarly adopted peer rating as their selection method to identify best submissions in their contests. Online peer rating represents a cheaper, more scalable way of evaluating submission and selecting winners, and therefore possibly it can be a powerful new way of evaluating submission quality in OI&C platforms ^[3, 4, 5]. In addition, peer rating empowers and engages users, which can expand the platform and maintain users in the OI&C community ^[6].

Despite of many advantages, peer rating also has its bias. Similar to scholarly peer review, peer rating in OI&C relies on the evaluation from one or multiple people of similar competence to the content producer. Lee et al. ^[7] identify different types of bias in peer evaluation process, including bias resulting from designer and evaluator characteristics. Among those characteristics, gender is a salient one. However, the results they have reviewed are mixed: some researchers find significant gender bias but others do not. Compared with scholars in peer review, participants in OI&C communities have an even bigger gap in terms of their knowledge and experience and thus non-quality factors such as gender may affect peer rating results. In sum, gender could be an influencing factor in online peer rating. In order to address this issue, we raise our research questions: do gender effects exist in peer rating in OI&C communities? If so, are gender effects moderated by the rating mode (single- or double-blind).

Gender effects in peer rating can be driven by two factors: judge gender (i.e., do male and female judges rate the same entry differently?) and the similarity between a judge's gender and a designer's (i.e., do judges with the same gender as the designer rate the entry differently from the one with different gender?). This study focuses on the later one while controls the first one and explores the research questions in the context of web design

because it is very common task in OI&C platforms such as 99.design, Taskcn, DesignCrowd, CrowdSpring and among many others. Because there are single-blind and double-blind peer rating, we consider and compare two rating modes: the designers' information including their gender is disclosed to judges (single-blind) and not disclosed (double blind). We conducted a quasi-experimental design and collected 1,585 evaluations of websites designed by college students, made by their fellow students. We find that, overall, judge gender has no significant effect on peer rating score, but gender similarity has a negative effect. Further examinations reveal that rating mode (single-blind or double-blind) may moderate such gender effects: male judges are predicted to give a higher rating score than females when the designer's information is disclosed while in double-blind peer rating, gender similarity reduces the peer rating score. Our findings suggest that gender effects must be considered when designing peer rating system in OI&C platforms.

This remainder of this paper is organized as follows. The next section provides a theoretical background by reviewing the relevant literature and discussing gender effects. Research method is described in Section 3 and results are presented in Section 4. The final section discusses practical implications, research limitations, and future research directions.

2. BACKGROUND

2.1 Peer rating as a selection system

Selection is the process of choosing candidates from a group of potentials based on some criteria. The selection system theory identifies three ideal types of selection systems based on the “selectors”: *market*, *peer*, and *expert* ^[1, 8]. In the market selection, consumers are selectors. In the peer selection, the selectors and the selected are part of the same group. In the expert selection, the selectors are neither producers nor consumers, but have the power to shape selection by virtual of their specialized knowledge and distinctive abilities. This literature adds that in practice a combination of selector types may occur, perhaps at different stages of the selection processes, but it is still useful to specify the dominant selectors.

As a form of peer rating, peer review has been introduced to the academia for many decades. It has provided a reliable form of scientific communication and ensured the quality of scientific research. Peer review is recognized as a required component of research validation, the academic reward system, and the scholarly publication process^[7]. Peer review could be single blind, double-blind, or open.

Recently, as the Internet provides a great way for open innovation, many OI&C platforms rely on peer rating for quality control, winner selection, and reward distribution. Peer rating is recognized as an important evaluation method in innovation contests^[2, 9]. Based on 33 articles and 57 real-world innovation contests, Bullinger & Moslein^[2] recognize the evaluation mechanism as one of the ten key design elements for innovation contests. According to them, innovation contests can use expert evaluation, peer rating, self-assessment, or mixed method. They also find that 31 out of 57 real-world innovation contests or platforms use community functionality such as commenting functions and forums and all of them comprise any form of peer rating. Despite of its great importance and relevance to open innovation, very few studies examine this issue in the context of OI&C.

2.2 Gender impact in peer rating

Gender plays an important role in human interaction. Previous studies have examined the effects of gender and gender similarity on various decisions and outcomes such as employment interviews and recruitment, job analysis, sales, and customer service, but their results are inconsistent and mixed ^[10, 11]. Those studies find positive, negative, or no gender effects. Therefore, those studies, we cannot infer how gender and gender similarity may affect peer rating result in OI&C.

Gender is a salient factor influencing peer rating results. Generally, the gender effect on peer rating consists

of two components: the judge gender and the similarity between a judge's gender and a designer's. Table 1 summarizes a few studies investigating the gender impact in peer rating under various research contexts. Both Girard & Pinar^[12] and Chen & Fang^[13] find that female reviewees receive a higher rating from peer reviewers. Pinar & Girard^[14] find that gender similarity increases peer rating results in one setting, but not in another setting.

Table 1. Research on gender effect in peer rating

Study	Main findings	Research context
[7]	This study identifies different types of bias in peer evaluation process. One of them is bias as a function of author characteristics such as nationality and gender.	Scholar publication
[12]	The gender of evaluators or presenters did not have any significant effect on presentation scores. Female students seem to be perceived as better presenters than male students.	Student presentation
[13]	Compared with male reviewers, female reviewers have a significantly higher number of high-quality reviews	Online reviews
[15]	Even though the estimates of the gender effect vary substantially across studies, men applicants have statistically significant greater odds of receiving grants than women by about 7%.	Grant application

2.3 Judge Gender

One driver of gender effect in peer rating is judge gender. The literature on marketing and psychology indicates that male and female process the same information in different ways. When judging a product, males were less oriented to visuals and more motivated extrinsically than females^[16]. Similarly, Wesley et al.^[17] posit that male consumers show significantly lower recreational consciousness and fashion consciousness than female consumers in their shopping activity. For example, Seock and Sauls^[18] find that male and female customers have different shopping orientation and use different criteria when evaluating apparel retail stores. Coontz^[19] finds significant differences between female and male judges when making their judicial decisions. However, Cooper et al.^[20] find that judge gender explain little of the variation in the ratings of the job evaluation.

2.4 Gender Similarity

The similarity between a judge's gender and a designer's may also affect peer rating results. However, there are two opposite arguments about how gender similarity impacts peer rating outcomes. On the one hand, the similarity-attraction theory posits that individuals tend to like and be attracted to others who are similar, rather than dissimilar, to themselves^[21]. Accordingly, gender similarity may increase peer ratings. In other words, a submission is expected to receive a higher rating score when the judge has the same gender with the designer.

On the other hand, gender dissimilarity is recognized as positive signal for social identification and will produce attraction between selectors and selectees^[22]. For example, Jones et al.^[23] find that customers are more likely to accept salespersons who are dissimilar to themselves. Accordingly, gender dissimilarity may yield a favorable peer rating score.

3. RESEARCH METHOD

In order to explore our research questions, we conducted a quasi-experiment design. Undergraduate students from multiple sections of a management information systems course at a major university were required to complete an individual project, which included two stages: designing a website and then evaluating websites designed by peer students. The instructor used the format provided by an open innovation platform to write a website design brief. At the first stage, students followed the brief to design a website on the same platform. At the second stage, students were randomly assigned to be a judge to evaluate 10-13 websites design by students

from a different section. The data was collected at the website evaluation stage. In total, 139 students participated in this project. Among them, 73 (52.5%) were male while the remaining were female. One website was dropped because of no accessibility. In total, 1585 evaluations were obtained from 138 websites and 139 judges.

Our dependent variable is overall peer rating score of each website. It ranges from 1 (very poor) to 5 (excellent). Among 1585 evaluations, 1454 include comments. There were 11,593 words in all these comments and, on average, each comment has 7.97 words. We use text mining techniques and generate a sentiment score based on judges' comments. In sentiment analysis, each comment receives a score describing it to be either positive or negative. This was conducted in RapidMiner 7.6. The sentiment score from each evaluation is considered as a supplementary dependent variable to check the reliability of our main dependent variable. Some sample comments and their sentiment score are shown in Table 2.

Table 2. Sample comments and their sentiment score

Comments	Sentiment Score
The photo gallery pics were not organized	-0.63
It's nearly the same with other websites	-0.36
This design was a little confusing for me	-0.33
The design of this website was quite boring	-0.25
I thought the website had some cool information, but the design of it was confusing, as I was not very sure what to look for.	-0.04
Very thorough website. It seems like you're involved in a lot. Maybe include some pictures to spice up the page a little bit.	0.12
Purpose isn't explicitly stated. Design is very well structured and easy on the eyes. Very thorough effort.	0.37
Good concept, purpose was a bit unclear	0.38
Loved the colors and design of this website!	0.50
This site was well put together and contained a lot of thorough information.	0.51
Author of website is enthusiastic!	0.63

Our independent variable is gender similarity (1: the judge and designer have the same gender; 0: the judge and designer have different genders). We control judge gender (1: male; 0: female), and website attributes, including purpose (how successfully this website serves a clear purpose such as introducing a person, a company, or a place or conducting business activities such as shopping or financial service), design (how successfully this website is well-designed such as well-organized content, appealing visuals, easy navigation), and originality (how successfully this website is distinguishable from other websites and gives you something that you cannot find elsewhere). These three attributes are usually considered as the top criteria to evaluate a website. Each of the three variables ranges from 1 (very unsuccessfully) to 5 (very successfully).

In addition, designers were allowed to voluntarily choose whether or not disclose their gender information such as biography and pictures on their websites. Doing so, we created an additional factor, rating mode, to indicate whether the evaluation is single-blind (i.e., the judge can view the designer's gender while the designer does not the judge's gender) or double-blind. We will test whether rating mode moderate gender effects on peer rating.

4. RESULTS

The mean and standard deviation of each variable and correlation coefficients among all the variables are

presented in Table 3. Each evaluation represents a designer-judge pair. About a half of 1585 evaluations are evaluated by male judges and 52% of designer-judge pairs have the same gender. Peer rating score has a significantly positive relationship with Sentiment score, indicating a good reliability of our main dependent variable.

We first run a linear regression model with the peer rating score as the dependent variable for all the sample. We find that gender similarity has a significant but marginal negative influence on peer rating score. We do not find a significant relationship between judge gender and peer rating score. In addition, the positive and significant regression coefficients of all the three controlled variables indicate that the three website attributes purpose, design, and originality play an important role in determining peer rating score.

Table 3. Descriptive statistics of seven variables

Variable	Mean	Std. Dev.	1	2	3	4	5	6	7
1. Judge gender	0.52	0.500	1.00						
2. Gender similarity	0.52	0.500	0.06	1.00					
3. Purpose	4.64	0.745	-0.11	-0.02	1.00				
4. Design	4.29	0.937	-0.09	-0.03	0.59	1.00			
5. Originality	4.43	0.848	-0.15	-0.05	0.52	0.58	1.00		
6. Peer rating score	4.47	0.809	-0.11	-0.06	0.75	0.83	0.74	1.00	
7. Sentiment score	0.20	0.222	0.00	0.00	0.27	0.31	0.23	0.34	1.00

Note: Sample size is 1585 for the first six variables and 1454 for the last variable, Sentiment score.

Table 4. Linear regression model with overall rating score as the dependent variable

Independent Variable	All sample	Single-blind	Double-blind
	Coefficient. (Std. Error)	Coefficient. (Std. Error)	Coefficient. (Std. Error)
Gender Similarity	- 0.050 (0.0158) **	- 0.042 (0.0333)	- 0.040 (0.0186) *
Controlled variables			
Judge Gender	0.016 (0.0160)	0.067 (0.0335) *	0.008 (0.0188)
Purpose	0.348 (0.014) ***	0.391 (0.025) ***	0.330 (0.017) ***
Design	0.397 (0.011) ***	0.358 (0.024) ***	0.409 (0.013) ***
Originality	0.289 (0.012) ***	0.309 (0.023) ***	0.287 (0.014) ***
Sample Size	1585	399	1150
Adjusted R Square	0.848	0.860	0.846

* p<0.05, ** p<0.01, *** p<0.001.

In order to check whether rating mode (single-blind or double-blind) moderate the effects of judge gender and gender similarity on peer rating score, we divide the whole sample into two groups and run the same model individually for each group. Interestingly, gender effects are moderated by rating mode. Specifically, when the designer information is not disclosed (double-blind rating), gender similarity still has a negative effect on peer rating score, which is consistent with the results derived from the whole sample. A potential reason is that a judge may get some cues from the web design elements such as the use of color or pictures to determine the designer's gender. This explanation is found from the research on scholarly reviews, which reveals that even though author anonymity can prevent reviewer bias in the double-blind review, reviewers can often identify the

author through their writing style, subject matter, or self-citation ^[7]. In our case, judges may identify the designer's gender based on their wording, colors, pictures, interests, or hobbies on their website even though they may not identify the designer.

When the designer information is disclosed to judges, the effect of gender similarity is insignificant, but the judge gender has a significant and positive effect on peer rating score, indicating that male judges give a higher rating to the designer no matter whether they have the same gender or not. Under the double-blind rating mode, the design of a website makes the largest contribution to the overall peer rating score, while under the single-blind rating mode, the purpose of a website is the greatest contributor. These comparisons suggest the existence of a potential moderation effect of rating mode (single-blind or double-blind), which can be further examined in future research.

When the values of controlled variables are fixed, Figure 1 presents the difference of designers' rating scores under two rating modes. Interestingly, no matter whether the designer's gender information is disclosed or not, female designers always receive the highest rating score from male judges, but the lowest rating from female judges. No matter whether the designer's gender information is disclosed or not, female designers always receive a higher rating from male judges than that from female judges. When the designer's information is disclosed, male designers receive a higher rating score from male judges than that from female judges. However, when the designer's gender information is not disclosed, male designers receive a higher rating from female judges than from the male judges. This finding indicates that the gender effect on male designers' rating is moderated by rating mode.

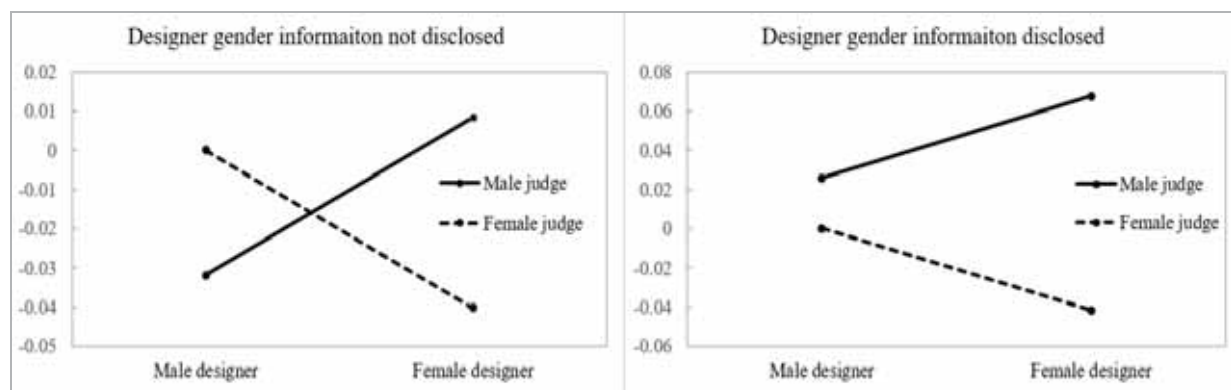


Figure 1. The difference of designers' rating scores under two rating modes

5. CONCLUSIONS AND DISCUSSIONS

This paper aims to investigate gender effects on peer rating in OI&C: whether judge's gender and the gender similarity between judge and designer may influence peer rating in the case of website design. Overall, we find that judge gender does not affect the peer rating score, but gender similarity decreases the peer rating score. Our case supports the negative impact of gender similarity, which may indicate that dissimilarity can attract higher peer ratings in OI&C. In addition, we find that the effects of judge gender and gender similarity on peer rating score seem to be moderated by rating mode (single-blind or double-blind): in single-blind peer rating, male judges are likely to give a higher rating score than females while in double-blind peer rating, gender similarity reduces the peer rating score.

This study has a few practical implications. Because gender effects do exist in the peer rating, no matter whether the rating is single-blind or double-blind, OI&C platforms should find a good way to avoid or minimize the gender effect. For example, they can keep a balance of male and female judges in each open innovation project. Compared with the overall rating score (main dependent variable), sentiment score (supplementary

dependent variable) is less sensitive to gender effect. Therefore, when evaluating an OI&C project, both objective rating based on evaluation rubric and subjective comments should be considered.

This study further confirms the significant and dominant role of website's purpose, design, and originality in website rating. A well-designed and creative website with a clear purpose always deserve a higher rating from peer judges. However, when all these factors are the same, gender effects cannot be ignored because even a marginal increment in peer rating can change the ranking of competitors in an OI&C contest.

This study also has a few limitations. First of all, we did not control the disclosure decision of designer's information in their submissions. Future experiment can add this treatment to systematically examine the effect of rating mode. Secondly, in our quasi-experiment, judges were randomly assigned to evaluate all 11-13 websites systematically, but in reality judges are voluntary to evaluate particular submissions in a non-systematic manner. Thirdly, most participants in this study are naïve, rather than professional designers, so it is possible that they disclosed their information unintentionally. All the limitations can be addressed in the future by designing a more strict experiment with rating mode as a factor, including other types of tasks such as logo design, or incorporating the real rating behavior data from an open innovation platform.

REFERENCES

- [1] Gemser, G., Leenders, M.A., and Wijnberg, N.M. (2008). Why Some Awards Are More Effective Signals of Quality Than Others: A Study of Movie Awards†. *Journal of Management*, 34(1): 25-54.
- [2] Bullinger, A.C., and Moeslein, K. (2011). Innovation Contests: Systematization of the Field and Future Research. *International Journal of Virtual Communities and Social Networking*, 3(1): 1-12.
- [3] Howe, J. (2008). *Crowdsourcing: How the power of the crowd is driving the future of business*. Random House.
- [4] Brabham, D.C. (2009). Crowdsourcing the public participation process for planning projects. *Planning Theory*, 8(3): 242-262.
- [5] Majchrzak, A., and Malhotra, A. (2013). Towards an information systems perspective and research agenda on crowdsourcing for innovation. *Journal of Strategic Information Systems*, 22(4): 257-268.
- [6] Semmann, M., and Grotherr, C. (2017). How to Empower Users for Co-Creation-Conceptualizing an Engagement Platform for Benefits Realization. *The 13th International Conference on Wirtschaftsinformatik* St.Gallen, Switzerland: AIS.
- [7] Lee, C.J., Sugimoto, C.R., Zhang, G., and Cronin, B. (2013). Bias in peer review. *Journal of the Association for Information Science and Technology*, 64(1): 2-17.
- [8] Wijnberg, N.M. (2011). Classification systems and selection systems: The risks of radical innovation and category spanning. *Scandinavian Journal of Management*, 27(3): 297-306.
- [9] Adamczyk, S., Bullinger, A.C., and Möslin, K.M. (2012). Innovation Contests: A Review, Classification and Outlook. *Creativity and Innovation Management*, 21(4): 335-360.
- [10] Graves, L.M., and Powell, G.N. (1995). The Effect of Sex Similarity on Recruiters' evaluations Of Actual Applicants: A Test of the Similarity - Attraction Paradigm. *Personnel Psychology*, 48(1): 85-98.
- [11] Hardin, J.R., Reding, K.F., and Stocks, M.H. (2002). The effect of gender on the recruitment of entry-level accountants. *Journal of Managerial Issues*: 251-266.
- [12] Girard, T., and Pinar, M. (2009). An exploratory study of gender effect on student presentation evaluations: Does gender similarity make a difference? *International Journal of Educational Management*, 23(3): 237-251.
- [13] Chen, L., and Fang, J. (2017). Users' Attributes Associated with High-Quality Review in Online Communities. *WHICEB 2017 Proceedings*, Wuhan, China: AIS.
- [14] Pinar, M., and Girard, T. (2006). Student Perceptions of Class Presentations: Does Gender Impact The Evaluations? , *Marketing Management Association 2006 Educators' Conference Proceedings*, pp. 28.

- [15] Bornmann, L., Mutz, R., and Daniel, H.-D. (2007). Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, 1(3): 226-238.
- [16] Holbrook, M.B. (1986). Aims, concepts, and methods for the representation of individual differences in esthetic responses to design features. *Journal of consumer research*, 13(3): 337-347.
- [17] Wesley, S., LeHew, M., and Woodside, A.G. (2006). Consumer decision-making styles and mall shopping behavior: Building theory using exploratory data analysis and the comparative method. *Journal of Business Research*, 59(5): 535-548.
- [18] Seock, Y.-K., and Sauls, N. (2008). Hispanic consumers' shopping orientation and apparel retail store evaluation criteria: An analysis of age and gender differences. *Journal of Fashion Marketing and Management: An International Journal*, 12(4): 469-486.
- [19] Coontz, P. (2000). Gender and judicial decisions: Do female judges decide cases differently than male judges? *Gender issues*, 18(4): 59-73.
- [20] Cooper, E.A., Doverspike, D., Barrett, G.V., and Alexander, R.A. (1987). Sex bias in job evaluation: The effect of sex on judgments of factor and level weights. *Educational and psychological measurement*, 47(2): 369-375.
- [21] Byrne, D.E. (1971). *The Attraction Paradigm*. New York: Academic Press.
- [22] Schaffer, B. (2016). Dissimilarity-Attraction in Teams: New Ideas for Workplace Diversity Research. *Academy of Management Proceedings: Academy of Management*, pp. 12344.
- [23] Jones, E., Moore, J.N., Stanaland, A.J., and Wyatt, R.A. (1998). Salesperson race and gender and the access and legitimacy paradigm: Does difference make a difference? *Journal of Personal Selling & Sales Management*, 18(4): 71-88.

A Text Mining Based Approach for Mining Customer Attribute Data on Undefined Quality Problem

Qing Zhu^{1}, Yiqiong Wu¹, Yuze Li², Renxian Zuo¹*

¹International Business School, Shaanxi Normal University, Xi'an, 710119, China

²Department of Mechanical and Industrial Engineering, University of Toronto, ON, Canada

Abstract: Understanding how the consumer perceives quality is a key issue in supply chain management. However, as the market structure continues to deepen, traditional evaluation methods using SEVRQUAL are unable identify all issues related to customer quality and unable to supply solutions. The maturation of data mining technology, however, has opened the possibilities of mining customer attribute data on quality problems from unstructured data. Based on the consumer perspective, this research uses an unsupervised machine learning text mining approach and the Recursive Neural Tensor Network to resolve the attribution process for undefined quality problems. It was found that the consumer quality perception system has a typical line-of-sight that can assist consumers quickly capture the logical structure of the quality problem. Although attributions related to quality problems are very scattered, a highly unified view was found to exist within each group, and a strategy to solve the undefined quality problem was agreed through group consensus by 61% of the consumers.

Keywords: text mining, supply chain management, quality control

1. INTRODUCTION

Jacoby, Olson and Haddock ^[1] identified consumer perceived quality as an important criterion for consumer feedback on the quality of goods. Based on Jacoby, Olson and Haddock's initial definition, Grönroos ^[2] extended the interpretation, as follows: when the quality of the goods are on the same level, due to impact factors such as previous experience and commodity characteristics, consumers develop different perceptions about the quality of the goods. As consumer perceptions of quality are affected by mood and trust, this could affect consumer satisfaction and loyalty ^{[3],[4]}. Lee and Lin ^[5] then extended this concept to marketing strategy, proposing that as consumer satisfaction influences consumer purchase intention, controlling the consumer's perception of good quality is an important part of enterprise marketing strategies ^[6].

At the same time, Folkes and Kotsos ^[7] claimed that if managers wanted to increase transaction success, they needed to minimize the differences between the consumer and the seller, and therefore supply chain managers needed to be able to understand product quality from the consumer's perspective ^[8] and be able to analyze consumer thinking to illuminate the shortcomings in their products and services, reduce potential company losses, and enhance profit margins ^[9], all of which would result in more effective, targeted supply chain management.

In support of Jacoby's idea of consumer perceived quality, Parasuraman, Zeithaml, and Berry sought to develop an appropriate consumer attribution management tool and proposed the SERVQUAL evaluation method ^[10], with the understanding that service quality depended on the gap between the consumer's expectations and the perceived quality. They divided service quality into 10 dimensions: tangibles, reliability, responsiveness, reliability, communication, politeness, security, understanding, and accessibility. In subsequent studies, service quality was integrated into 5 dimensions (22 items); materiality, reliability, responsiveness, assurance and empathy, after which this evaluation method was widely used in many fields to measure perceived consumer quality. For example, Marek and Nowacki ^[11] used these methods to evaluate the tourism quality at the Rogalin

* Corresponding author. Email: zhuqing@snnu.edu.cn

Museum, from which they were able to identify its competitive advantages and disadvantages and offer comprehensive guidance for managing perceived customer quality.

While it has been widely used, the reliability of the SERVQUAL method has been questioned due to its subjectivity, robustness, and variability. A series of revised models on perceived quality evaluations were then developed. Cronin and Taylor^[12], for instance, developed the SERVPERF method, in which consumer service expectations and consumer service perceptions were two separate measures. Peter, Churchill, and Brown^[13] proposed a non-difference valuation method as they felt that the SERVQUAL method did not account for previous service experience when measuring consumer service expectations, thus weakening the validity of the difference evaluation method.

While the validity of traditional quality perception measurements has been proven and part of the SERVQUAL application can pass the Kaiser-Meyer-Olkin (KMO) and Bartlett tests of sphericity, mainstream measurement methods based on SERVQUAL all tend to formulate project scores from the supply chain management's perspective and fail to directly address consumer concerns or measure consumer expectations. Therefore, the information measured using traditional quality measurement methods is limited both in perspective and effectiveness.

As supply chain managers understand the complete supply chain operation, they consider all quality control links in the supply chain^{[14],[15]}. However, the information asymmetry between consumers and management^[16] and the lack of consumer information about the nature of goods during production and sales have resulted in significant differences between consumers and management in terms of the causes of the quality problem attributions^[17]. Further, the SERVQUAL tool designs amplify such differences. When quality problems occur, consumers usually attribute the problem to the perceivable end of the supply chain, and do not perceive the overall supply chain. The main cause for the inability of traditional methods such as SERVQUAL to adapt is a difference in perspective: that is, understanding the quality perception of the consumer from the management's perspective can only amplify the differences, which further highlights the invalidity of the tool.

Since 2012, non-structured big data processing technology such as text mining has begun to mature and be applied in areas such as quantitative strategy, market segmentation, prediction and group behavior intervention, and other fields. Therefore, there are new methods now available for measuring the consumer quality perspective. Based on Folkes and Kotsos, this paper uses an unsupervised machine learning algorithm and a recursive neural network method to identify consumer quality perceptions from unstructured data, analyze consumer responses about quality problems, and describe consumer expectations and behavior, while confirming that mainstream measurement methods such as SERVQUAL are not optimal.

2. METHODOLOGY

2.1 Text mining

Text mining is a special form of data mining. It can discover and extract implicit valuable information from massive amount of unstructured data, and form knowledge that is easy for users to understand. The implementation of text mining is divided into two steps: text preprocessing and knowledge extraction. Text preprocessing transforms unstructured text into term-document matrix. Knowledge extraction derives facts and knowledge from term-document matrix. Based on different purposes, the task of text mining can be broadly divided into five categories: text classification, text clustering, association rule mining, automatic summarization and topic detection.

2.2 Text preprocessing

In this study, a series of cleaning and feature representation of the text data are carried out by means of NLP and TM packages in R.

2.2.1 Characteristics representation

This study used a Vector Space Model (VSM) to represent the text^[18], the fundamental principle for which was to assign different weights to each word, thereby allowing the characteristics vector to be represented as a weighted text, as follows: for text set, a particular text $T = \{t_i\}_{i=1}^n$, a particular text $t_i (i \leq n)$ can be represented as $t_i = w_{i1}, w_{i2}, \dots, w_{im}$. In which, m is the number of characteristics, w_{ij} is the weight of j th characteristic in text t_i .

A classic weight assignment method when constructing a vector space model is the TD-IDF method developed by Salton and Buckley^[19], which is able to calculate the importance of a particular word in a text and therefore has more accurate representation and clustering results. The formula for the TD-IDF weight assignment method is as follows:

$$W_j = TF_j * IDF_j \quad (1)$$

in which,

$$IDF_j = \log \left(\frac{n}{DF_j} \right) \quad (2)$$

where W_j is the weight of the j th characteristic, TF_j is the frequency of the j th characteristic in the current text, and DF_j is the frequency of the j th characteristic in the overall text set. In a real application, to avoid too broad a variable value range, this study normalized the vector so that the average was 0 and the square difference was 1.

2.2.2 Similarity analysis between texts

Before conducting text clustering analysis, we need to measure the degree of similarity and difference between texts. This document uses hierarchical, k-means and spectral clustering for text clustering.

When conducting text hierarchical clustering, the distance between clusters is optimized by the Ward method. The Ward method requires the degree of separation to be calculated by the Euclidean distance. The Euclidean distance between i th text and j th text is calculated using the following formula:

$$d(i, j) = \sqrt{\sum_{k=1}^m (w_{ik} - w_{jk})^2} \quad (3)$$

Subject to,

$$d(i, i) = 0$$

$$d(i, j) = d(j, i)$$

Dhillon and Modha^[20] claimed that the cosine distance was superior to the Euclidean distance for measuring text clustering similarity. Therefore, k-means clustering, spectral clustering, and cosine similarity were used to assess document similarities and obtain the document similarity matrix. The cosine similarity between the i th and the j th documents was determined using the following formula:

$$\text{sim}(d_i, d_j) = \frac{d_i \cdot d_j}{|d_i| \times |d_j|} = \frac{\sum_{k=1}^m w_{ik} \cdot w_{jk}}{(\sum_{k=1}^m w_{ik}^2)(\sum_{k=1}^m w_{jk}^2)} \quad (4)$$

Subject to,

$$\text{sim}(d_i, d_j) = 1$$

$$\text{sim}(d_i, d_j) = \text{sim}(d_j, d_i)$$

2.3 Text clustering

Text clustering is unsupervised learning that involves aggregating massive text data into several classes without prior knowledge or assumptions, thereby ensuring as high a similarity of text data as possible and as

low a similarity as possible across the classes. As there is no precise definition for clustering, the clustering algorithm varies with the results.

2.3.1 Hierarchical cluster

Hierarchical clustering has been a common clustering method. In clustering analysis, the basic principle is to select the two classes with the highest similarity aggregation of all the classes. This step is repeated until all data are grouped into a class. Compared with other clustering methods, hierarchical clustering can be applied to arbitrary shapes and attribute data set types^[21]; however, the time complexity of the algorithm is relatively high and therefore not suitable for clustering massive amounts of text data^[20]. From the notion of basic hierarchical clustering, Rohlf proposed an MST-algorithm based on the minimum spanning tree that was able to optimize hierarchical clustering performances^[22]. In this paper, an optimized hierarchical clustering algorithm was adopted, the algorithm for which was as follows:

1. With the known text set $T = \{t_i\}_{i=1}^n$ and the differences between documents, d ;
2. Let every text be a cluster, and then initialize the output table such that $out \leftarrow []$;
3. For any $x \in T$, $size(x) \leftarrow 1$;
4. Suppose $\arg \min (d(t_x, t_y))$ is (t_a, t_b) , then combine t_a, t_b into a new cluster $(t_a \cup t_b)$:

$$out \leftarrow out + (t_a, t_b, (t_a \cup t_b)) \quad (5)$$

$$T \leftarrow C_T t_a \cup C_T t_b \cup (t_a \cup t_b) \quad (6)$$

5. Use the Ward method^[23] to update the inter-cluster distance:

$$d(t_a \cup t_b, t_k) = \sqrt{\frac{(n_{t_a} + n_{t_k})d(t_a, t_k) + (n_{t_b} + n_{t_k})d(t_b, t_k) - n_{t_k}d(t_a, t_b)}{n_{t_a} + n_{t_b} + n_{t_k}}} \quad (7)$$

6. $size(t_a \cup t_b) \leftarrow size(t_a) + size(t_b)$
7. Repeat steps 4 to 6 until a cluster of size n is obtained

2.3.2 K-means clustering

K-means clustering is a common clustering method based on centroids^[21] that has a lower time complexity and a higher computational efficiency; however, the algorithm is not suitable for non-convex data, does not have robustness, is more sensitive to outliers, and can easily fall into a local optima. Therefore, the clustering results are more susceptible to the influence of the number of predefined clusters^[21]. Pelleg and Moore^[24] proposed an X-Means algorithm that could automatically determine the number of K clusters using optimization. Therefore, this study used the split level algorithm, the steps for which were as follows:

1. With a known text set $T = \{t_i\}_{i=1}^n$;
2. Initialize the number of K clusters;
3. Randomly select the clustering centroid $C = \{c_k\}_{k=1}^K$;
4. Cluster the text objects into the nearest-located cluster and obtain K classes $\{W_k\}_{k=1}^K$, as defined in the equation (8):

$$W_k = \{t_i \in T | k = \arg \min_{j=1, \dots, K} \|t_i - c_j\|\} \quad (8)$$

Subject to, $W_k \subset T, W_l \cap W_q = \emptyset, \cup_{k=1}^K W_k = T$;

5. Use equation (9) to update each centroid

$$c_k = \frac{1}{|W_k|} \sum_{t_i \in W_k} t_i \quad (9)$$

6. Repeat step 4 and 5 until a stable clustering result is obtained.

The algorithm optimizes the clustering result using the iterative method in the following equation (10) to minimize the sum of square errors $E_T(C)$:

$$E_T(C) = \sum_{k=1}^K \sum_{t_i \in W_k} \|t_i - c_k\|^2 \quad (10)$$

2.3.3 Spectral clustering

The essence of the spectral clustering algorithm is to use the eigenvector of the Laplace matrix. The relationships between the texts are used to build graph $G = (V, E, W)$ with n nodes, with the vertex $V = \{1, \dots, n\}$ representing each text. The edge $E \subseteq V \times V$ in the graph illustrates the relationships between texts, and the weight of the edges $W = (w_{ij})_{n \times n}$ shows the strength of the relationships between the texts. The goal of spectral clustering is to divide the graph model into a number of subgraphs and minimize segmentation losses [25]. The spectral clustering algorithm has the ability to converge the clustering results to a global optimum and is not sensitive to outliers. However, the spectral clustering time complexity and the number of clusters k needs to define in advance are high [21]. The algorithm is as follows:

1. Obtain a similarity matrix F_{ij} between the texts;
2. The Laplace matrix $L = D^{-1/2}FD^{-1/2}$ is constructed, in which D is the diagonal matrix of the diagonal elements $D_{ii} = \sum_{j=1}^n F_{ij}$;
3. The eigenvectors s_1, s_2, \dots, s_k that correspond to the minimum eigenvalues of the first k of L are calculated, and the matrix $S = [s_1, s_2, \dots, s_k] \in R^{n \times k}$ is obtained;
4. Consider each line in the S as a point in space R^k , and use the k-means clustering algorithm to obtain k text clusters.

2.4 Sentiment analysis - recursive neural tensor network

As grammar rules are recursive, Socher, Perelygin, and Wu [26] combined them with a corresponding algorithm to fully analyze a text. Then, based on the existing recursive neural network (RNN) and matrix-vector recursive neural network (MV-RNN) models, they proposed a recursive neural tensor network (RNTN) for fine grained sentiment text classification. The fine grained sentiment classifications had 5 emotional levels; very negative, negative, neutral, positive, and very positive. For fine grained sentiment analysis, the algorithm increased the accuracy from 44.4% to 45.7%.

2.4.1 Neural network calculation process

In recursive neural models, the compositional vector representations for phrases of different length and syntactic type can be computed in a bottom up recursive fashion using different compositionality functions.

Based on the current RNN model, Socher, Perelygin, and Wu proposed a new model called the RNTN, which was able to compute a sentence tree with detailed emotional information using the recursive combination between words and phrases

Figure 1 gives an example of a three tensor layer neural network. The computation process for a single tensor layer is as follows:

1. Each word in the sentence is represented as a d -dimensional vector. All word vectors are initialized by random sampling each value from a uniform distribution $U(-0.0001, 0.0001)$;
2. The output of a tensor product $h \in R^d$ is defined as:

$$h = \begin{bmatrix} b \\ c \end{bmatrix}^T V^{[1:d]} \begin{bmatrix} b \\ c \end{bmatrix} \quad (11)$$

where $V^{[1:d]}$ is the tensor that defines the multiple bilinear forms;

3. The first parent vector $p1$ is computed:

$$p1 = f \left(\begin{bmatrix} b \\ c \end{bmatrix}^T V^{[1:d]} \begin{bmatrix} b \\ c \end{bmatrix} + W \begin{bmatrix} b \\ c \end{bmatrix} \right) \quad (12)$$

where W is the sentiment classification matrix;

4. The next parent vector $p2$ in the tri-gram is computed

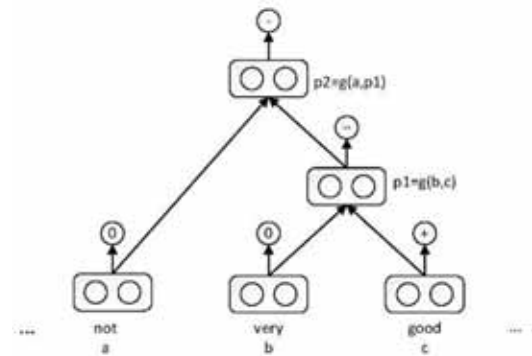


Figure 1. An example of a recursive neural tensor network

using the same weights:

$$p2 = f \left(\begin{bmatrix} a \\ p1 \end{bmatrix}^T V^{[1:d]} \begin{bmatrix} a \\ p1 \end{bmatrix} + W \begin{bmatrix} a \\ p1 \end{bmatrix} \right) \quad (13)$$

5. Steps 3 and 4 are repeated and each parent vector is computed in a bottom up fashion until the top parent vector is reached and the final sentiment orientation determined.

2.4.2 Model training

When the syntax tree is generated, the model trains a softmax classifier from top to bottom through the vector labels of each node. This semantic distribution relationship can be expressed as:

$$y^a = \text{softmax}(W_s a) \quad (14)$$

in which, $W_s \in R^{5 \times d}$ is the emotional classification matrix, and a is the operation node for the classification.

2.5 Price decomposition model

From the sentiment classification prediction model of the text, the sentimental trends in each sentence are integrated into the sentiment fluctuation in which t represents the number of sentences in the current comment; however, the t values of each text are not equal. To identify similar trends in the sentiment fluctuations, a price decomposition model was applied to divide each comment into positive sentiment fluctuations E_{inc} and negative sentiment fluctuations E_{dec} . The price decomposition model, which was first proposed by Oscar in 1972^[27], decomposes price into a rise and a fall and allows for the asymmetric effect of demand to be studied. For a comment containing t sentences, the specific decomposition formulas are as follows:

$$\begin{cases} E_{inc} = \sum_{i=1}^{t-1} \max \{0, (E_{i+1} - E_i)\} \\ E_{dec} = \sum_{i=1}^{t-1} \min \{0, (E_{i+1} - E_i)\} \end{cases} \quad (15)$$

3. DATA COLLECTION

As the object of this study was the general end consumer, it does not include the “industrial market”, “raw material market” or “intermediate manufactured goods market” as these could result in consumer ambiguity. Further, to ensure scientific questionnaire validity, a real online shopping situation was simulated that applied real evaluation rules. Study objects were required to provide comments on three evaluation categories; the quality of the goods, the logistics service, and consumer service attitudes (total simulation of an Alibaba shopping scenario) and used Likert scales ranging from -2 to 2 to represent their level of satisfaction with -2 being very unsatisfactory and 2 being very satisfactory.

Hoffman^[28] claimed that education, income, gender, occupation, and other factors affected online consumer shopping behavior. However, with the increased popularity of network technology, demographic characteristics are expected to gradually decline. For example, Zellner^[29] found that gender, income, and education levels did not contribute to online shopping differences, and Doolin, Dillon, and Thompson^[30] also found no significant correlations between a consumer’s age and online shopping behavior. Therefore, as it has been repeatedly shown that demographic characteristics were less related to consumer online shopping behavior, it is reasonable to surmise that the experimental results were not biased.

This study conducted a questionnaire survey posted electronic questionnaire online. A total of 788 questionnaires were collected and 508 valid samples obtained. After data collection was completed, the score items and text comments were separated and stored, and were then read separately into the software for analysis.

4. DATA ANALYSIS

The data analysis was divided as follows: (1) an analysis of the perceived quality measurements based on the SDERQUAL structure and the scores for the quality of the goods, the logistics service, and the customer

service attitude; (2) a description of the consumer perception system and the problem detection expressions; (3) presentation and measurement of consumer attributions; (4) analyses of consumer strategies and expected actions; and (5) an analysis of the relevant consumer logic and emotional consumer attributions.

4.1 Score item analysis

In this study, the three dimensions expressed using the Likert scale were turned inside out, so that the degree of dissatisfaction and traditional attribution expressions could be displayed, as shown in Figure 2, with the distance between the midpoint and the center indicating the degree of satisfaction; that is, the shorter the distance, the higher the product satisfaction. In this paper, the three scores (the quality of the goods, the logistics service, and consumer service attitudes) from the 508 individual subjects were connected in a closed triangle, in which the degree of overlap is shown in the color transparency.

Table 1 shows the correlation coefficients for the score items, with all relevant significant relationships showing horizontal dominance. The correlation coefficient between quality of goods and logistics was shown to be negative, indicating that most participants perceived the poor quality to be attributable to either the goods or the logistics. Therefore, as a generally negative evaluation was given for the goods and logistics services under this endogenous evaluation system, the accuracy of the attributions could be questioned. However, in real management activities, as the usual practice is to review the system from management to the operating level in response to negative feedback, management cannot generally respond quickly.

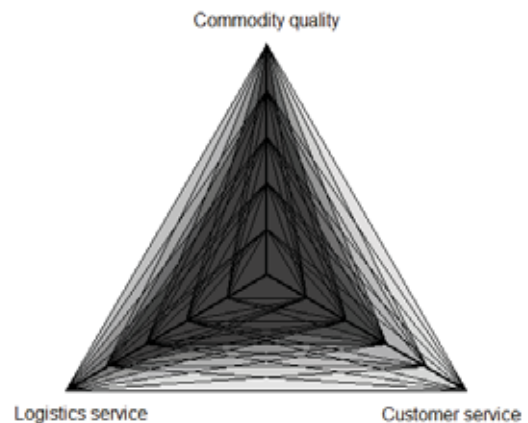


Figure 2. Score item analysis

Table 1. Correlation coefficients between score items

Correlation coefficients	Commodity quality	Logistics service	Customer service
Commodity quality	1	-0.198**	0.132**
Logistics service	-0.198**	1	0.329**
Customer service	0.132**	0.329**	1

Note: **: significantly correlation on 0.01 level.

4.2 Text mining

4.2.1 Cause attribution measurements

After separately preprocessing the text comments, the word frequency statistics were counted. Excluding entries that had less than 15 words, the resulting word cloud is shown in Figure 3.

Through an effective combining of the words and word clouds, a series of elements surrounding the quality problems; “broken hole”, “customer service”, “logistics”, “quality” and “seller”; were clearly exposed. In contrast to traditional methods for measuring perceived quality (Figure 2), managers are able to quickly locate the immediate causes for quality problems, and take measures to prevent the problem from further deteriorating. This direct effect was demonstrated in a preliminary analysis of the output structure of the consumer perception system, from which it was observed that to indirectly express the consumer cause attribution component, the system could quickly and accurately capture the main quality issues without the assistance of a non-difference

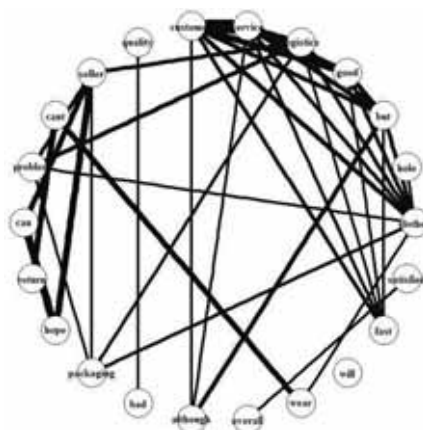
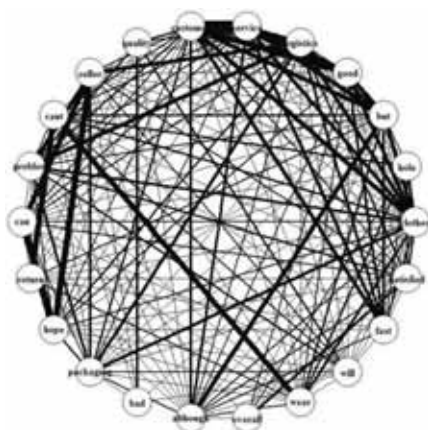


Figure 3. Word cloud.

4.2.2 Word association analysis

To maintain a maximum level of information entropy in the consumer perception system, based on the cause attribution estimates, a consumer perceived quality logic and uses *findAssocs()* function was constructed in the tm Package in R to determine the associations between 22 entries.

Based on characteristics such as the word set sparsity and information redundancy, word entries with frequencies greater than 40 were selected, and a relationship graph drawn that showed the combined logical relationships between the high frequency words (Figure 4 (a)). To highlight the entry logics that had strong associations, the associations less than 0.2 were removed between the 22 entries, resulting in Figure 4 (b).



(a) The logical structure.

(b) The simplified logical structure.

Figure 4. Word association analysis.

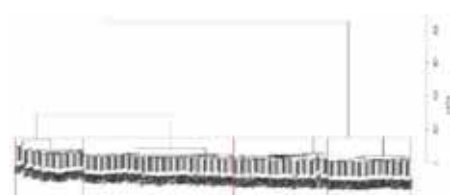
Figures 4 (a) and (b) appear to indicate that consumer perceived quality cannot be measured using SERVQUAL, its algorithms, or using abstract management theories. The basic cause attribution principle is to construct a unique consumer preference structure by emphasizing certain aspects in the logical structure shown in Figure 4. In addition, as the cause and results are consistent, this means that the cause is also the result. Nikhashemi and Tarodfer^[31] found a high degree of similarity between consumer preferences and consumer perceived quality and predicted that the high latitude and endogeneity in the consumer cause attribution structure could be responsible. In this paper, the ordinary least square (OLS) estimation result was deemed unacceptable, which also indirectly reflected the two elements integrated structure. However, as language structure is highly logical and self-consistent, the mixing of causal factors may be normal.

Using Figure 4 and the grammar rules, the associated entries were combined into the phrases that the consumers paid more attention to: “hole”, “can’t wear”, “can’t return”, “poor quality” and “seller does not return”. The relevance was further differentiated based on Figure 4 (b), and the logically isolated components

removed so as to obtain a two logic structure that had certain associations; “nice clothes”, “but”, “clothing hole”, “can’t wear” and “seller does not return” and “no return”. As “poor quality” was an isolated structure, it was not seen as affecting the consumer tendencies towards cause attribution. Compared with the results of the measurement analysis in Section 2.1, consumers do not pay much attention to summarizing and criticizing their reasons for poor quality and are also unable to deduce strong attribution factors from “poor quality”. However, phrases such as “the seller does not return”, “no return” or “demand return” are at the center of the review comment logic, which revealed that consumers tend to have a certain strategy when assessing quality perception. Traditional perception methods have failed to identify consumer strategies and expected behaviors, and generally, consumer enthusiasm for quality management activities has also been misunderstood.

4.2.3 Text clustering

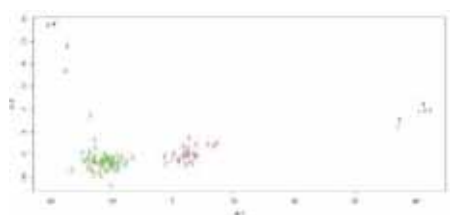
To determine which solution was satisfactory for most consumers, a text clustering method was used to classify consumer comments. As there were individual deviations in the clustering results because of the randomness of the clustering algorithm, hierarchical clustering, k-means clustering and spectral clustering were combined to derive a general proportion for the total number of consumer comments.



(a) Hierarchical clustering



(b) K-means clustering



(c) Spectral clustering

Figure 5. Visual clustering results

Figure 5 shows the results of these three clustering methods. Figure 5 (a) is the visual result for the hierarchical text clustering, from which it can be seen that the consumer reviews were clustered into four distinct categories. Table 2 displays the typical comments in each category; the 1st type simply objectively describes the quality problem and suggests a desired solution (return and

exchange); the 2nd type believes the seller is responsible for the quality problem; the 3rd type believes the logistics are the cause of the problem; and the 4th type points out the quality problem directly and demands a return.

As k-means clustering and spectral clustering require the number of clusters to be customized, the number of clusters in the hierarchical clustering were referred to and initialized as 4 to facilitate comparisons across the categories. Figure 5 (b) shows the visual results for the k-means clustering and Table 3 gives a representative evaluation of the 4 categories. The 1st consumer type attributed the commodity problem to the common responsibility of both the seller and the logistics; the 2nd consumer type simply pointed out the commodity quality problems; the 3rd consumer type simply described the problem and demanded a return; and the 4th consumer type described the problem and also gave positive evaluations for the logistics and customer service.

Table 2. Typical comments from each category (hierarchical clustering)

Category	Comment
1	There are holes in the packaging and clothing. Whether it is a seller problem or a logistics problem, the customer should be able to return the item. Because I paid for the clothes, I should receive it in good condition. I am satisfied with the clothes except for the hole. The clothes feel comfortable when I wear it and it is the size recommended by customer service. I hope the seller can negotiate with the logistics and make me satisfied. I will also accept a replacement if it can't be returned. I

	still trust the clothing quality.
2	I hope it won't happen again. The seller needs to improve greatly and correct the problem. It needs to know that the customer is good.
3	It's good overall except for the fact that the logistics caused the hole in the clothes.
4	The quality of this cloth is terrible, I need a sales return.

As k-means clustering and spectral clustering require the number of clusters to be customized, the number of clusters in the hierarchical clustering were referred to and initialized as 4 to facilitate comparisons across the categories. Figure 5 (b) shows the visual results for the k-means clustering and Table 3 gives a representative evaluation of the 4 categories. The 1st consumer type attributed the commodity problem to the common responsibility of both the seller and the logistics; the 2nd consumer type simply pointed out the commodity quality problems; the 3rd consumer type simply described the problem and demanded a return; and the 4th consumer type described the problem and also gave positive evaluations for the logistics and customer service.

Table 3. Typical comments from each category (k-means clustering)

Category	Comment
1	There is a hole in the packaging as well as the clothing. So I think the logistics should pay more attention. This hole is very obvious and I can't wear the clothes at all. I need to give a bad rating for the logistics. Also I hope the seller can pay more attention and package it better so it won't break entirely. Also I hope the seller can label the fragile packages to eliminate these kinds of problems. I can only give an okay review to the seller and a bad review for the logistics.
2	The clothing quality is bad.
3	There is a hole and I can't return the clothes. I am very disappointed.
4	Overall is okay but there is a hole which requires careful consideration, customer service is very good .

Figure 5 (c) gives the visual result for the spectral clustering. Table 4 gives the typical comments from each category classified using the spectral clustering; the 1st consumer type attributed the problem to the seller and logistics; the 2nd consumer type gave a positive review for the logistics, seller, and customer service despite the quality problems; the 3rd consumer type pointed out the hole in the clothes and suggested several solutions (return and replacement); however, as the holes seriously affected the consumer's overall impression, the 4th consumer type attributed the problem to either the seller or to logistics.

Table 4. Typical comments from each category (spectral clustering)

Category	Comment
1	There is a hole in the packaging as well as the clothing. So I think the logistics should pay more attention. This hole is very obvious and I can't wear the clothes at all. I need to give a bad rating for the logistics. Also I hope the seller can pay more attention and package it better so it won't break entirely. Also I hope the seller can label the fragile packages to eliminate these kinds of problems. I can only give an okay review to the seller and a bad review for the logistics.
2	Customer service is very good and the clothes match the descriptions. There are no color differences and the size is appropriate The packaging and clothes are torn.
3	There was a hole in the clothes when it was sent and I can't return it I can't wear it at all so bad rating.
4	The clothes are good and the customer service is very good. But the package is broken and the clothes are torn as well. It is torn, it is torn. I'm applying for compensation with logistics.

Table 5 shows the numbers from each clustering method.

Table 5. Numbers for each clustering method

Category	Hierarchical clustering	K-means clustering	Spectral clustering
1	193	115	84
2	87	25	120
3	121	62	232
4	107	306	72

A comparison of the results from the three clustering methods indicated that the internal group structure and logical structure were relatively stable. Although the three clustering results were not highly consistent, the three typical cluster structures effectively subdivided the consumer comments into typical sub classes.

The similarities between the groups indicated that overall, most consumers felt that they needed to solve the undefined quality problems, with 60% of total consumer comments agreeing with this perspective. A further 25% of consumers believed the problem was caused by logistics, 21% believed the problem was caused by product quality, and 54% did not attribute the problem to any causes. However, no consumers attributed the quality problems to warehousing or any point prior to warehousing on the supply chain, and no other strategies besides “return” and “exchange” were mentioned; therefore, when managers are seeking to deal with undefined quality problems, controlling outcomes could be better than controlling the process ^[32].

Therefore, this study cautiously concluded the following. First, there were obvious disparities and frames of reference in the consumer perception system. In this study, of the samples (55 cases), all of which were from consumers with a higher education in business, management or economics, none were found that extended the quality problem to the supply chain before warehousing. Second, the attribution expressions were deemed to be insufficient for the consumer perception system as only 47% of the sample gave any cause attribution, with nearly half only stating the facts and asking for a return. Third, the inter-group consumer analyses suggested that strategy and behavior expectations were unified, a return was a clear proposition, and there was a general group consensus; however, there were no strategies other than “return” and “exchange”. In comparison to conclusions obtained from traditional perception quality analysis methods, this study of consumer reviews revealed the consumer’s requirements for after-sales service, which in turn revealed the problems with traditional perception quality analysis methods. From the comment information analyses, it was determined that the consumer was dissatisfied with the customer service. The contradiction between traditional perception evaluations and the text mining method proved that consumers have cognitive dissonance when using traditional methods to evaluate perceived quality.

4.2.4 Sentiment analysis

Unstructured text data contains logical information and sentiment information (tendencies and fluctuations). This paper explored the correlations between logic and sentiment information in text reviews to analyze the relationships between consumer sentiment and the consumer perception system. From the database and equations (15), the sentiment tendency E_{inc} and E_{dec} were calculated for each sentence in 508 reviews.

E_{inc} and E_{dec} were then used as cluster variables to cluster the comments and derive the cluster visualization results (Figure 6).



Figure 6. Sentiment clustering results

To compare the results of the hierarchical text clustering, it was divided into 4 categories. Figure 7 shows the comparisons for the numbers of comments in each category for the hierarchical text clustering and the sentiment analysis clustering, with the number of connections indicating the number of comments in the two classes. Figure 7 indicated that there was no statistical significance in the correlation between the logical hierarchical clustering results and the sentiment analysis clustering results. Jang and Numkang^[33] found that consumer sentiment had a significant impact on consumer preferences and perception. While this study used different analytical angles and methods, it was not possible to obtain evidence to support these conclusions.

There are several possibilities for these results.

1. Although the virtual experimental environment maximized the extent to which the consumer perception process and consumer behavior were reproduced, because the consumer results were simulated, the psychological defense mechanism possibly weakened the emotional volatility [34] and was therefore not able to accurately measure the dynamic wave data.

2. Related demand urgency research has highlighted that demand driven consumption could affect consumption mood fluctuations, which can reduce psychological defense mechanisms and cause consumer mood swings. The virtual consumer environment and objects designed in this study did not take account of the urgent demand characteristics, so there were insufficient strong emotional volatilities.

3. As the consumer perception system has separate logical attribution and emotional volatility forms under certain constraints, it is surmised that there is no correlation between the logical structure and emotional fluctuation when there is a general constraint condition. When the constraint condition changes, the violent fluctuations in emotions may replace or change the logical structure.

4. The algorithmic tool results in a loss of the time series characteristics, resulting in structural deviations.

5. CONCLUSIONS

This research explored the consumer quality perception system using text mining technology. Different from traditional perception evaluation methods, this research captured consumers' highly unified solutions to undefined quality problems, and proved that the consumer quality perception system had a typical line-of-sight, with consumers usually attributing undefined quality problems to only those aspects of the supply chain they can perceive. It was found that customer service control was more efficient than process control. However, determining more accurate ways to assess the sentiment tendencies in texts to explore the relationship between consumer sentiments and their perceptions requires further study.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China Committee 71350007, the Science and Technology Department of Shaanxi Province Government 2013KCX(2)008, the Shaanxi Land Construction Group SXDC201208.

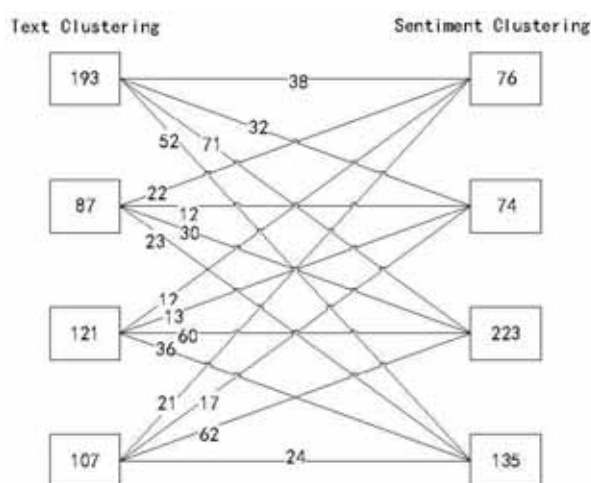


Figure 7. Text clustering and sentiment clustering comparison

REFERENCES

- [1] Jacoby J, Olson J, Haddock R. (1971). Price, brand name, and product composition characteristics as determinants of perceived quality. *Journal of Applied Psychology*, 55(6): 570-579.
- [2] Grönroos C. (1984). A Service Quality Model and its Marketing Implications. *European Journal of Marketing*, 18(4): 36-44.
- [3] Smith A, Bolton R. (2002). The effect of customers' emotional responses to service failures on their recovery effort evaluations and satisfaction judgments. *Journal of the Academy of Marketing Science*, 30(1): 5-23.
- [4] Marakanon L, Panjakajornsak V. (2017). Perceived quality, perceived risk and customer trust affecting customer loyalty of environmentally friendly electronics products. *Kasetsart Journal of Social Sciences*, 38(1): 24-30.
- [5] Lee G, Lin H. (2005). Customer perceptions of e-service quality in online shopping. *International Journal of Retail & Distribution Management*, 33(2): 161-176.
- [6] Steenkamp J. (1990). Conceptual model of the quality perception process. *Journal of Business Research*, 21(4): 309-333.
- [7] Folkes V, Kotsos B. (1986). Buyers' and Sellers' Explanations for Product Failure: Who Done It? *Journal of Marketing*, 50(2): 74-80.
- [8] Brochado A, Pereira C. (2017). Comfortable experiences in nature accommodation: Perceived service quality in Glamping. *Journal of Outdoor Recreation & Tourism*, 17: 77-83.
- [9] Buzzell R, Gale B. (1989). The PIMS principles : linking strategy to performance. *Journal of Marketing*, 53(2): 126.
- [10] Parasuraman A, Zeithaml V, Berry L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49: 41-50.
- [11] Marek M. (2005). Evaluating a museum as a tourist product using the servqual method. *Museum Management & Curatorship*, 20(3): 235-250.
- [12] Cronin J, Taylor S. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing*, 56(3): 55-68.
- [13] Peter J, Churchill G, Brown T. (1993). Caution in the Use of Difference Scores in Consumer Research. *Journal of Consumer Research*, 19(4): 655-662.
- [14] Styliadis K, Madrid J, Wickman C, & Söderberg R. (2017). Towards Overcoming the Boundaries between Manufacturing and Perceived Quality: An Example of Automotive Industry. *Procedia CIRP*, 63: 733-738.
- [15] Choi T, Eboch K. (1998). The tqm paradox: relations among tqm practices, plant performance, and customer satisfaction. *Journal of Operations Management*, 17(1): 59-75.
- [16] Özer Ö, Raz G. (2011). Supply chain sourcing under asymmetric information. *Production & Operations Management*, 20(1): 92-115.
- [17] Davis D, Allen J, Crosenza R. (1988). Segmenting local residents by their attitudes, interests, and opinions toward tourism. *Journal of Travel Research*, 27(2): 2-8.
- [18] Gerard S, Wong A, Yang C. (1974). A vector space model for automatic indexing. *Communications of the Acm*, 18(11): 613-620.
- [19] Gerard S, Christopher B. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5): 513-523.
- [20] Dhillon, Inderjit S, Dharmendra S. (2001). Concept decompositions for large sparse text data using clustering. *Machine Learning*, 42(2): 143-175.
- [21] Xu D, Tian Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2), 165-193.
- [22] Rohlf F. (1982). 12 single-link clustering algorithms. *Handbook of Statistics*, 2(82): 267-284.
- [23] Ward J. (1963). Hierarchical grouping to optimize an objective function. *Publications of the American Statistical Association*, 58(301): 236-244.

- [24] Dan P, Moore A. (2000). X-means: Extending K-means with Efficient Estimation of the Number of Clusters. Seventeenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc, 727-734.
- [25] Luxburg, Ulrike. (2007). A tutorial on spectral clustering. *Statistics & Computing*, 17(4): 395-416.
- [26] Socher R, Perelygin A, Wu J, Chuang J, Manning C, & Ng A, et al. (2013). Recursive deep models for semantic compositionality over a sentiment treebank, *EMNLP*, 1631: 1631-1642.
- [27] Oscar R. (1972). Positivistic measures of aggregate supply elasticities: some new approaches: comment. *American Journal of Agricultural Economics*, 54(3): 530-531.
- [28] Hoffman D, Kalsbeek W, Novak T. (2000). Internet and web use in the united states: baselines for commercial development. *Communications of the Acm*, 39: 36-46.
- [29] Zellner, Wendy. (2000) Wooing the Newbies.(efforts of online retailers to attract new customers). *Business Week*, 3681: 116.
- [30] Bill D, Stuart D, Fiona T, Jim C. (2005). Perceived risk and the internet shopping experience in online purchasing behaviour. *Journal of Global Information Management*, 13(2): 66.
- [31] Nikhashemi S, Tarofder A, Gaur S, Haque A. (2016). The effect of customers' perceived value of retail store on relationship between store attribute and customer brand loyalty: some insights from malaysia . *Procedia Economics & Finance*, 37: 432-438.
- [32] Liu S, Wang L, & Huang W. (2017). Effects of process and outcome controls on business process outsourcing performance: moderating roles of vendor and client capability risks. *European Journal of Operational Research*, 260(3): 1115-1128.
- [33] Jang S C, & Namkung Y. (2009). Perceived quality, emotions, and behavioral intentions: application of an extended mehrabian–russell model to restaurants. *Journal of Business Research*, 62(4): 451-460.
- [34] Kuftyak E. (2014). Psychological defense and coping in the prognosis of children mental adaptation. *Procedia - Social and Behavioral Sciences*, 159: 180-183.

The Influence of Media News Frame on Consumer's Brand

Attitudes in the Context of Product-harm Crisis

Silan Li, Shiyu Liu, Mengzhi Xian*

Hengda School of Management, Wuhan University of Science and Technology,
Wuhan, 430065, China

Abstract: This article examines the relationships between the media news frame and consumer brand attitude after the product-harm crisis. Based on the past classification about negative events, this article divides the media news frame into two kinds: disability type and immoral type. By a 2(media frame: disability VS immorality) * 2 (severity of crisis: high VS low) experiment, the findings show that in the case of product harm crisis, the customer brand attitude is more significantly affected by the immoral type media news frame. In the meanwhile, the significance of this frame effect is affected by cognitive need and severity of crisis. For the consumers with high cognitive need, the frame effect is not obvious. But for the ones with low cognitive need, the frame effect is obvious. The customer brand attitude is more obviously affected by the media news frame in a circumstance of high crisis than low crisis. Therefore, when a product-harm crisis happened, the firms should care about the news reporting format so that the customer brand attitude will not be seriously damaged and the brand image can be well restored.

Keywords: frame effect, brand attitude, product-harm crisis, cognitive need

1. INTRODUCTION

The term 'product-harm crisis' refers to well-known events related to product defects or harm associated with some brands^[1]. In fact, every enterprise cannot be perfect, so the product harm crisis is almost a problem that each enterprise can not avoid. Once an enterprise is confronted with a severe product-harm crisis, it is likely to experience a disaster. In order to help enterprises with product-harm crisis better respond to consumers' questioning and repair its image, many studies have been carried out in this field in recent years and have achieved fruitful results. Nonetheless a lot of them mainly concentrated on exploring the impact of coping style or consumers' responses. Nowadays news media exert great influence on social orientation of public opinion. Especially with rapid growth of social media, negative publicity may even completely destroy a brand. When product-harm crisis event---as a negative publicity happened, consumers will actively search for crisis related information, attribute who should be responsible for the incident^[2]. Different reporting style may have different effects on consumers, thus affecting consumer's brand attitudes after crisis event---may be slightly, or may be very seriously. For the same negative publicity, different media will report from different perspectives so that individuals can form different cognition. When different media expressions give rise to consumers' different cognition, "frame effect" is generated. As to frame effect, the research achievements are fruitful both in domestic and abroad at present, which also includes the research of negative publicity. But in the case of product harm crisis---as a typical negative publicity, the question how media frame affects consumers brand attitudes remains unsolved in the current study. Obviously, after crisis event, a quick and efficient way for consumers to get to know the progress of that event is through media coverage. So we can say consumer's brand attitudes are subject to different types of reporting frame. In this article, we hope to explore the possible regulatory factors in the process of media news frame influencing consumers' attitudes to brand in the occurrence of product-crisis. From the theoretical point of view, this research intends to build a model illustrating the media news frame effect and

* Silan Li. Email: lsl200168@126.com

on exploring the related factors that may affect consumers' brand attitudes. From the practical point of view, this research intends to provide enterprises and society the theoretical basis on how to effectively use the media to spread information and later repair its image after product-harm crisis.

2. LITERATURE REVIEW

The theoretical study of the frame was derived from <The Frame Analysis> published by Goffman in 1974, which clearly illustrated the frame. He believes that the frame can quickly help people understand the surrounding social reality and translate the reality into subjective thought and give people a chance to form their own cognition^[3]. Based on the sociological perspective, Goffman studied the frame, believing that the frame is partly caused by the individual's past experience and also be influenced by the social cultural consciousness. Research on frame effect originated from Kahneman and Tversky's 'Asian disease' experiment, it proved that people will make different decisions when an event is expressed by two ways of positive and negative, so in this case, people are affected by frame effect, making themselves no longer a rational decision maker base on the facts^[4]. The frame effect has long been studied in the field of Psychology, while in the field of Marketing, researches are relatively rare. This paper intends to broaden the study field of the theory of frame effect to marketing and product-harm crisis circumstances.

3. CONSTRUCTIONS AND MODEL

3.1 Two different media news frame: disability and immorality

Compared with positive events, a lot of researches are more inclined to study the effect of information frame in the context of negative events. In the context of corporate negative events, most of the previous studies are carried by the ability frame and morality frame. There are different types of media frame, for example, the media tends to blame the corporate immorality, or to blame the corporate disability. Disability and immorality, are often the two criticisms of a business after a negative event happened, we can see them as two different angles of judgment. The enterprise's crisis may due to the lack of ability to provide products, for example, the production process is not well controlled, therefore lead to quality problems; It may also due to the immoral problems of the enterprise, such as using inferior or harmful material or parts in the process of production in order to reduce costs^[5]. Therefore, consumers will also evaluate the product-harm crisis events from the two dimensions of disability and immorality: disability type (such as product quality is not good, technology is not good enough), immoral (such as lack of basic morality, make decision only by self-interests). Theoretically, these two kinds of crisis events can be defined as a crisis formed by the lack of ability to fulfill their promises to consumers(disability), and a crisis formed by conflicts with the moral standards established by consumers or society(immorality).

Base on this standard, the media news frame that affects consumers is divided into disability frame and immorality frame. In other words, if news media pay more attention to condemning the enterprises lacking of ability after product-harm crisis, consumers will tend to rationally attribute that event to enterprise lack of ability, conversely to condemn the enterprise lack of morality, consumers may perceptually attributed that event to the enterprise lack of business ethics.

3.2 Cognitive need

Cognitive need refers to an individual's tendency to think^[6]. It reflects the tendency of individuals to engage in and enjoy thinking activities. It exerts great effects on the depth and breadth of information extraction and judgment. Cacioppo believes the essence is that different individual's cognitive motivation is different. people with high cognitive need tend to pay more attention to the essence of information, and are willing to

devote more energy to process information, analyze and process information more comprehensively, thus forming a relatively more objective cognition. Therefore, when people with high cognitive need receive information, they will think more deeply and collect some related data to understand and analyze the information more comprehensively, depend more on their own understanding of information. People with lower cognitive need, on the contrary, are reluctant to spend more energy to process information, and are more dependent on others heuristic cognitive or social comparison process. Since they rely on the outside world but not to analyze the information through their own effort, people with low cognitive need are more susceptible to the influence of external information. In the context of product-harm crisis, every consumer has different tendency in information processing. Some consumers may be particularly susceptible to the frame effect, and some may not. Therefore, the impact of brand attitudes caused by different types of media frames be regulated by consumers' cognitive need.

3.3 Consumer brand attitudes

Attitude is a very important concept both in the field of psychology and marketing. In marketing, consumer attitudes refers to the acquisition tendency of consumers for specific objects in a specific situation, which can guide consumers to form relatively stable positive or negative behaviors towards the object^[7]. Today, scholars define attitude from the perspective of structure and mechanism of action, so there is a more comprehensive view of attitude. Attitude is a lasting system including cognition, emotion, and behavior tendency. Thus we can say, it is the integration of cognition, emotion and behavior that form consumer attitude. In the view of the mechanism of the attitude, consumers evaluate the brand through cognitive experience and emotional experience, then produce the behavior of buying or not buying. It shows that the consumer's purchase behavior is completely done in good emotional and cognitive conditions. Therefore, once the crisis events happened, the first consideration for an enterprise is to save the consumers' emotion and cognition, minimize the negative emotions of the public, and make time for enterprises to think out a way to repair consumers' purchase intention. Therefore, in this study, given that media news frame will exert influence on consumers, we will measure the brand attitudes of consumers by two dimensions of cognition and emotion. To sum up, the research frame and main variables of this paper are shown in Figure 1.

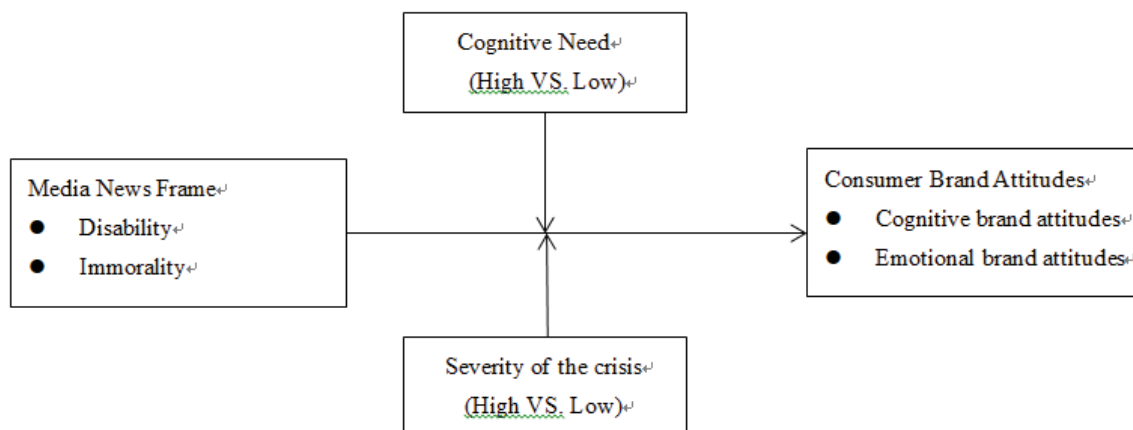


Figure 1. The research frame

4. THEORETICAL DEDUCTION AND HYPOTHESIS

4.1 Different reporting frames cause different consumer's brand attitudes

Pulling (2006) divided the negative information into two types: ethics-related and product-related^[8]. Media's reports on product-harm crisis is obviously a negative one. Therefore, based on the classification of

enterprise negative information, we divide the news reporting frame into disability frame and immorality frame. Of course, sometimes a crisis occurs because of both lack of morality and ability. But according to the related research of media news frame, the media will selectively report and highlight some truth, thus forming the frame of consumers' judgment. The researches show that in the positive information frame, satisfaction in individual decision-making process and confidence on result is significantly higher than those in negative information frame. Empirical research pointed out that when faced with a crisis, consumers are subject to be influenced by the way of problem expression. Research on negative events of disability and immorality also found that violation of emotional morality is more incline to generate a severe and lasting impact than lack of ability. According to these researches the following assumptions is made:

H1: In the context of product harm crisis, compared with disability frame, consumer's brand attitudes is influences more significantly by immorality frame.

4.2 The regulatory role of cognitive need

The existing research shows that under the positive frame, cognitive need can significantly affect individual decision outcomes. Under the negative frame, for people with high cognitive need, frame effect is not as significant as that with low cognitive need. When cognitive need and frame conditions are considered as dependent variables respectively---do not consider their interaction, have significant impacts on the individuals' decision-making. Similarly, the negative publicity of enterprises have been studied from the moral perspective, and the regulatory role of cognitive need in immoral negative publicity has been discovered^[9]. This research found that the negative spillover effect of the people with low cognitive need is more significantly influenced by negative publicity of the brand than those with high cognitive need. What's more, consumers with low cognitive need are more subject to the influence of immoral issues. Based on these studies, we infer that consumers with high cognitive need are less affected by frame effect than those with low cognitive need. Therefore, the following hypothesis is made:

H2: In the context of product-harm crisis, compared with consumers with high cognitive need, consumers with low cognitive need is more influenced by the media news frame.

4.3 The regulatory role of the degree of crisis

Vassilikopoulou and other scholars(2009) defined the severity of crisis as the number and extent of the casualties of people, animals and the environment that caused by the crisis. When consumers are unable to accurately get the data caused by the crisis, they will speculate the severity of the event. Therefore, a crisis event which can not be quantified through data, its severity is likely to be different to consumers. The more serious the crisis is, the higher consumer's perceived risk^[10]. Both enterprises and consumers are affected by different degrees of crisis events. In the studies of the impact of negative events on consumer confidence, scholars proved that no matter in what kind of negative events (integrity or ability), as long as the severity of that incident get higher, consumer's trust will be more significant damaged. As a result, consumers brand attitudes may also be significantly affected in the context of high degree of crisis. To sum up, the following assumption is made:

H3: Compared with the lower degree of crisis, consumer's brand attitudes are more affected by media news frame in the context of higher degree of crisis.

5. EXPERIMENT AND STATISTIC ANALYSIS

5.1 Pretest

The purpose of pretest is to design and test the context of product harm crisis in the empirical research, control the degree of severity, and design the media news frame, control the way when report the disability and

immorality. According to the hypothesis, the experimental scenarios are 2(media frame: disability VS immorality) * 2 (severity of crisis: high VS low).

The first step in the experiment is to identify a crisis-generated industry. After studying 148 product-harm crisis events in all kinds of industries, we sum up 8 classic cases, finally chose three representative industries from the perspective of social awareness. These industries are food manufacturing industry, dairy manufacturing industry and drinks manufacturing. As food and drinks are closely related to the health of consumers, they are easy to attract the attention of the media and consumers. Therefore, this article designed a situation related to the beverage industry, which is more likely to arouse the situation association of respondents.

The second step is to consider the type of media news frame. In order to prevent consumers from being too subjective to the scene, we described the scenario more like news. By referring to some main internet news, we organized the event in our own statements to make the text more clear. For the construction of two media reporting frames of disability and immorality, we described the company of being "restricted by technology" and criticized the company "reduce cost and violate the rule". By such keywords to make consumers have a basis for judgment.

The third step is to construct crisis. Since what we need to construct is media news reporting frame of product-harm crisis, therefore the brand and company names were fictitious. So participants' prior attitudes would not affect their responses. In summary, we chose to construct a high/low product harm crisis in the beverage industry respectively on our own. Through the analysis of scholars, we can find that both the media and the consumers are more concerned about the crisis that is harmful to human health, so we set the crisis into two situations that may cause safety issue to human health. At the same time, considering consumers may pay much too attention to their health so that they may also see some small harm to health as serious health threats, and then judge that crisis as very serious, we emphasized "mild allergy" in the construction of low risk scenarios, which is greatly contrasted with "severe diarrhea and vomiting" in high risk situations. According to the ideas and construction methods above, four kinds of questionnaires were designed.

In total, 40 pretest questionnaires were retrieved. In order to exclude the effects of area and age, the questionnaires were distributed in real time through the network, and the communication with the respondents was maintained. For the news frame, the accuracy rate of the report of disability and immorality reached 100%, which shows that everyone has made correct judgment. Therefore, it is sure that the media coverage frame operated successfully under the four scenarios. For high degree crisis event, the mean of Likert 7 scale is $6.7 > 4$, which indicates that the degree of the crisis is high. For low degree crisis event, the mean is $3.3 < 4$, which indicates the crisis is of low degree. From these 2 results, the control of new frame and severity of crisis is successful.

5.2 Formal experiment

In this experiment, 164 questionnaires were collected. After excluding invalid questionnaires, the remaining valid questionnaires were 138, and the effective rate was 84%, including 66 men and 72 women. The experiment is a group-between of 2 (media news frame: disability VS immorality) * 2 (severity of crisis: high versus low). Subjects were randomly assigned to one of the scenarios. In order to get rid of the interference, we didn't mention the brand name of that beverage in the experiment scenario.

First of all, at the top of the questionnaire, a news scene is designed. There are some words like "the following is news, according to media reports..." The product-harm crisis is described in the tone of news report. The causes of the crisis were first explained, and the underlying causes of the crisis were revealed in an in-depth investigation. The description of the whole news will form a frame for the subjects. After reading the background information, the subjects were asked to answer the questionnaire. The questionnaire was composed

of the Freedman's mature brand attitude scale and the 18 items of cognitive need scale which is revised by Cacioppo in 1984.

5.2.1 Frame effect hypothesis test

The experimental group was divided into disability group and immorality group. By using the mature Freedman (2001) classic attitude measurement model and the ANOVA analysis we got some data as table1 showed. In the immorality group, subjects' brand attitudes was $M_{\text{Immority}}=1.60$ on average. While the subjects's brand attitude of disability group was $M_{\text{Disability}}=2.85$ ("1-7 points" indicates "very dissatisfied -- very satisfied"). $F=52.881$, $P=0.000<0.05$, which indicates that the difference of consumers' brand attitudes result from the impact of the two media news frame is statistically significant. H1 is proved.

Table 1. Media news frame effect

	SS	df	MS	F	P
Group-between	54.219	1	54.219	52.881	0.000
Group-in	139.442	136	1.025	—	—
Total	193.661	137	—	—	—

5.2.2 The regulatory role of cognitive need test

We took the median of cognitive needs score as the midpoint, divided the subjects into two groups: high cognitive need and low cognitive need. Through ANOVA variance analysis, we got the following conclusions: For high cognitive need group, $M_{\text{Immority}}=2.37$, $M_{\text{Disability}}=2.44$, $F=2.64$, $P=0.22>0.05$; For low cognitive needs group, $M_{\text{Immority}}=1.87$, $M_{\text{Disability}}=2.01$, $F=6.005$, $P=0.016<0.05$, this data showed that for high cognitive need group, the frame effect is not significant but for the low cognitive need group, the frame effect is significant. H2 is proved.

Table 2. The regulatory role of cognitive need

		SS	df	MS	F	P
Group-between	High	6.335	1	5.832	2.64	0.22
	Low	8.877	1	8.877	6.005	0.016
Group-in		201.029	136	1.478	—	—
Total		209.906	137	—	—	—

5.2.3 The regulatory role of the degree of crisis test

Dividing experimental groups into high-risk group and low risk group according to the designed crisis scenario, we examined the difference of consumer brand attitudes between these two cases. Though ANOVA analysis, the following table3 was obtained. Under high crisis situation, $M_{\text{Immority}}=1.92$ and $M_{\text{Disability}}=2.50$, $F=8.786$ and $P=0.004<0.05$; Under low crisis situation, $M_{\text{Immority}}=2.97$ and $M_{\text{Disability}}=3.14$, $F=5.526$ and $P=0.34>0.05$; these data showed that the frame effect is significant under high crisis situation but insignificant under low crisis one. H3 is proved.

Table 3. The regulatory role of the degree of crisis

		SS	df	MS	F	P
Group-between	High	11.752	1	11.752	8.786	0.004
	Low	9.757	1	10.833	5.526	0.34
Group-in		181.910	136	1.338	—	—
Total		193.661	137	—	—	—

6. CONCLUSIONS

6.1 Contributions

This research explored the influence of media news frame on consumer brand attitudes in the context of product harm crisis. It is a new development for the existing product harm crisis research, and it also extends the research of information frame effect to product-harm crisis context. Previous researches on frame effect have been used as the basis to further explore the information frame effect. Taking negative events product-harm crisis as background, this study explored whether different reporting ways of news media on the same crisis would affect consumers' brand attitudes. Through empirical studies, we found consumers are really affected by the information frame. Furthermore, we introduced the cognitive need from the perspective of individual consumers and the level of the crisis as regulatory variables and proved that the consumers with high cognitive need is less affected by the frame effect than those with low cognitive need, the higher level of the crisis, the more significant frame effect. It provided the boundary of the frame effect functioning.

6.2 Implications of Management

This study provided a valuable reference for enterprise who is involved in product-harm crisis to response. The enterprise can use correct information release to make people generate attribution which is benefit to it. This study found that the immoral reporting frame is easy to make consumer's brand attitudes more negatively affected. Lack of ability can be forgiven by consumers, but lack of morality can hardly win the trust of consumers again. The enterprise should monitor the media's report after product-harm crisis. When negative publicity happened, if it is really the enterprise's mistake, the firm should immediately admit the mistake to the public and make promise to compensate. Then control the spread of crisis in a proper way, especially when the media criticize the enterprise's immorality. So after the product harm crisis, the enterprise should not only pay attention to its own repair measures to consumers, but also pay attention to the media.

This study proved that when the enterprise has a very high degree of crisis, it should control the media that criticize the immoral problems of the enterprises, which will not make consumers to have very strong moral condemnation to enterprises and thus more negative attitudes to the brand. This study also found that people with different cognitive need may be differently affected by media news frame. Consumers with high cognitive need will not be affected by media news frame and may only pay attention to the facts. For consumers with low cognitive need, companies can affect such groups through the media frame. Although it is difficult for the enterprise to distinguish the people with high cognitive need from those with low cognitive need, it can do distinguish the different style of media reporting. For example, the information with more pictures may arouse the attention of people with lower cognitive but the information with more words describe may arouse the attention of people with higher cognitive. So enterprises pay more attention to the former style of information because the people with low cognitive need may be more likely affected by the media new frame.

6.3 Limitations and Directions

All the exploratory studies have some unavoidable limitations, and so does this study. First of all, though these scenarios have been verified by the pretest and proved that they are in line with the requirements, and all the questionnaires designed in this study were made under the circumstances controlled by the experimenter. It may result in inevitable external validity problem. In the future, the researcher may use internet big data to make the study media news frame effect in the context of product-harm crisis.

This research investigated media news frame effect and the regulatory role of cognitive need and severity of crisis that news frame affected consumer's attitude in the context of product-harm crisis. But the underlying mechanism how consumer's attitude is affected by new frame is unsolved. In the future, the underlying psychological mechanism is a valuable research direction.

ACKNOWLEDGEMENT

This research was supported by the Humanities and Social Science Research Planning Foundation of Ministry of Education of China under Grant 17YJA630043 and the Humanities and Social Science Research Project of Hubei Provincial Education Department of China under Grant 16D009.

REFERENCES

- [1] Siomkos G, Kurzbard G. (1994). The hidden crisis in product-harm crisis management. *European Journal of Marketing*, 28(2): 30~41
- [2] Seon-Kyoung A, Karla K G, Seung H C. (2011). Level of crisis responsibility and crisis response strategies of the media. *Journal of Communication Management*, 15(1): 70~83
- [3] Goffman E. (1974). *Frame analysis an essay on the organization of experience*. New York, paper&Row,.
- [4] Tversky A, Kahneman D. (1981). The Framing of Decisions: The Psychology of Choice. *Science*, 211(4481): 453-458
- [5] Xie Y, Peng S. (2009). How to repair customer trust after negative publicity: The roles of competence, integrity, benevolence, and forgiveness. *Psychology & Marketing*, 26(7): 572~589
- [6] Cacioppo J T, Petty R E. (1982). The Need for Cognition. *Journal of Personality and Social Psychology*, 42(1): 116~131
- [7] Schiffman, Kanuk. (2000). *Consumer Behavior*. 7th ed. Upper Saddle River. NJ: Prentice-Hall Inc.
- [8] Pullig C, Netemeyer R G, Biswas A. (2006). Attitude basis, certainty, and challenge alignment: A case of negative brand publicity. *Journal of the Academy of Marketing Science*, 34:528~542
- [9] Zhuang A L, Yu W P. (2011). Empirical research on the spillover effect of negative publicity of moral related brand: the regulatory role of events type and cognitive need. *Journal of Business Economies, China*, 240(10): 60~67.
- [10] Vassilikopoulou A, Siomkos G, Chatzipanagiotou K, et al. (2009). Product-harm crisis management : Time heals all wounds ?[J]. *Journal of Retailing and Consumer Services*, 16(3):174-180.

A Study on Consumers' Learning Effect in the Price

Reduction Auction: a case study of Gongtianxia

Yanan Xu, Bo Yang*, Rong Zhang, Pengfei Xie

Information Institute, RENMIN UNIVERSITY of CHINA, Beijing 100872, China

Abstract: GongTianXia's "7 days shot" and "15 minutes shot" is a very typical online price reduction auction. By studying the learning effect of consumers in this online price reduction auction, this paper hopes to provide some suggestions for the formulation of marketing strategy of GongTianXia. Through the research of related literature and the deep analysis of consumer purchasing decision theory, the research problem is divided into two aspects. First, with the increase in the number of participants, whether the proceeds of the auction goods are gradually increasing. Second, with the increase in the number of participants, whether the grasp of the timing is more and more accurate. Through the statistical analysis of the auction data of GongTianXia, this paper verified the existence of consumer learning effect in the price reduction auction.

Keywords: Price reduction auction; Learning effect; Consumer purchase decision theory

1. INTRODUCTION

From the beginning of 2014, Gongtianxia.com carried out agricultural and sideline products price-reduction auctions, which were mainly divided into "7-day auction" and "15-minute auction" two types^[1]. 7-day auction gets new goods on shelf at 10 pm on the first day, and sets this time as 1st day of the auction of this goods with the initial auction price. Then the auction price will be reduced to a certain price by the platform at 10 pm in the next 6 days, which happened only once at 10 o'clock. It is possible for consumers to bid the good at that certain price during this auction period until 10 pm of the next day. When auction comes to 10 pm of the seventh day, auction price will drop to 1 yuan for all goods. There is a special rule that the total amount of an auction good is limited, which means auction can be stopped in advance if all the auction goods are sold out. The rule of 15-minute auction is similar to 7-day auction. It differs from 7-day auction's rule in the price-dropping interval, and the price will drop to 1 yuan at the last auction period. Both "7-day auction" and "15-minute auction" are 'an efficient marketing method to stimulate the consumers' purchasing behavior^[1].

The definition of *Learning Effect* is staffs can accumulate experience about product production, technical design or management during long-term production process in order to enhance their productivity and reduce average production cost^[3]. Through rough observation of our collected data of Gongtianxia.com, we found one user of platform took a gradually higher bid and stayed stable on a particular price tier during his auction experience^[5]. There still remains necessity to test whether a single user's auction purchase behavior is significantly influenced by learning effect. We hope our research can analyze and explain the learning effect in this novel marketing mode, thus provide some suggestions on making marketing strategies for auction E-commerce platform.

2. RESERCH CONTENT

This paper aimed at research on the learning effect present by individual consumers in the price-reduction auction process of agricultural and sideline products, mainly expanded by the following two problems.

The first problem is with the increase of the participating number of auctions, whether the auction gain is gradually increasing or not, which is represent by the difference between initial auction price (1st price) and the

* Corresponding author. Email: yangbo_ruc@126.com (YANG Bo)

final bidding price^[3]. In other words, consumer can decrease their final bidding price payment by choosing a higher auction tier, to get a larger auction gain. This issue aims at proving that there exists a learning effect in the auction of the agricultural and sideline products for consumers essentially. In order to show whether there is a learning effect, we have established a multiple linear regression model. For the linear regression model, we have added the appropriate control variables to make the results more reliable according to the relevant literature and theoretical research. And then the data and then detailed statistics and analysis to prove that in the agricultural and sideline products price reduction in the consumer whether there is a corresponding learning effect.

The second one refers to the grasp accuracy of the bidding timing with the increasing participating times. We can also divide this problem into two issues. One is the estimation accuracy of the last auction tier (because auction will be stopped before 7th tier if goods are sold out). The other is about the time length used for finalizing their bidding on a particular price tier^[3]. Due to the limited amount of remaining goods, the faster a consumer bids for the good, the more possible he can finalize their bidding for this auction.

3. LITERATURE REVIEW

The price-reduction auction, also known as the Dutch auction and originated in the Dutch flower market, is a typical auction type for flower sales^[1]. As for sales promotion, Ding and Guo (2005) attached importance to analysis on Dutch auction from economic perspective, and took advantage of empirical research^[1]. Jennifer Neujahr and Qiao (2006) discussed the problems existed in the auction^[9]. Zhang (2007) analyzed the existing situation and some unique characteristics of Chinese domestic flower auction market in detail, and summarized the shortcomings of it^[9]. Qin (2012) systematically and comprehensively studied the law of possible price fluctuation in Chinese domestic agricultural auction market and the related great influencing factors^[13].

Learning Effect dated back to 1930s. Wright (1936) found that the unit time of the production aircraft would be reduced by 20% when the aircraft production was doubled in the study of aircraft production costs^[8]. For a relatively long period of time, studies of learning effects have been mostly concentrated in the production of manufacturing enterprises. Nowadays, the study of learning effect has been extended to many industry fields. In recent years, there has been some articles about the research on consumer behavior in online auctions affected by learning effect. PB Goes, GG Karuga, AK Tripathi (2012) set research object as consumer behavior in online auctions and explained the decision process by an experience model^[10]. Their research found that consumers' decision was influenced by auction desires, previous related experience and a few parameter design in auctions. H Zheng, KY Goh, KW Huang (2011) analyzed consumer behaviors under "1-cent auction" mode, and achieved optimization by adjusting limit conditions^[14]. Ma (2014) constructed a mathematical model based on influence of learning effect made on enterprises' promotion activities^[14]. However, most related researches focused on learning effect on enterprise level instead of individual behaviors.

4. CHANGES OF CONSUMERS' BIDDING GAIN WITH THE INCREASE OF AUCTION PARTICIPATING TIMES

Consumers generate learning effects in order to obtain a more favorable price advantage in the auction^[8]. In the initial auction, due to restrictions on auction total amount and inadequate understanding of the auction, so consumers are willing to give up part of price advantage to grab inexpensive goods, that is, consumers might bid on a relatively low price tier (high price) from their beginning use of the auction E-commerce platform. With the increase in the number of auctions consumer had participated, consumers had a better understanding of the auction process, then he will try to adjust his own auction prices to increase their bidding gain. We recognized this adjustment as an expression of auction learning effect. We built a multiple linear regression model in order to verify our basic hypothesis quantifiably and intuitively.

4.1 Variable selection

Firstly, we decided on the selection of dependent and independent variables. Based on literature review, we knew learning effect made great effect on consumer's auction behavior, so it could be verified if the auction behavior were influenced largely by previous bidding. As for consumer's auction behavior, the final bidding price is an important measure of consumer auction behavior. However, different goods have different prices, so it was not feasible to set the price directly as a variable. Considering the bidding rules of both 7-day auction and 15-minute auction, we found standard seven price tiers could represent the auction prices of various goods, which simplified the standardized process. But as a typical categorical variable, it should not be applied to linear regression. As a result, we introduced price reduction extent, which seen as a bidding gain for consumer from another prospective. In summary, we set bidding gain (price reduction extent for purchased good in this time) in current auction as dependent variable, and bidding gain in previous auction.

Secondly, suitable control variables contributed much to the accuracy of regression model, which is related to consumers' bidding psychology. Consumer purchase decision-making theory refers to the evaluation and final selection of the properties of goods during the purchase process of a specific type of goods^[15]. The whole process includes the determination of consumer needs, the emergence of the purchase motive, the analysis and selection of a variety of commodity purchase options, and following series of practical test after the purchase. The information available from the data collected on Gongtianxia E-commerce platform is primarily relevant to the marketing activities of the price-reduction auction. For the sake of the bidding gain obtained in each auction, all historical, current and future price were still the first consideration of consumers to determine their purchase^[17]. As a result, we decided to add seven price-reduction extent to our model as control variables, which were processed in advantage of seven price tiers. 1st price-reduction extent was equal to 0 constantly, so there were six price-reduction extents in total actually. In addition, total remaining amount is limited in one price-reduction auction, stimulating and reminding consumers to bid as soon as possible, so our research considered remaining number of goods as a control variable. Finally, according to the habits of consumers to buy goods through online E-commerce, whether set free for delivery or not, should be included into the linear regression as a control variable.

4.2 Linear multiple regression model of consumer's bidding gain

We constructed following linear multiple regression model to test whether current consumer's bidding gain is connected with the previous one on the basis of explanation in 4.1.

$$BG(i, j, t) = \beta_0 + \beta_1 * BG(i, j, t-1) + \alpha_1 * PR_2(i, j, t) + \alpha_2 * PR_3(i, j, t) + \alpha_3 * PR_4(i, j, t) + \alpha_4 * PR_5(i, j, t) + \alpha_5 * PR_6(i, j, t) + \alpha_6 * PR_7(i, j, t) + \alpha_7 * AMT(i, j, t) + \alpha_8 * DEL(i, j) + \varepsilon(i, j, t) \quad (1)$$

Among them, $BG(i, j, t)$ means the bidding gain obtained by the consumer_i bids for j^{th} kind of goods for the t^{th} time, while $PR_n(i, j, t)$ means the n^{th} price reduction extent when the consumer_i bids for j^{th} kind of goods for the t^{th} time. $AMT(i, j, t)$ and $DEL(i, j)$ means the remaining amount of the goods and whether it was free for delivery when consumer_i bids for j^{th} kind of goods for the t^{th} time. $\varepsilon(i, j, t)$ is the random error.

4.3 Regression result

4.3.1 7-day auction mode

Table 1. Abstract of regression models

Model	R	R ²	Adjusted R ²	Standard estimation error	Change statistics				
					Change of R ²	Change of F	Degree of freedom 1	Degree of freedom 2	Change of Sig. F
1	.709 ^a	0.502	0.502	0.1065168	0.502	4672.713	1	4628	0.000
2	.765 ^b	0.586	0.586	0.0971946	0.083	931.342	1	4627	0.000
3	.774 ^c	0.599	0.598	0.0956798	0.013	148.670	1	4626	0.000
4	.776 ^d	0.603	0.602	0.0952095	0.004	46.818	1	4625	0.000
5	.777 ^e	0.604	0.604	0.0950419	0.001	17.327	1	4624	0.000
6	.778 ^f	0.606	0.605	0.0948867	0.001	16.137	1	4623	0.000

a. Estimation variable: (Constant), BG(t-1)

b. Estimation variable: (Constant), BG(t-1), PR6

c. Estimation variable: (Constant), BG(t-1), PR6, AMT

d. Estimation variable: (Constant), BG(t-1), PR6, AMT,DEL

e. Estimation variable: (Constant), BG(t-1), PR7, AMT,DEL,PR5

f. Estimation variable: (Constant), BG(t-1), PR7, AMT,DEL,PR5, PR7

Table 2. ANOVA analysis result

Model		Sum of squares	Degree of freedom	Mean square	F	Sig.
1	Regression	53.016	1	53.016	4672.713	.000 ^b
	Residual	52.509	4628	0.011		
	Sum	105.524	4629			
2	Regression	61.814	2	30.907	3271.693	.000 ^c
	Residual	43.710	4627	0.009		
	Sum	105.524	4629			
3	Regression	63.175	3	21.058	2300.296	.000 ^d
	Residual	42.349	4626	0.009		
	Sum	105.524	4629			
4	Regression	63.599	4	15.900	1754.014	.000 ^e
	Residual	41.925	4625	0.009		
	Sum	105.524	4629			
5	Regression	63.756	5	12.751	1411.630	.000 ^f
	Residual	41.768	4624	0.009		
	Sum	105.524	4629			
6	Regression	63.901	6	10.650	1182.898	.000 ^g
	Residual	41.623	4623	0.009		
	Sum	105.524	4629			

a. Dependent variable: BG(t)

b. Estimation variable: (Constant), BG(t-1)

c. Estimation variable: (Constant), BG(t-1), PR6

d. Estimation variable: (Constant), BG(t-1), PR6, AMT

e. Estimation variable: (Constant), BG(t-1), PR6, AMT,DEL

f. Estimation variable: (Constant), BG(t-1), PR7, AMT,DEL,PR5

g. Estimation variable: (Constant), BG(t-1), PR7, AMT,DEL,PR5, PR7

As shown in Table 1, we adopted a stepwise regression method, and the last bidding gain (BG(t-1)) was the first independent variable to enter the regression equation, earlier than several price reduction extents, remaining amount and free delivery or not^[17]. And the contribution of last bidding gain made to the change of R^2 is relatively largest among all variables, which indicated the last bidding gain (BG(t-1)) had the most important influence on current bidding gain earned at this time in 7-day auction mode. R^2 increased gradually with the entrance of other control variables, and reached largest at the 6th regression model. Through the ANOVA result shown in Table 2, the whole regression model is significant.

Table 3. Coefficient results

Model	1	Non-standardized coefficient		Standardized coefficient	t	Sig.
		B	Standard error	Beta		
1	(Constant)	0.258	0.006		40.686	0.000
	BR(t-1)	0.650	0.010	0.709	68.357	0.000
2	(Constant)	0.077	0.008		9.354	0.000
	BR(t-1)	0.444	0.011	0.485	40.492	0.000
	PR6	0.455	0.015	0.365	30.518	0.000
3	(Constant)	0.042	0.009		4.829	0.000
	BR(t-1)	0.428	0.011	0.467	39.291	0.000
	PR6	0.551	0.017	0.442	33.087	0.000
	AMT	0.000	0.000	-0.132	-12.193	0.000
4	(Constant)	0.053	0.009		6.003	0.000
	BR(t-1)	0.416	0.011	0.453	37.844	0.000
	PR6	0.564	0.017	0.453	33.827	0.000
	AMT	-9.7E-05	0.000	-0.108	-9.509	0.000
	DEL	-0.023	0.003	-0.070	-6.842	0.000
5	(Constant)	0.056	0.009		6.355	0.000
	BR(t-1)	0.417	0.011	0.455	38.016	0.000
	PR6	0.759	0.050	0.610	15.262	0.000
	AMT	0.000	0.000	-0.112	-9.871	0.000
	DEL	-0.028	0.004	-0.085	-7.855	0.000
	PR5	-0.227	0.055	-0.158	-4.163	0.000
6	(Constant)	-0.335	0.098		-3.431	0.001
	BR(t-1)	0.412	0.011	0.449	37.314	0.000
	PR6	0.806	0.051	0.648	15.801	0.000
	AMT	0.000	0.000	-0.122	-10.492	0.000
	DEL	-0.037	0.004	-0.112	-8.802	0.000
	PR5	-0.278	0.056	-0.193	-4.967	0.000
	PR7	0.412	0.103	0.047	4.017	0.000

In addition, Table 3 summarized the coefficients of both independent and control variables on dependent variable. The relevance of the consumers' current bidding gain with last bidding gain for the same kind of goods is 0.449 with very great significance. The 2nd, 3rd and 4th price reduction extent was excluded from the regression model at last, which might be connected with the habit that consumers mostly showed much more concern with higher price tier (lower bidding price). Previous four price tiers did not present so much attraction for consumers.

Through Table 3, we got the linear regression model of 7-day auction mode (1).

$$BG(i, j, t) = \beta_0 + 0.449 * BG(i, j, t-1) - 0.197 * PR_5(i, j, t) + 0.648 * PR_6(i, j, t) + 0.047 * PR_7(i, j, t) - 0.122 * AMT(i, j, t) - 0.112 * DEL(i, j) + \varepsilon(i, j, t) \quad (2)$$

We found that the last bidding gain had a positive effect on current bidding gain, which indicated that consumers could adjust current auction behavior positively by learning from last auction. The larger the last bidding gain was, the greater current bidding gain was.

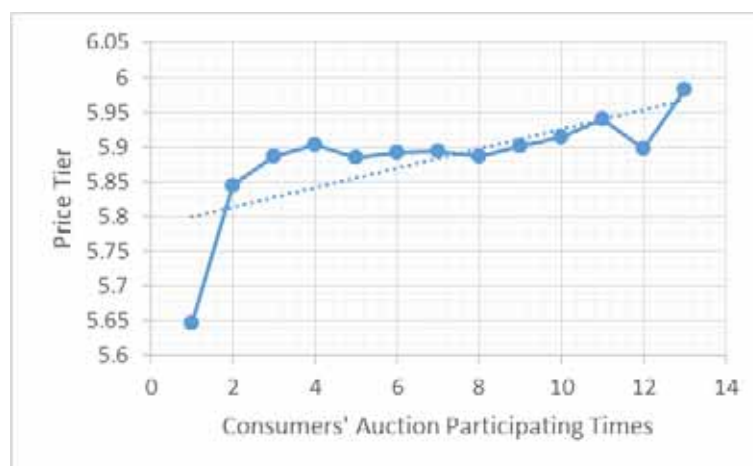


Figure 1. Change of price tier with the increase of auction participating times (7-day)

Next, we needed to focus on the changing trend of bidding price tier with the increase of consumers' auction participating times. Considering the crossover learning impact between different types of goods and the limited amount of auction data, we would not categorize the type of auction goods in the research process of descriptive statistics. Finally we could describe the relationship between bidding price tier and auction participating times shown in Figure 1. In 7-day auction, consumers tended to bid on a gradually higher price tier to obtain lower bargain price at the beginning. When they took part in the auction for more than 4 times, average price tier started to become steady. Fluctuation would occur when times reached more than 10, which might be relevant to exposure of remaining amount and other external pressure. As a whole, bidding price tier presented a upward trend with the increase of consumers' auction participating times in the 7-day auction.

4.3.2 15-minute auction mode

Table 4. Abstract of regression models

Model	R	R ²	Adjusted R ²	Standard estimation error	Change statistics				
					Change of R ²	Change of F	Degree of freedom 1	Degree of freedom 2	Change of Sig. F
1	.816 ^a	0.666	0.666	0.1160903	0.666	12822.825	1	6419	0.000
2	.873 ^b	0.761	0.761	0.0982156	0.095	2550.067	1	6418	0.000
3	.876 ^c	0.767	0.767	0.0970944	0.005	150.069	1	6417	0.000
4	.878 ^d	0.770	0.770	0.0963346	0.004	102.621	1	6416	0.000
5	.879 ^e	0.772	0.772	0.0960342	0.001	41.212	1	6415	0.000
6	.879 ^f	0.772	0.772	0.0959877	0.000	7.209	1	6414	0.007

a. Estimation variable: (Constant), BG(t-1)

d. Estimation variable: (Constant), BG(t-1), PR6, PR3, DEL

b. Estimation variable: (Constant), BG(t-1), PR6

e. Estimation variable: (Constant), BG(t-1), PR6, PR3, DEL, PR7

c. Estimation variable: (Constant), BG(t-1), PR6, PR3

f. Estimation variable: (Constant), BG(t-1), PR6, PR3, DEL, PR7, PR2

Table 5. ANOVA analysis result

Model		Sum of squares	Degree of freedom	Mean square	F	Sig.
1	Regression	172.813	1	172.813	12822.825	.000 ^b
	Residual	86.509	6419	0.013		
	Sum	259.321	6420			
2	Regression	197.411	2	98.706	10232.500	.000 ^c
	Residual	61.910	6418	0.010		
	Sum	259.321	6420			
3	Regression	198.826	3	66.275	7030.134	.000 ^d
	Residual	60.495	6417	0.009		
	Sum	259.321	6420			
4	Regression	199.779	4	49.945	5381.754	.000 ^e
	Residual	59.543	6416	0.009		
	Sum	259.321	6420			
5	Regression	200.159	5	40.032	4340.629	.000 ^f
	Residual	59.163	6415	0.009		
	Sum	259.321	6420			
6	Regression	200.225	6	33.371	3621.894	.000 ^g
	Residual	59.096	6414	0.009		
	Sum	259.321	6420			

a. Dependent variable: BG(t)

b. Estimation variable: (Constant), BG(t-1)

c. Estimation variable: (Constant), BG(t-1), PR6

d. Estimation variable: (Constant), BG(t-1), PR6, PR3

e. Estimation variable: (Constant), BG(t-1), PR6, PR3, DEL

f. Estimation variable: (Constant), BG(t-1), PR6, PR3, DEL, PR7

g. Estimation variable: (Constant), BG(t-1), PR6, PR3, DEL, PR7, PR2

As shown in Table 4 and 5, last bidding gain is still the first variable entering the regression equation, earlier than price reduction extents, remaining amount and delivery condition, indicating the last bidding gain was the greatest influencing factor in 15-minute auction mode similarly. The goodness of fit of this linear regression model can reach 0.772, a little higher than that of 7-day auction mode, and the regression result is strongly significant.

Table 6. Coefficient results

Model		Non-standardized coefficient		Standardized coefficient	t	Sig.
		B	Standard error	Beta		
1	(Constant)	0.145	0.005		29.428	0.000
	BR(t-1)	0.812	0.007	0.816	113.238	0.000
2	(Constant)	-0.005	0.005		-0.986	0.324
	BR(t-1)	0.514	0.008	0.517	60.721	0.000
	PR6	0.520	0.010	0.430	50.498	0.000
3	(Constant)	-0.049	0.006		-7.926	0.000
	BR(t-1)	0.503	0.008	0.506	59.744	0.000

Model		Non-standardized coefficient		Standardized coefficient	t	Sig.
		B	Standard error	Beta		
4	PR6	0.705	0.018	0.583	38.650	0.000
	PR3	-0.192	0.016	-0.164	-12.250	0.000
	(Constant)	-0.018	0.007		-2.604	0.009
	BR(t-1)	0.489	0.008	0.491	57.741	0.000
5	PR6	0.676	0.018	0.558	36.811	0.000
	PR3	-0.164	0.016	-0.140	-10.388	0.000
	DEL	-0.026	0.003	-0.064	-10.130	0.000
	(Constant)	0.387	0.063		6.097	0.000
6	BR(t-1)	0.475	0.009	0.478	54.540	0.000
	PR6	0.622	0.020	0.514	30.934	0.000
	PR3	-0.115	0.018	-0.098	-6.569	0.000
	DEL	-0.028	0.003	-0.070	-11.038	0.000
7	PR7	-0.392	0.061	-0.047	-6.420	0.000
	(Constant)	0.373	0.064		5.863	0.000
	BR(t-1)	0.476	0.009	0.479	54.632	0.000
	PR6	0.658	0.024	0.544	27.256	0.000
8	PR3	-0.253	0.054	-0.215	-4.661	0.000
	DEL	-0.030	0.003	-0.074	-11.362	0.000
	PR7	-0.374	0.061	-0.045	-6.103	0.000
	PR2	0.102	0.038	0.093	2.685	0.007

We can conclude that last bidding gain significantly affect current bidding gain positively in consumers' auction participating process since it passed both T-test and F-test. At the same time, 4th and 5th price reduction extent did not pass the test, which might related to the auction scenario where consumers paid more attention to the starting and ending stage of the auction due to the high-valued property of goods in 15-minute auction. The linear regression equation is:

$$BG(i, j, t) = \beta_0 + 0.479 * BG(i, j, t-1) + 0.093 * PR_2(i, j, t) - 0.215 * PR_3(i, j, t) + 0.544 * PR_6(i, j, t) - 0.045 * PR_7(i, j, t) - 0.074 * DEL(i, j) + \varepsilon(i, j, t) \quad (3)$$

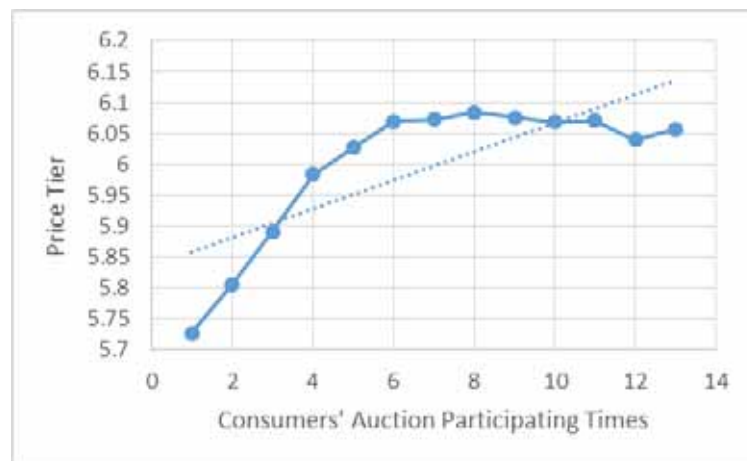


Figure 2. Change of price tier with the increase of auction participating times (15-minute)

Comparing Figure 1 and Figure 2, we found the very similar trend of bidding price tier with the increase of consumers' auction participating times in the 15-minute auction, which went upward firstly and then kept steady. However, the steady pattern appeared after 6 times, later than that in 7-day, 4 times.

5. DESCRIPTIVE ANALYSIS OF CONSUMERS' GRASP ACCURACY OF THE BIDDING TIMING

After the linear regression model, we also wanted to study the trend of consumers' grasp accuracy of the bidding timing as the number of auction participating times increase. Because the linear regression model can only prove that the consumer's last auction behavior significantly affect the current auction behavior, thus proving the existence of consumer learning effect, but it cannot clearly describe the consumer's chosen bidding price tier with the increasing number of auction times. Descriptive statistic method is primarily used for statistical analysis for a situation or data as a whole or a potential connection between them. Therefore, our research also needs to take advantage of this method in order to see whether consumers can grasp the bidding timing more and more accurately when times of participating auctions increase.

Descriptive analysis is divided into two parts: First, we studied the consumer's grasp accuracy of the final price tier (the price tier when all auction goods have been sold out) for the auction. Second, we did some descriptive research on the bidding time from the beginning of the auction at the final price tier.

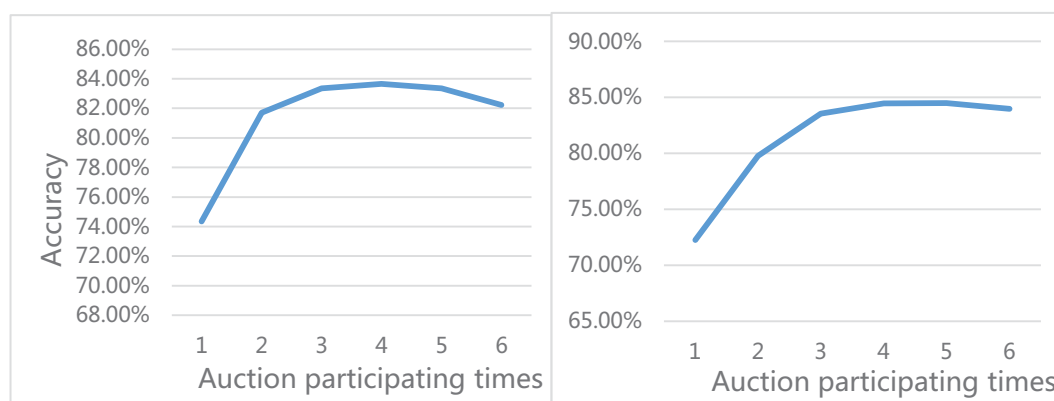


Figure 3. Consumer's grasp accuracy of the final price tier (7-day / 15-minute)

We can see that in both 2 auction modes, with the increase of the number of consumers' participating auctions, consumers' forecast or grasp for the final price tier in the auction was more and more accurate until it arrived at a relatively stable certain level. The curve also complied with the typical learning curve mentioned in the previous introduction. The above figure illustrates the existence of the consumers' learning effect in the price reduction auction of agricultural and sideline products from the view of the consumer's grasp accuracy of the final auction price tier.

Secondly, we were going to verify that with the increase in the number of auction participating times, consumers' bidding reaction time would be shortened or not. The question is raised because the number of goods which sold at a gradually reduced price is limited. This limit probably led to a large number of consumers to bid at the same price tier when auction went to a tipping point, and only the consumers with shorter reaction time was likely to bid the goods successfully.

After extracting the starting time and the consumers' bidding time at the final price tier, we calculated the time span between these two time points, which called as "auction reaction time span". We got the following results eventually.

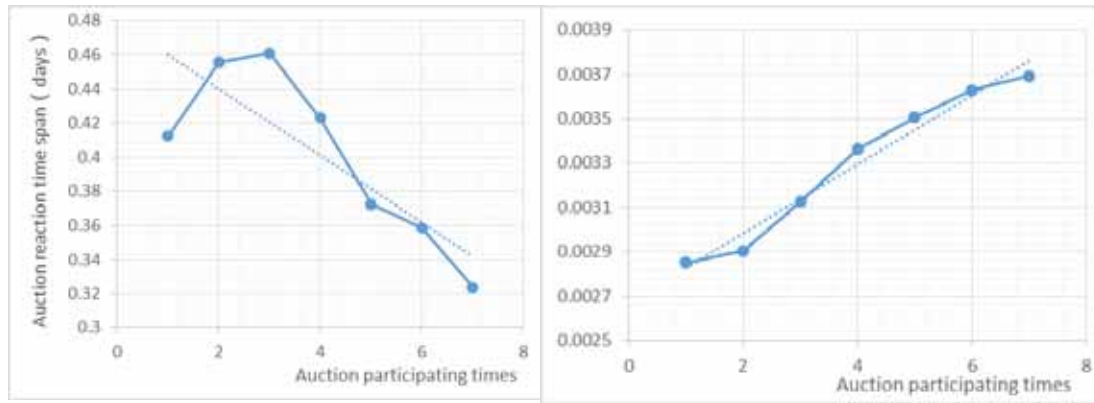


Figure 4. Consumer's auction reaction time span in the final price tier (7-day / 15-minute)

In the 7-day auction, with the increase in the number of auction participating times, the consumers' average auction reaction time span was indeed going down, that is, consumers grasped the bidding timing increasingly accurately. Although there's an increase from the first to the third participating time, the existence of such a fluctuation is mainly because of a certain degree of randomness in the first two times. The overall trend is declining, improving the learning effect on consumers' grasp of bidding time.

However, in the 15-minute auction, the consumer's auction reaction time span increases as the number of auction participating times increased. This result was somehow inconsistent with our guess and 7-day auction result. We attributed the possible reason to the much shorter auction interval compared with 7-day mode and consumer's personal bidding features. For example, auction reaction time spans of consumer A were 9,8 and 7 minutes and so on, and those of consumer B were 5,4 and 3 minutes. If A participated more than 3 auctions, and B did not, then the results will show the 4th average auction time is high. So we adopted an enhanced data processing method, which was to standardize reaction time span according to that of the first auction participating experience for each single consumer, as shown in formula (4) as below. $RTS_{i,t}$ means the reaction time span in the t^{th} participating experience of the i^{th} consumer. And we could draw Figure 4 to describe the relationship between processed reaction time span and auction participating times for 15-minute auction mode.

$$\sum_{i=1}^N \frac{RTS_{i,t} - RTS_{i,1}}{RTS_{i,1}}, t = 1, 2, \dots, n \quad (4)$$

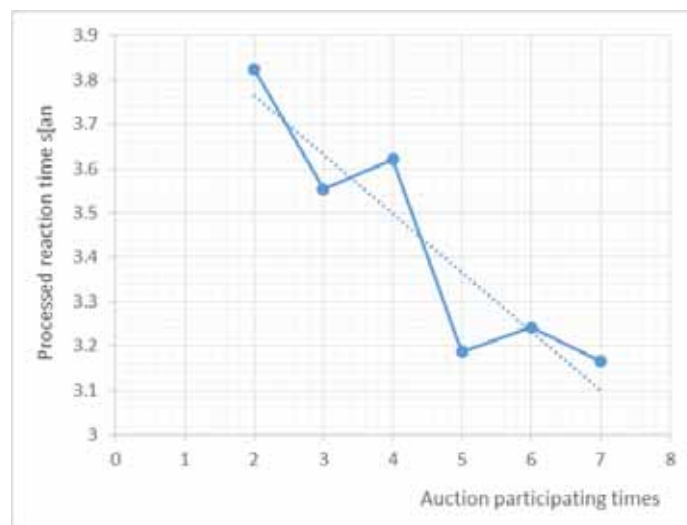


Figure 5. Processed consumer's auction reaction time span in the final price tier (15-minute)

From the above Figure 5, we can see that the consumer's auction time span decreases with the increase in the number of auction participating times, proving the existence of consumers' learning effect for grasping final bidding timing to some extent^[19].

6. Conclusion and Prospects

6.1 Research conclusions

This paper analyzed and summarized the collected data, and gives the relevant verification results for the two questions raised at the beginning of the study. Now, we summarize them as follows.

(1) We validated the presence of consumers' learning effects in both 7-day and 15-minute auction by linear regression model, which showed last bidding gain had significantly positive impact on current bidding gain with consideration of other control variable factors.

(2) We verified the consumer's auction timing grasp accuracy would be improved with the increase in the number of auction participating times. This conclusion can be reflect in more accurate grasp of the final price tier and shorter reaction time span when consumers take part in more auctions.

6.2 Research contributions

The contribution of this paper is divided into two parts. The first part is the theoretical contribution. This article introduced the learning effect to the online price reduction auction of agricultural and sideline products in the E-commerce platform, which fills the blank of relevant research field.

The second part is about practical contribution. This paper confirms the existence of consumer learning effect in the price reduction auction of agricultural and sideline products from two prospective, which especially reflected in the promoted bidding gain of consumers. This finding can potentially provide referential suggestions for price-reduction auction E-commerce platform to concern more about this learning effect in pricing strategy and auction scheme design.

6.3 Research prospects

(1) To improve the data information required for the study

Because the data information we collected is limited in consumers' personal information details, we hope to follow up and collect more complete data information for further study.

(2) To promote the experimental results of this study

Because this article is mainly carried out on the platform provided by Gongtianxia.com, so it is still necessary to find out whether our experimental results in this article is also applicable to other price-reduction auction.

ACKNOWLEDGEMENT

This research was supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China 18XN0002. New media enterprise organization and service change driven by technology: Based on the perspective of the communication theory, innovation theory and big data.

REFERENCE

- [1] Li Qingqing. Analysis of the price-reduction auctions of agricultural products in China[J]. Modern Business, 2007, (17): 169-170. (in Chinese)
- [2] Xu Yongguo. An overview of the economic theory of auction[J]. Economic Research Journal, 2002, (09): 84-91, 95.

- (in Chinese)
- [3] Zhu Miaohui, Qin Kaida. Price Anomaly and Winner's Strategic Behavior: Evidence from Rose Auctions[J]. Kunming University of Science and Technology, 2016, (02): 119-126. (in Chinese)
 - [4] Zhao Guoshun. Research on stock price trend prediction based on time series analysis[D]. Xiamen University, 2009. (in Chinese)
 - [5] Ye Chunming. Study on a new model of scheduling problems for manufacturing companies based on learning effect[J]. Enterprise Economy, 2015, (04): 5-10. (in Chinese)
 - [6] Hu Tianshi. Research on electronic commerce model of agricultural products in China[D]. Chinese Academy of Agricultural Sciences, 2005: 134. (in Chinese)
 - [7] Xing Hao. Research on the design of multi-item online auction mechanism based on Agent[D]. Beijing University of Chemical Technology, 2015. (in Chinese)
 - [8] Zhang Xuefei. The game scheduling problems with learning effect[D]. Qufu Normal University, 2015: 29. (in Chinese)
 - [9] Ding Meisheng, Guo Suwen. Economic analysis of price reduction auction[J]. Commercial Times, 2005, (18): 87-89, 32. (in Chinese)
 - [10] Goes P B, Karuga G G, Tripathi A K. Bidding behavior evolution in sequential auctions: characterization and analysis[J]. Mis Quarterly, 2012, 36(4):1021-1042.
 - [11] Jennifer N, Qiao Wei. How can the auction industry in Holland develop sustainably[J]. China Flowers & Horticulture, 2006. (in Chinese)
 - [12] Zhang Hailiang. Analysis and research on the operation of Chinese flower auction market[D]. Kunming University of Science and Technology, 2007. (in Chinese)
 - [13] Qin Kaida. Study on the price fluctuation in the auction market of fresh agricultural products[D]. Kunming University of Science and Technology, 2012. (in Chinese)
 - [14] Zheng Hanxiong, K. Y. Goh, and K. W. Huang. "Evaluating the Impacts of Auction Bidding Restrictions on Consumer Surplus and Behaviors - An Empirical Study of Penny Auctions." International Conference on Information Systems, Icis 2011, Shanghai, China, December DBLP, 2011:3546-3556.
 - [15] Ma Xichen. Analysis of the promotion effect based on the theory of consumers' learning[D]. Shanxi University of Finance and Economics, 2014. (in Chinese)
 - [16] Li Xingmei, Zhang Qian, Zhao Xinchao, Qi Jianxun. Research on the project network model based on the double effect of learning effect and time conversion constraint[J]. Journal of Industrial Engineering and Engineering Management, 2015, (02): 63-69. (in Chinese)
 - [17] Hu Ziwei. The application of mathematical statistics in data analysis[J]. Economic & Trade, 2016. (in Chinese)
 - [18] She Lingling. Consumers' psychological factors effect on the result of two different sales promotion tools[D]. Huazhong Agricultural University, 2010: 84. (in Chinese)
 - [19] Liu Yang, Zhu Hui. Research on consumer behavior under modern electronic commerce[J]. China Journal of Commerce, 2010, 25:133-134. (in Chinese)
 - [20] Katok E, Kwasnica A M. Time is money: The effect of clock speed on seller's revenue in Dutch auctions[J]. Experimental Economics, 2008, 11(4):344-357.

Consumer Coupon Redemption Behavior Prediction on B2C E-commerce

Wenmin Qi¹, Li Li², Ruibo Yao^{3*}

^{1,2} School of Economics and Management, Nanjing University of Science and Technology, Nanjing, 210094, China

³ Focus Technology Co., Ltd., Nanjing, 210061, China

Abstract: How to recognize the tendency of the coupons among the users who receive the coupons and then send the coupon reminder to improve the coupon redemption rate and reduce the marketing cost has become an important issue in the coupon decision-making process. Based on the log data and transaction data in enterprise database, this study combined with the demographics, past purchasing behavior, past coupon usage behavior and the visiting behavior during the coupon validity period to construct the e-coupon redemption behavior prediction model. The model is constructed to help e-commerce enterprises identify the target users who have the coupon proneness after the coupons are issued, so as to send coupon reminders in time and enhance the effectiveness of coupon marketing.

Keywords: E-commerce, behavioral analysis, logistic regression, coupon redemption

1. RESEARCH QUESTION

We propose a prediction model of e-coupon redemption behavioral based on user behavior data and transaction data. The model combines consumer demographics, past purchasing behavior, past coupon usage behavior and the visiting behavior features during the validity period.

2. MAJOR RESEARCH FINDINGS

2.1 Model

We intends to adopt the logistic model as the modeling basis to predict consumer e-coupon redemption:

$$P(Y = 1|x) = \frac{1}{1 + e^{-\beta_0 - \sum_{i=1}^k \beta_i x_i}} \quad (1)$$

Where $P(Y=1|x)$ represents the probability that the consumer who has an access behavior during the coupon validity uses the coupon, x_i is the extracted i characteristic variables based on the factors of affecting consumer coupon usage after the combined analysis of literature and experimental data.

2.2 Results

In order to optimize the model, the feature variables were screened by a stepwise optimization method. The results of model fitting after feature selection are shown in Table 2.

Table 2. Original model results

	β_0	GENDER	AGE	N_GETCO P	N_USECOP	COP_AMT _TOAL	PRICE_MA X	N_VISIT_P AGE
Estimate	-2.183037 ***	-0.376732*	0.018942*	-0.655871 ***	5.554697** *	0.043365** *	0.02918**	3.692188** *

In terms of demographics, we can find that women are more likely to use coupons than men, and the older are more likely to use coupons than the younger. This is consistent with existing research results^[1].

By analyzing the negative correlation between consumer history coupon acquisition and consumer

* Corresponding author. Email: lily691111@126.com(Li Li), woshiqiwenmin@126.com(Wenmin Qi)

e-coupon redemption, we can hold the opinion that consumer perceptions and attitudes toward coupons can be used as intermediary variables to explore the relationship between the historical coupon acquisition and coupon redemption behavior. Besides, the number of consumer historical coupons usage and the consumer historical monthly average amount of coupon usage have a positive correlation with consumer e-coupon redemption.

In terms of past purchasing behavior, consumers who spend large amounts of money in a single consumption often have a relatively high purchasing power and are therefore more likely to use coupons. Consumers who spend large amounts of money in a single consumption often have a relatively high purchasing power and are therefore more likely to use coupons.

The number of pages visited during the coupon validity period is positively correlated with the consumer e-coupon usage, indicating that the consumers having more pages to visit during the coupon validity period have a relatively higher interest in products and are more possible to use the coupons.

It can be concluded that the e-coupon redemption behavior prediction model is:

$$P(Y=1|x) = \frac{1}{1+e^{g(x)}} \quad (2)$$

$$g(x) = -2.183037 - 0.376732x_1 + 0.018942x_2 - 0.655871x_3 + 5.554697x_4 + 0.043365x_5 + 0.029180x_6 + 3.692188x_7 \quad (3)$$

Where x_1 represents the gender of the user, x_2 represents the age, x_3 represents the monthly average of the history coupon x_4 represents the monthly average usage amount of historical coupons, x_5 represents the average monthly usage amount of historical coupons, x_6 represents the historical purchase maximum price, and x_7 represents the total number of pages visited after standardization.

3. CONCLUSIONS

Based on log data and transactional data, etc., we construct a e-coupon redemption behavior prediction model from the level of individual users. The study shows that female consumers are more likely to use coupons than male ones, older consumers are more likely to use coupons than their younger ones. We discuss the consumer's historical behavior and access behavior, which are seldom mentioned in the existing literature. Results found that the number of historical coupons obtained is negatively correlated with the user's coupon redemption, while the quantity and amount of consumer historical coupons used, the total amount of discounts, the highest historical purchase price and the number of pages visited during the coupon validity were positively correlated with the user e-coupon usage behavior. This research is of guiding significance for the study using the access data and other data to predict their behavioral intentions.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71771122 and 71271115.

REFERENCES

- [1] Inman, C. W. J. J. (2008). Do shoppers like electronic coupons?: a panel data analysis. *Journal of Retailing*, 84(3), 297-307.
- [2] Khajehzadeh, S., Oppewal, H., & Tojib, D. (2014). Consumer responses to mobile coupons: the roles of shopping motivation and regulatory fit ☆. *Journal of Business Research*, 67(11), 2447-2455.

Measurement on Short-term Effect and Purchase Conversion

Mechanism of Online Advertising

Weiwei Liao¹, Li Li², Ruibo Yao^{3}*

^{1,2}School of Economics and Management, Nanjing University of Science and Technology, Nanjing, 210094, China

³Focus Technology Co., Ltd., Nanjing, 210061, China

Abstract: With the development of multimedia, the form of online advertising has become more diversified. In the increasingly competitive environment between e-commerce enterprises, the proportion of online advertising investment in total advertising investment continues to grow. However, the operating costs of enterprises are limited, in order to maximize the effectiveness of operating costs, evaluating the effect of various forms of online advertising effectively to determine the scientific online advertising budget allocation is crucial. Based on the enterprise's Web log data, we extract the characteristic variables that represent consumer's behavior including click behavior, visit behavior and purchase behavior. On the one hand, using multivariate linear regression method to assess the short-term effect of three forms of online advertising in an overall level. On the other hand, we use logistic regression models to investigate the impact of various consumer behaviors on the purchase conversion of consumers who enter the website from different advertising channels.

Keywords: Online advertising, Short-term effects, Purchase conversion

1. INTRODUCTION

The rapid development of Internet technology and the improvement of social informatization level provide a favorable environment for the development of enterprises. Under the trend of accelerating the integration of traditional media and new media, enterprises have realized the significance of Internet in the marketing system, and gradually shift the focus of marketing to the Internet. "2017 Annual Report on Online Advertising Market of China" released by IResearch indicates that revenue scale of online advertising increased from 33.71 billion RMB in 2010 to 290.27 billion RMB in 2016 with an annual growth rate of 43.1% ^[1]. It can be seen from the report that the revenue of online advertising is far higher than that of traditional advertising, consequently, it is more likely for enterprises to invest online advertising and expand input scale in the future.

However, the cost of online advertising keeps continuous rising, which has become the largest expenditure of enterprises' operating expenses. Furthermore, diversification of forms of online advertising has exacerbated the rise of advertising costs. Therefore, it is crucial for enterprises to figure out the effectiveness of different forms of online advertising so that they can determine scientific budget allocation to get the maximum profits. Besides, based on online environment, enterprises are capable to collect various user behaviors. For consumers who enter the website from different advertising sources, the impact of their behavior on purchase are different. The enlightenment gained from the analysis of consumer behavior is more valuable to marketing management.

2. LITERATURE REVIEW

2.1 Online advertising

In recent years, a large amount of the existing research focus on influential factors of specific advertising form's effectiveness. For example, Rutz and Trusov explained how position and content of paid search

* Corresponding author. Email: lily691111@126.com (Li Li)

advertising influence its effectiveness^[2]. Lambrecht and Tucker found that the effectiveness of retargeting depends on consumers' product preference^[3]. These studies neglect to evaluate the effects of various forms of online advertising in an overall level and thus, are of limited value in enterprises' budget allocation decisions. Nottorf indicated that consumers are often influenced by more than one specific type of online advertising when they click advertising links to enter the website^[4]. Consumers are usually exposed to several online advertisements before purchase. Then, what contribution will consumers' click behavior of various online advertisements have on following purchase? An overall assessment about the effects of various online advertising is needed.

Previous studies have shown that advertising has immediate and carryover effects on sales. The short-term effects of online advertising refer to a consumer's reaction to advertising exposure within a short period of time, such as clicks, purchases^[5]. Bass and Clarke said that the largest impact of advertising may not necessarily be in the initial moment but work after a certain time^[6], which emphasized the significance of carryover effects. Many studies worked on how to treat time lag, which only partially reflected short-term effects of advertising. Breuer and Brettel evaluated short- and long-term effects of display, e-mail, price comparison advertising using multiple linear regression models combined with time lag^[5]. Haan *et al.* compared the long-term effectiveness of nine forms of online advertising by establishing a structure-vector autoregressive model and a restrictive impulse response, but their results did not indicate the evidence of time-varying parameters^[7]. However, for enterprises, the short-term effects of online advertising can stimulate consumer purchases in a short period of time and bring sales revenue to enterprises more directly. This is undoubtedly a more attractive research topic for enterprises.

2.2 Short-term effects of online advertising

In recent years, researchers have noticed to evaluate effects of online advertising based on clickstream data. Yet, for commercial websites, clickstream data usually involves commercial secrets, thence studies may be limited by data availability and the findings in this area are still rare. Concluding from the existing research results, the evaluation of short-term effect mainly starts from click-through rate and conversion rate.

Click-through rate is the percentage of advertising clicks to total impressions. Among them, last-click model has been widely used in actual business management. However, it does not consider consumer's behavior and neglect other information before consumer's last click on online advertising. Although click-through rate shows an insufficient description, it still reflects the effectiveness of advertising to some extent. Some scholars also point out that click-through rate can be used as an intermediary variable to reflect other influential factors. Some studies explored impact of consumer's behavior on click possibility under different online advertising exposures. Chatterjee *et al.* built a measurement model of click probability by means of logistic regression to discuss the impact of advertising exposure period and impressions on click probability^[8]. Nottorf investigated consumer click possibilities for display, redirect, paid search and video advertising using a binary logit with Bayesian mixture of normal based on similar consumer advertising browsing and clicking data^[4].

Further than click-through rate, some scholars start to consider online advertising effects from the perspective of purchase intention^[7]. Manchanda *et al.* measured impact of display advertising on purchase probability by establishing a survival model^[9]. Summers *et al.* compared differences between effectiveness of targeted advertising based on online and offline environment by purchase intention. Compared with click-through rate, purchase intention as evaluation index of online advertising effect is closer to actual demand of enterprises^[10].

Conversion rate represents the ratio of consumer conversion from potential customer to newers under the influence of online advertising. Rutz and Trusov designed a two-stage user transformation model to describe the

process from click to purchase^[2]. Similar to Rutz, Montgomery *et al.* used purchase conversion rate to further convert clicks into purchase possibilities and tried several models including multiple predictive model to fit consumer's browsing and clicking data^[11]. Xu *et al.* studied dynamic interaction between online advertising clicks by using translation probability to discuss effects of online advertising^[12]. Among these studies, the dataset they used are simple statistics of page browsing and advertising impressions before clicks. Few studies have examined short-term effects of online advertising from following visit behavior after consumers enter the website.

Considering that consumers' clicks on online advertising to enter the site is triggered by current advertising, thus subsequent behavior is what we believe to be affected by short-term effects of the advertising. And reviewing existing research, the question of connection between consumer behavior and purchase under the influence of different types of online advertising has not been discussed in depth. Therefore, based on data of a third-party insurance agent network platform, we chose to examine short-term effects of various types of online advertising and set up evaluation model based on purchase intention to construct the relationship of consumers' visit behavior and purchase behavior after they enter the site through online advertising. We concentrated on three types of online advertising—paid search, e-mail and short message service advertising, measured short-term effects of each advertising and figured out the specific contribution of various consumer behavior to purchase.

3. MODEL

The mainstream types of online advertising include display, paid search, e-mail and short message service advertising. As display advertising is increasingly perceived as disruptive to users, enterprises have curtailed their display advertising investment. Accordingly, our modeling work of short-term effects of online advertising and purchase conversion mechanisms is centered around paid search, e-mail and short message service advertising.

3.1 Short-term effects measurement

Based on previous studies, short-term effects of online advertising refer to behavior reaction of consumers in a short period of time after advertising exposures, which includes following click and purchase behavior. Considered actual demands of enterprises, direct response of advertising effects that enterprises can notice is the sales growth. In other words, the most intuitive reflection of short-term effects is on consumer purchase intention.

Earlier, Srinivasan and Weir proposed a direct aggregation approach that examines the advertising effect by establishing the relationship between advertising and sales^[13], as follows:

$$S_t = a + bAdv_t^* + e_t \quad (1)$$

where S_t represents sales volume on day t , Adv_t^* represents advertising stock of each advertising on day t and the estimated parameter b indicates short-term effects of each advertising. The model takes advertising stock as an independent variable and considers the contributions of each advertising inputs to sales volume. The contributions are interpreted as short-term effects of each advertising. Since the model was proposed under offline environment, it was hard to track the profit that each consumer made under the impact of advertising. Therefore, it takes sales volume of the entire enterprise as dependent variable, which results that the sales income can't be differentiated from advertising and non-advertising caused. Currently, enterprises can track consumers' visit tours based on Web log so that effects of online advertising can be examined from consumer-level by modeling each consumer's behavior record. Studies have shown that user engagement, such as interaction with advertising, can lead to stronger predictive effects. Compared with advertising stock,

consumers' clicks on advertising represent their willingness on advertising and predict possible purchase better, besides, it also confirms that subsequent purchase behavior is affected by advertising. Consequently, we try to establish connection between advertising clicks and purchase in short order. We consider consumer's clicks on each online advertising as independent variables and purchases volume as dependent variable to investigate short-term effects, where short-term effects are interpreted as impact of advertising click on purchase in short order. Proposed model is shown as equation (2):

$$S_{it} = \alpha + \beta_1 SEA_{it} + \beta_2 SMS_{it} + \beta_3 EMA_{it} + e_{it} \quad (2)$$

where S_{it} means purchases volume of user i on day t ; SEA_{it} , SMS_{it} , and EMA_{it} respectively represents clicks on paid search, short message service and e-mail advertising of user i on day t ; β_1 , β_2 , β_3 are parameters to be estimated, which represents short-term effects of paid search, short message service and e-mail advertising.

3.2 Purchase conversion mechanism

In addition to an overall assessment of short-term effects of three types of online advertising, we attempt to construct purchase conversion mechanisms for various types of online advertising, which discusses the impact of subsequent visit behavior after consumers enter the website through online advertising on purchase. So, we consider a logistic regression model is suitable with whether consumer purchases or not as dependent variable, and a plurality of characteristic variables describing consumer's visit behavior as independent variables.

Reviewing previous studies, we extracted characteristic variables that describe customers' behavior from their visit tours. In marketing management, website purchase funnel is often used to describe a consumer's journey. Website purchase funnel mainly consists of four stages, first, a consumer usually lands on home page or product pages through advertising links. And then, he searches the website according to his need and views various product pages. In this process, he stores products that are possible to purchase in the shopping basket. After this, he would like to compare the products and finally make a purchase. By studying the literature describing this process systematically and combining actual experience, we summarize consumers' behavior that may influence purchase behavior during their stay on the site, including search, visit and interest behavior. Considering the functional design of the website and the actual data, eight characteristic variables of consumer behavior were extracted, including number of filtering, number of viewing pages, average time of viewing each page, number of viewing product pages, average time of viewing each product page, number of viewing "School" pages, number of viewing "Topic" pages, number of viewing "Toptag" pages. The specific model proposed is as equation (3):

$$h(Y^*) = p(\text{purchase} | Y^*) = \text{Logit}(\beta Y^* + \varepsilon) = [1 + \exp(\alpha + \beta_1 y_1 + \beta_2 y_2 + \dots + \beta_8 y_8)]^{-1} \quad (3)$$

where $h(Y^*)$ represents purchase flag; Y^* represents behavior characteristic variables of consumers from different advertising sources, y_i represents behavior characteristics; β_i is the parameter to be estimated, is also the influence coefficient, which is used to judge the impact of each user behavior on the final purchase behavior.

4. EMPIRICAL STUDY

4.1 Data pretreatment

The data we used is the Web log data of a third-party insurance agent network platform in Nanjing. The time period ranges from January 1st, 2017 to April 30th, 2017, and the total amount of original data exceeds 15 million user records. Web log data records consumer's click behavior, where each record represents a click

behavior of the consumer. We aim to construct connection between consumer's visit behavior and purchase behavior. User identification is required to decide which records belong to the same person. After this, session segmentation helps us to figure out consumer's behavioral sequence from entering the site to leaving. Finally, we extract characteristic variables that we proposed above from each session of each user. The data preparation process is as follows:

(1) Data acquisition: We obtain single-day PC terminal data from Web log and extract research related field information, such as IP, visit date, visit time, agent, cookie, etc.

(2) User identification: According to the extracted field information, we distinguish the users based on cookie, IP and agent field. There is a need to illustrate that cookie represents the web cache information generated by user visit behavior, while agent represents the user's web browser.

(3) Session segmentation: We divide session by time threshold set between two adjacent access requests, and the threshold value is set to 1800 seconds. What needs to be explained here is that since the first model is an overall assessment, we believe that fetching variables on a daily basis is sufficient.

(4) Feature extraction: We extract users' behavior characteristics of each session that is mentioned in the theoretical models. The specific variables are shown in table 1.

Table 1. Characteristic variables description

Variable	Name	Variable	Name
S_{it}	Purchases volume	y_4	Number of viewing product pages
SEA_{it}	Paid search advertising clicks	y_5	Average time of viewing each product page
SMS_{it}	Short message service advertising clicks	y_6	Number of viewing "Study" pages
EMA_{it}	E-mail advertising clicks	y_7	Number of viewing "Topic" pages
y_1	Number of filtering	y_8	Number of viewing "Toptag" pages
y_2	Number of viewing pages	$h(Y^*)$	Purchase flag
y_3	Average time of viewing each page		

4.2 RESULT

4.2.1 Short-term effects measurement

Based on the theoretical model (2), we extracted the clicks of paid search, short message service, e-mail advertising and purchases volume, and constructed the relationship between the clicks of each type of online advertising and purchases. The descriptive statistics and estimated parameters are respectively shown in table 2 and table 3.

Table 2. Descriptive statistics

	S_{it} (Purchases volum)	SEA_{it} (Paid search advertising clicks)	SMS_{it} (Short message service advertising clicks)	EMA_{it} (E-mail advertising clicks)
Sum	25404	610848	174	573

Table3. Parameter estimation of short-term effects

Parameter	Estimate	t value	Pr(> t)
α (Intercept)	-0.0063	-3.3790	0.0007
β_1 (Paid search advertising)	0.0463	38.1550	0.0000
β_2 (Short message service advertising)	0.0564	1.2840	0.1992
β_3 (E-mail advertising)	0.0811	3.5320	0.0004

From Table 3, we can see that P values of β_1 and β_3 are less than 0.05. It means clicks on paid search and e-mail advertising have a significant impact on purchase. Therefore, our resulting model is shown as equation (4).

$$S_{it} = -0.063 + 0.0463 \times SEA_{it} + 0.0811 \times EMA_{it} \quad (4)$$

Based on the estimation of β_1 and β_3 (0.0463: 0.0811), we find that impact of both paid search and email advertising on purchase are positive and the influence of email advertising is stronger than paid search advertising. We consider paid search advertising to be triggered by consumers' spontaneous searching behavior, which is driven by their interests, and therefore information provided by paid search advertising is related to consumers' preferences and current needs. E-mail advertising focuses on inferring consumers' preferences and distributing information to specific groups based on their visit or purchase history. In other words, paid search advertising is result of consumers' searching when they are aware of their own demands, thus its role is to stimulate consumers' willingness to make a purchase. While e-mail advertising can stimulate consumers to generate demand, and based on past visits and purchases, once existing consumers generate interests, they will be more prone to purchase. Therefore, we believe that e-mail advertising is more incentive to purchase than paid search advertising, and the result also indicates that the enterprise has done a good job of targeting e-mail advertising.

In term of short message service advertising, its impact of short message service advertising clicks on purchase is not significant. Combining Table 2, we can see that we have only got 174 total clicks of short message service advertising. This may be due to the limited investment of short message service advertising so that the clicks are few; otherwise, the distribution of the enterprise does not have a strong pertinence, that is, the content of short message service advertising does not match the preference of whom received the message. The enterprise spent advertising investment did not receive the expected effect, resulting in a waste of funds.

4.2.2 Purchase conversion mechanism

According to the theoretical model (3), we extract user behavior variables and try to respectively establish purchase conversion mechanism of paid search, email and short message service advertising. When we tried to construct purchase conversion mechanism of short message service advertising, all variables were insignificant, and considering its insignificant clicks on purchase, we will not discuss this purchase conversion mechanism.

1) Paid search advertising

We get dataset after data pretreatment and model records of consumers who enter by paid search advertising links to construct purchase conversion mechanism, the estimated parameters are shown in table 4.

Table 4. Parameter estimation of paid search advertising

Parameter	Estimate	t value	Pr(> t)
α (Intercept)	-4.9600	-208.760	0.0000
β_1 (Number of filtering)	-0.5185	-27.389	0.0000
β_2 (Number of viewing pages)	0.3943	143.241	0.0000
β_3 (Average time of viewing each page)	0.0004	5.018	0.0000
β_4 (Number of viewing product pages)	-0.3623	-59.622	0.0000
β_5 (Average time of viewing each product page)	-0.0004	2.050	0.0403
β_6 (Number of viewing "Study" pages)	-2.392	-26.257	0.0000
β_7 (Number of viewing "Topic" pages)	-0.6628	-2.911	0.0036
β_8 (Number of viewing "Toptag" pages)	-1.5740	-15.901	0.0000

As shown in Table 4, all parameters are significant, which means that all the behavioral characteristics have a significant impact on purchase conversion. Therefore, the resulting model is as equation (5):

$$h(Y^{SEA}) = \left[1 + \exp \left(\frac{-4.9600 - 0.5185y_1 + 0.3943y_2 + 0.0004y_3 - 0.3623y_4}{-0.0004y_5 - 2.3920y_6 - 0.6628y_7 - 1.5740y_8} \right) \right]^{-1} \quad (5)$$

The parameters of number of viewing pages and average time of viewing each page (0.3943: 0.0004) are positive, which means that with more pages and longer stay that consumer viewing, purchase possibility will get higher, and the impact of former is much stronger than the later. Consumers who click on paid search advertising to enter a site may be unfamiliar with the site, they are in the process of exploring and searching for products, and therefore, the increase in the number of viewing pages and average time of viewing each page means that the functional design or product of the website causes consumer's interest, the purchase possibility will also rise.

In addition, all other variables have negative parameters, which means that all the other behaviors will have a negative impact on purchase intention.

The increase in the number of filtering means that the consumer constantly changes the filter conditions to search the product that meets their expectations, that is, most products may not satisfy the consumer, leading him to adjust priority of their demands to get a better choice. For example, a consumer needs to purchase travel insurance with a selected brand of China Ping An and the security coverage includes flight delays and money robberies. When he finds that there is no product that fits all the requirements, he has to re-determine filter conditions and may choose other brands to check the results. As a result, as the number of adjustments continues to increase, the consumer is less likely to find the desired product, and purchase probability falls.

As for number of viewing product pages and average time of viewing each product page on purchase, their impacts are just in contrast to number of viewing pages and average time of viewing each page, which we think is not contradictory. For consumers who land on the website by paid search advertising links, they are exploring, thus the increase in number of viewing pages is more likely caused by other functional pages rather than product pages, which represents a rise of consumers' interests to the website itself. It will promote purchase intention. While an increase in number of viewing product pages or average time of viewing each product page indicates that consumers are more hesitate to make a purchase. Therefore, it will lead to a decrease in purchase possibility.

Besides, the featured sections of the website are generally avenues for consumers seeking help when they doubt about some product or service, so the more pages viewed, the less likely a consumer's question will be resolved and the odds of a purchase occurring get lower.

2) E-mail advertising

We identify the records of consumers who enter the website through e-mail advertising links to establish the purchase conversion mechanism of e-mail advertising, where we find the number of viewing "Topic" pages among all records is null. Thus, we do not discuss this variable. The estimated parameters are shown in table 5.

Table 5. Parameter estimation of e-mail advertising

Parameter	Estimate	t value	Pr(> t)
α (Intercept)	-4.7250	-8.067	0.0000
β_1 (Number of filtering)	0.1175	0.495	0.6218
β_2 (Number of viewing pages)	0.1418	4.894	0.0000
β_3 (Average time of viewing each page)	0.0015	0.594	0.5525
β_4 (Number of viewing product pages)	0.2431	2.009	0.0445
β_5 (Average time of viewing each product page)	-0.0132	-1.446	0.1481
β_6 (Number of viewing "Study" pages)	-0.1629	-0.010	0.9922
β_7 (Number of viewing "Topic" pages)	0.6643	0.000	0.9999

It can be seen from table 5 that except β_3 and β_5 , all the other parameters are not significant, which means that for consumers who click on the email advertising to enter the website, only the number of viewing pages and product pages are influential to purchase intention. Therefore, the resulting model is as equation (6):

$$h(Y^{EM}) = [1 + \exp(-4.7250 + 0.1418x_2 + 0.2431x_4)]^{-1} \quad (6)$$

Both number of viewing pages and product pages (0.1418: 0.2431) are positive, which means that consumers will be more likely to purchase while browsing more pages, and the impact of the later than the former. Here, we consider that customers who click on an e-mail advertising link to enter the site often click the advertising caused by their interests to merchant activities or specific product information. And based on their past experience of using the website, they prefer to check specific products. As a result, the increase of viewing pages is mainly caused by product pages. In other words, an increase in product page views means that consumers are increasingly interested in the content of the e-mail advertising links, and the likelihood of their purchase conversion rises.

In general, the impact of consumers' visit behavior on purchase conversion depends on whether the behavior can increase the probability that consumers will find the product that meets his expectation. The higher the probability, the more likely to make a purchase. For enterprises, the above purchase conversion mechanisms can provide some advice for the website design and function adjustment. For example, in the real-time recommendation and coupon distribution, if a consumer clicks through an email advertising, the managers can set an appropriate amount of coupons to promote the deal when the browsing time exceeds a certain threshold.

5. CONCLUSION

In order to explore short-term effects of online advertising further, we start from the perspective of purchase conversion to consider and use Web log data provided by a third-party insurance agency platform in Nanjing to attempt our proposed models. We first analyze short-term effects of three types of online advertising - paid search, short message service and e-mail advertising in an overall level. Besides, we put forward purchase conversion mechanisms of each online advertising to analyze the influence of consumers' behavior on purchase intention for consumers from different advertising resources. Our results help enterprises obtain marketing implications for consumers from different advertising sources. Due to the limited data provided by the platform, we only discussed short-term effects of three types online advertising. As the type of online advertising evolves, more types of online advertising may be considered in the future. And following study can extent our model, add more detailed description of the consumers' behavior characteristics to support enterprises making marketing strategy decisions.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71771122 and 71271115.

REFERENCES

- [1] IResearch. (2017-04-27). 2017 Annual Report on Online Advertising Market of China. <http://www.iresearch.com.cn/report/2980.html>.
- [2] Rutz O J, Trusov M. (2011). Zooming in on Paid Search Ads--A Consumer-Level Model Calibrated on Aggregated Data. *Marketing Science*, 30(5):789-800.
- [3] Lambrecht A, Tucker C E. (2013). When does Retargeting Work? Information Specificity in Online Advertising. *Journal of Marketing Research Jmr*, 50(5):561-576.

- [4] Nottorf F. (2014). Modeling the clickstream across multiple online advertising channels using a binary logit with Bayesian mixture of normals. *Electronic Commerce Research and Applications*, 13(1):45-55.
- [5] Breuer R, Brettel M. (2012). Short- and Long-term Effects of Online Advertising: Differences between New and Existing Customers. *Journal of Interactive Marketing*, 26(3):155-166.
- [6] Bass F M, Clarke D G. (1972). Testing Distributed Lag Models of Advertising Effect. *Journal of Marketing Research*, 9(3):298-308.
- [7] Haan E D, Wiesel T, Pauwels K. (2016). The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework. *International Journal of Research in Marketing*, 33(3):491-507.
- [8] Chatterjee P, Hoffman D L, Novak T P. (2003). Modeling the Clickstream: Implications for Web-Based Advertising Efforts. *Marketing Science*, 22(4):520-541.
- [9] Manchanda P, Dubé J P, Goh K Y, Chintagunta P K. (2006). The Effect of Banner Advertising on Internet Purchasing. *Journal of Marketing Research*, 43(1):98-108.
- [10] Summers C A, Smith R W, Reczek R W. (2016). An Audience of One: Behaviorally Targeted Ads as Implied Social Labels. *Journal of Consumer Research*, 43(1): ucw012.
- [11] Montgomery A L, Li S, Srinivasan K, Liechty J C. (2004). Modeling Online Browsing and Path Analysis Using Clickstream Data. *Marketing Science*, 23(4):579-595.
- [12] Xu L, Duan J A, Whinston A. (2014). Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion. *Management Science*, 60(6): págs. 1393-1412.
- [13] Srinivasan V, Weir H A. (1988). A Direct Aggregation Approach to Inferring Microparameters of the Koyck Advertising-Sales Relationship from Macro Data. *Journal of Marketing Research*, 25(2):145-156.

How do Patients with Chronic Diseases Make Usage Decisions

Regarding Mobile Health Monitoring Services?

Fanbo Meng^{1,2}, Xitong Guo^{1}, Kee-hung Lai², Xinli Zhao¹*

¹School of Management, Harbin Institute of Technology, Harbin, China.

²Faculty of Business, The Hong Kong Polytechnic University, Kowloon, Hong Kong.

Abstract: The increasing population of patients with chronic diseases generates great challenge of chronic disease management. The occurrence of mobile health monitoring service is beneficial to chronic disease prevention and health promotion. The objective of this study is to investigate how the patients with chronic diseases make usage decisions on mobile health monitoring service. A survey comprising 261 subjects were conducted to validate the research model and proposed hypotheses. The results revealed that health severity positively influences mobile health monitoring service use intention, while negative health emotions do not. Health uncertainty avoidance strengthens the effect of health severity but weakens the effect of negative health emotions on mobile health monitoring service use intention. Limitations and implications for research and practice are discussed.

Keywords: Mobile health monitoring service, chronic diseases, health severity, negative health emotions, health uncertainty avoidance.

1. INTRODUCTION

With the rapid approach of the aging society, patients with chronic diseases make up a considerable proportion of the global population ^[1]. According to a recent report by the World Health Organization, for instance, chronic diseases account for 60% of all deaths worldwide ^[2] and 85% of those in China ^[1]. Expenses incurred for chronic diseases comprise about 75% of the total healthcare expenses in the United States ^[3]. Despite the serious impacts resulting from chronic diseases, there are no effective measures to alleviate the patients' health problems. The mobile health monitoring service (MMS) is a widely-used approach for chronic disease management ^[4]. However, there is a lack of knowledge on how this special group makes usage decisions on this monitoring service, practically and theoretically. The current literature on this issue has thus far mainly adopted a technology acceptance perspective and has largely ignored the health-related factors ^[5, 6]. Therefore, our research integrates health-related factors to investigate: How do patients with chronic diseases make usage decisions regarding the mHealth monitoring service? Exploring patients' decision-making from a health perspective will not only add to the current knowledge on health technology adoption, but also provide a basic understanding of how patients with chronic diseases make health-related decisions.

Due to the long-term influences of the chronic diseases, chronically ill patients may generate negative emotions, such as anxiety and depression, when facing their health issues ^[7]. These experiences will induce them to have a highly severe perception of their health conditions. Therefore, this study draws on negative health emotions and health severity to manifest the physical and emotional effects of the chronic diseases on patients, respectively. Prior evidence has indicated that services or transactions through virtual channels based on ICTs can provoke many uncertainties and potential risks ^[8]. Individuals possess different tolerances of uncertainty in their daily decision-making, which is determined by their characteristic of uncertainty-avoidance ^[9]. Accordingly, when they make health-related decisions, their health uncertainty avoidance can sharpen their decision processes. Therefore, to gain a better understanding of the role of negative health emotions and health severity in patients' usage of the MMS, this research further explores the contingency role of health uncertainty

avoidance.

To address the aforementioned questions, a theoretical research model is developed and empirically tested by a survey among patients with chronic diseases. In doing so, this research contributes to the extant literature in several aspects. First, our research is possibly one of the first to explore how users decide to use the mHealth monitoring service mainly from a health perspective. Second, by explicitly investigating the contingency role of health uncertainty avoidance, our findings shed light on the relative importance of the physical and emotional conditions for chronically ill patients' mHealth usage decisions. Third, our research focuses on a special group, i.e. patients with chronic diseases, and explores their special responses to mHealth services. We also contribute practical implications arising from this research to guide mHealth practitioners and providers.

The remaining sections of the paper are organized as follows. In the next section, we will review the literature on the mHealth monitoring service, the characteristics of patients with chronic diseases, and health uncertainty avoidance. Following the review, the research model is proposed. This is followed by an overall description of the research methodology and results, after which the key findings and implications are discussed. Finally, we end with a conclusion.

2. LITERATURE REVIEW

2.1 The mHealth monitoring service

mHealth service can be defined as the use of mobile information and communication technologies (ICTs) endowed with the capability of managing and delivering health information timely, between end-users and health professionals to improve patient safety and the quality of healthcare^[10]. mHealth service is a promising platform for chronic care due to the advantages provided by mobile ICTs. The long-term nature of chronic diseases promotes the need for health services while also leading to expense issues^[11]. A possible solution for these problems is to empower patients with the skills for self-health management^[12]. Thus, the MMS is important in their daily lives to keep them informed of their health conditions^[11]. Some researchers have verified the role of MMS in chronic care, such as by increasing clinical outcomes^[13, 14], decreasing patient costs^[15, 16], and improving patient-provider compliance^[14, 16].

As the MMS is an emerging service, behavioral studies on this topic are still scarce. Although the current research has mainly drawn on a technology perspective by applying the well-developed theories and models in the technology acceptance area, the health-related factors are largely ignored, leaving the mHealth diffusion research incomplete. To narrow this gap in the mHealth context, this study is designed to explore the effects of health-related factors, i.e. physical and emotional conditions on the diffusion of the MMS.

2.2 Characteristics of patients with chronic diseases

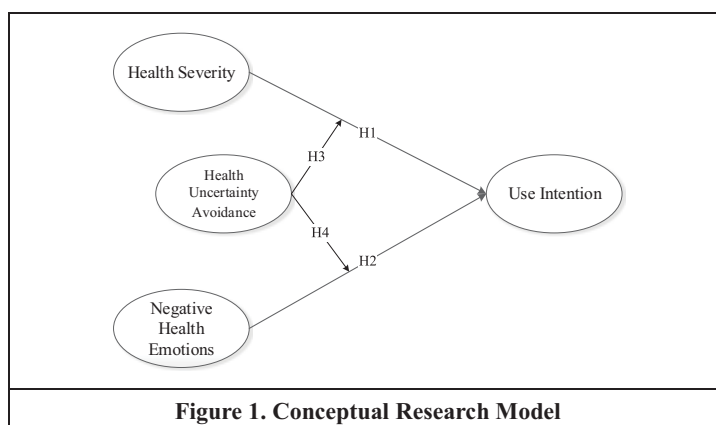
Chronic diseases are long-term conditions and cannot be cured easily and instantly, compared with the acute diseases^[17]. Patients with chronic diseases face many challenges for managing their health conditions in their daily lives. They would feel more threatened, which may, in turn, give rise to more negative emotions on their health, such as anxiety and depression^[7]. As chronic diseases are usually associated with substantial distress and functional limitations, patients have to deal with many changes in their daily lives, such as a feeling of discomfort, loss of potential opportunities, and the facing of financial constraints^[18]. Recognizing that these changes are caused by chronic diseases, which cannot be easily treated, they generally consider that their health condition is serious when making health-related decisions.

2.3 Uncertainty avoidance and health

Uncertainty avoidance refers to the extent to which individuals feel threatened by uncertain situations^[19]. With a strong sense of uncertainty avoidance, individuals will experience stress and anxiety in their daily routines, which may also influence their decision-making regarding new services and technologies^[20]. Since it is manifested as a general feeling of anxiety when individuals face uncertainty and challenges, uncertainty avoidance is widely seen as a personality of individuals that moderates their adoption decisions regarding various services and technologies^[20-22]. Accordingly, we define health uncertainty avoidance as the tendency to avoid any unexpected or unknown risks and uncertainty in health-related decisions. Due to the outcome randomness, information ambiguity, and treatment complexity, most health-related decisions are uncertain^[23]. Thus, when making health-related decisions, individuals are faced with many uncertainties, especially in the case of chronically ill patients. However, how patients make decisions under conditions of uncertainty and whether this special characteristic shapes their decision-making remain underexplored. To narrow this gap, our research proposes and tests the contingency role of health uncertainty avoidance in patients' MMS usage decisions.

3. RESEARCH MODEL

To address the previously mentioned research gaps, a conceptual research model with four hypotheses is proposed. Figure 1 is an illustration of the model. We explore the effects of health severity and health negative emotions on use intention and how such effects are shaped by health uncertainty avoidance among patients with chronic diseases.



Health severity refers to the patients' perceptions of whether their health problems, i.e. chronic diseases, are serious issues. In situations of high health severity, patients with chronic diseases will devote more attention to their health and are more likely to adopt more preventive health behaviors and purchase more preventive healthcare products and services^[24]. For these patients, the MMS can help them to self-monitor and self-manage their daily healthcare, which can be an effective measure for preventive healthcare^[25]. Therefore, we expect that when chronically ill patients perceive high severity regarding their health conditions, they are more likely to choose to use the MMS for health prevention. Thus, we hypothesize that:

H1: Health severity is positively associated with MMS usage intentions.

Patients' emotions are closely related to their health conditions and the long-term chronic experiences generally lead patients to experience negative emotions such as anger, worry, and depression^[26, 27]. Such feelings can induce them into generating aversion of their chronic conditions, and they will be more positive towards the risk-seeking choices for their health conditions^[28]. Hence, patients with chronic conditions and facing negative emotions on their health conditions will be more likely to adopt new services, e.g. the MMS, as

a potential approach to keep the diseases at a distance. Therefore, we propose that the negative health emotions of patients will induce them to use the MMS. Thus, we hypothesize that:

H2: A negative health emotion is positively associated with MMS usage intention.

With a high sense of health uncertainty avoidance, individuals will feel upset concerning the unknown or uncertain health situations ^[20]. Patients with highly serious health conditions will face many uncertainties regarding their physical conditions. In this situation, the high uncertainty avoidance will lead them to focus more on finding ways to reduce the uncertainties. Therefore, when making decisions on whether to use the MMS for their chronic diseases, there is a higher possibility for them to choose usage to reduce their feelings of health uncertainty. On the other hand, patients feeling a low sense of health uncertainty avoidance will pay less attention to their health conditions and feel less upset. They are less likely to use the MMS compared with those sensing high uncertainty avoidance. Thus, we hypothesize that:

H3: Health uncertainty avoidance positively moderates the relationship between health severity and use intention.

A negative health emotion induces patients into averting the current conditions and hence, they experience a higher tendency to make risk-seeking decisions to improve their health conditions ^[28]. On the contrary, uncertainty avoidance causes individuals to choose more certain decisions and avoid uncertain decisions. Patients possessing feelings of high uncertainty avoidance on their health conditions are less likely to make risk-seeking decisions, and those feeling a sense of low uncertainty avoidance will rely more on their inner negative feelings to make health decisions. Therefore, health uncertainty avoidance will weaken the effect of negative health emotions on user decisions. Thus, we hypothesize that:

H4: Health uncertainty avoidance negatively moderates the relationship between negative health emotions and use intention.

4. METHODOLOGY

4.1 Measures and data collection

To test the research model, a survey is conducted among chronically ill patients by means of a questionnaire. As the theoretical constructs are widely measured and used in previous empirical studies, this research, therefore, adopts these measures and adapts them to our research context. Then the content validity was tested by several scholars. The measures and their original sources are presented in the Appendix.

Of 504 participants, 213 subjects experiencing one or more chronic diseases were considered as valid participants. An incentive of about US\$3 was provided as a form of motivation. Of these respondents, 159 of them have one chronic disease while 54 have more than one such disease. Of these, 40.4% are female. More than half of the respondents are in their forties, while 16.9% are in their fifties, and 3.8 % are in their sixties. To test our model, we conducted measurement model and structural model analyses subsequently.

4.2 Measurement model

Smart PLS was used to test our measurement model. For the reflective constructs, the reliability, convergent validity, and discriminant validity of the measurement model were examined as indicators of the goodness of the measurement model. The reliability of the measurement model was assessed by examining Cronbach's alpha, composite reliability (CR) and average variance extracted (AVE) ^[29]. The results are presented in Table 1. In our study, the threshold values of CRs and AVEs were .70 and .50 respectively, consistent with those of Chin ^[30]. According to Nunnally ^[31], a value of at least .70 of Cronbach's alpha indicates adequate reliability. Composite reliabilities for these constructs ranged from .822 to .938, and the average

variances extracted varied from .571 to .835. These results suggest that all indicators are above the cut-off values, indicating good construct reliability^[29]. All item loadings on expected constructs were greater than their cross-loadings on other constructs, and the correlations of the constructs were significantly smaller than the square roots of the AVE of each construct, indicating that the constructs have good discriminant validity.

Table 1. Correlations and Discriminant Validity

	Cronbach's Alpha	CR	AVE	UI	NHE	HSEV	HUAE
UI	.901	.938	.835	.913			
NHE	.742	.822	.608	.188	.779		
HSEV	.828	.839	.571	.149	.414	.755	
HUAE	.863	.906	.709	.416	.026	.102	.842

Note: The diagonally arranged bold numbers are the square roots of AVEs.

4.3 Structural model

The structural model was assessed by checking the significance of path coefficients (β) between various factors. The results show that negative health emotion does not significantly influence use intention ($\beta=.092$, $t=1.291$) and health severity significantly influences use intention ($\beta=.150$, $t=2.603$). Therefore, H1 was supported, while H2 was not. Regarding the moderating effects, health uncertainty avoidance positively moderates the relationship between health severity and use intention ($\beta=.133$, $t=1.890$), thus supporting H3. Health uncertainty avoidance negatively moderates the relationship between negative health emotion and use intention ($\beta=-.133$, $t=1.890$), and therefore H4 is supported. These factors fully explain 24.3% of the variance of use intention.

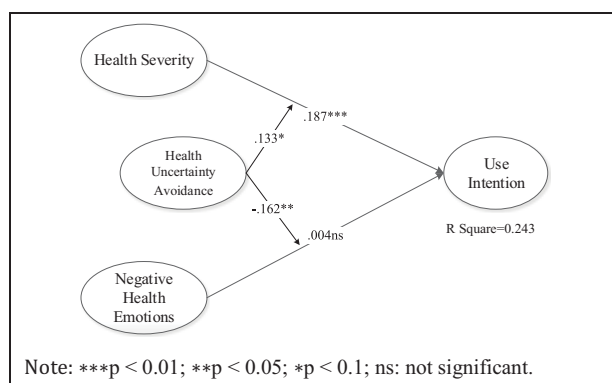


Figure 2. Results of the Research Model

5. DISCUSSION

5.1 Key findings

By testing the proposed research model and hypotheses, our research highlights four aspects of the key findings. First, the health severity perceptions of patients with chronic diseases enhance their use of the MMS for health monitoring. The relationship between health severity and use intention makes sense for these patients in that, if they perceive their health problems as serious, they will feel a strong need to take health preventive measures.

Second, while we propose that negative health emotions can motivate the patients into using the MMS, the empirical results do not support. This result seems controversial, but it is acceptable given the special participants of our study. Different from the general population, the patients with chronic diseases could be

influenced by the long-term effects of their diseases. Negative health emotions should be normalcy to those with chronic disease and do not have significant effects on their behavioral intentions toward MMS.

Third, we find that health uncertainty avoidance strengthens the effects of health severity on use intention. This indicates that if patients feel a sense of high uncertainty avoidance regarding their health conditions, the effects of their serious perceptions on their health status will have a stronger effect on their use intentions. This is because once the patients experience a high tendency to avoid health uncertainty they will feel a strong demand to take measures to deal with their chronic diseases.

Finally, health uncertainty avoidance weakens the effects of negative health emotions on use intentions. This negative moderating effect indicates that when patients feel a high sense of health uncertainty avoidance, their negative health emotions will deter them from using new services to monitor their health conditions. The direct effects of the negative health emotion are not significant but its interaction with health uncertainty avoidance is negatively significant. This indicates that patients with chronic diseases are more likely to make a more risk-avoidance decision when they are presented with MMS.

5.2 Theoretical implications

Our research contributes several implications to the extant literature. First, ours is possibly one of the first to explore the diffusion of the mHealth service from a health perspective. Although the diffusion of the mHealth or other services has been widely explored in recent decades, most prior studies have adopted a technology perspective and investigated the well-developed technology acceptance theories and models mainly from a theoretical lens. However, the health-related factors have been largely neglected. When exploring patients' responses to health services, their physical and emotional conditions, as well as the tendencies of health decision-making can play a critical role in determining their behavior.

Second, this research explores the important contingency role of health uncertainty avoidance. Our results indicate that health uncertainty avoidance positively moderates the effects of health severity and negatively moderates the effects of negative health emotions. The special contingency role of health uncertainty avoidance indicates that it plays a significant role in shaping patients' decision-making regarding health services. By explicitly investigating the contingency role of health uncertainty avoidance, our findings shed light on the relative importance of physical and emotional conditions for patients' mHealth usage decisions among different patient groups.

Third, little research attention has been devoted to the MMS usage behavior of this special group. By investigating the effects of their unique responses from their physical and emotional conditions, our research provides a basic understanding of how chronically ill patients respond to mHealth monitoring service.

5.3 Practical implications

This study also has several practical implications. First, service providers should be aware of the factors that cause patients to use their services. As patients' serious perceptions of their health conditions positively influence their use intentions, the providers are advised to exert efforts on informing the patients about the hazards of their chronic diseases. Hence, they can target their potential customers, patients with serious chronic conditions, who have also realized the hazards of their diseases. Second, in their role of facilitating health uncertainty avoidance, they can segment such customers in their marketing campaigns. To assist patients with serious health concerns, providers can assist them by facilitating health uncertainty avoidance, with measures such as paying more attention to those with a higher sense of health uncertainty avoidance. Providers can alleviate the problems faced by patients with negative emotions on their health by devoting more attention to those with a lower sense of health uncertainty avoidance. In this way, the service providers will experience a

higher possibility of transforming these patients with chronic diseases from potential customers to actual ones.

6. CONCLUSIONS

The increasing population of patients with chronic diseases has promoted the use of mobile ICTs in mHealth monitoring services. While the MMS is suitable for these patients, we know very little about how they make usage decisions regarding the MMS, and whether their health-related features cause them to make different decisions. To address our research questions, this study creates a theoretical model to test the effects of health severity and negative health emotions on the usage intentions of the MMSs and the contingency role of health uncertainty avoidance. The model was tested employing a survey among patients with chronic diseases. We find that health severity positively influences usage intentions while health uncertainty avoidance plays different moderating roles on the effects of health severity and negative health emotions. This research contributes to the understanding of the diffusion of the mHealth service, the role of health-related factors in decision-making, and the unique decision-making processes of patients with chronic diseases.

Appendix

Measurement Items		
Construct	Measure	Source
Use Intention	I plan to use the mHealth services in the next 3 months	20
	I predict I will use the mHealth service software in the next 3 months	
	I plan to use the mHealth services in the next 3 months	
Negative Health Emotion	I feel furious about my present health status	22
	My current health status is a real inconvenience	
	My present health problems fill me with dread	
	I feel disgusted with my current state of health	
Health Severity	My health issues are severe	20
	My health issues are serious	
	My health issues are significant	
Health Uncertainty Avoidance	If I use the mHealth services, I will increase my effectiveness on avoidance of any uncertainty or unknown situations related to my health status.	21
	If I use mHealth services, I will spend less time feeling concerned about any uncertainty or unknown situations related to my health status.	
	If I use Health services, I will improve the quality of avoidance of any uncertainty or unknown situations related to my health status.	
	If I use mHealth services, I will increase the quantity of output for the same amount of effort in avoiding any uncertainty or unknown situations related to my health status.	

REFERENCES

- [1] Xiaohui Y, Han H, Du Jiadong W, et al. Washington, DC: China Academy of Telecommunication Research of MIIT and Center for Technology Innovation at Brookings, 2014.
- [2] Pollettini J T, Baranauskas J A, Ruiz E S, et al. Surveillance for the prevention of chronic diseases through information association [J]. BMC Medical Genomics, 2014, 7(1): 7-17.
- [3] Walls J. Wireless changing US healthcare [J]. Washington Post, 2013,

- [4] Mochari-Greenberger H, Peters A, Vue L, et al. A Nationally Scaled Telebehavioral Health Program for Chronic Pain: Characteristics, Goals, and Psychological Outcomes [J]. *Telemedicine and e-Health*, 2017, 23(8): 640-8.
- [5] Guo X, Zhang X, Sun Y. The privacy–personalization paradox in mHealth services acceptance of different age groups [J]. *Electronic Commerce Research and Applications*, 2016, 16(55-65).
- [6] Sun Y, Wang N, Guo X, et al. Understanding the acceptance of mobile health services: a comparison and integration of alternative models [J]. *Journal of Electronic Commerce Research*, 2013, 14(2): 183.
- [7] Andenæs R, Kalfoss M H, Wahl A K. Coping and psychological distress in hospitalized patients with chronic obstructive pulmonary disease [J]. *Heart & Lung: The Journal of Acute and Critical Care*, 2006, 35(1): 46-57.
- [8] Vishwanath A. Comparing online information effects: A cross-cultural comparison of online information and uncertainty avoidance [J]. *Communication Research*, 2003, 30(6): 579-98.
- [9] Hofstede G. *Culture's consequences* [M]. Thousand Oaks, CA; Sage Publications. 2001.
- [10] Consulting V W. *mHealth for development: the opportunity of mobile technology for healthcare in the developing world* [M]. Washington Dc and Berkshire, UK. 2009.
- [11] Mirza F, Norris T, Stockdale R. Mobile technologies and the holistic management of chronic diseases [J]. *Health Informatics Journal*, 2008, 14(4): 309-21.
- [12] Radin P. “To me, it's my life”: medical communication, trust, and activism in cyberspace [J]. *Social Science & Medicine*, 2006, 62(3): 591-601.
- [13] Ramachandran K, Husain N, Maikhuri R, et al. Impact of a comprehensive telephone-based disease management programme on quality-of-life in patients with heart failure [J]. *The National Medical Journal of India*, 2006, 20(2): 67-73.
- [14] Shetty A S, Chamukuttan S, Nanditha A, et al. Reinforcement of adherence to prescription recommendations in Asian Indian diabetes patients using short message service (SMS)—a pilot study [J]. *J Assoc Physicians India*, 2011, 59(711-4).
- [15] Leong K C, Chen W S, Leong K W, et al. The use of text messaging to improve attendance in primary care: a randomized controlled trial [J]. *Family Practice*, 2006, 23(6): 699-705.
- [16] Ostojic V, Cvoriscec B, Ostojic S B, et al. Improving asthma control through telemedicine: a study of short-message service [J]. *Telemedicine Journal & e-Health*, 2005, 11(1): 28-35.
- [17] Deng X, Khuntia J, Ghosh K. Psychological Empowerment of Patients with Chronic Diseases: The Role of Digital Integration [M]. *Thirty Fourth International Conference on Information Systems*. Milan. 2013.
- [18] Taylor S E, Aspinwall L G. *Psychosocial aspects of chronic illness* [M]//PT C, GR V. *Psychological aspects of serious illness: chronic conditions, fatal diseases, and clinical care*. Washington, DC; American Psychological Association. 1996.
- [19] Geert H. *Cultures and organizations: Software of the mind* [M]. New York: McGraw-Hill, 1991.
- [20] Srite M, Karahanna E. The role of espoused national cultural values in technology acceptance [J]. *MIS Quarterly*, 2006, 30(3): 679-704.
- [21] Hwang Y, Lee K C. Investigating the moderating role of uncertainty avoidance cultural values on multidimensional online trust [J]. *Information & management*, 2012, 49(3): 171-6.
- [22] De Meulenaer S, De Pelsmacker P, Dens N. Power Distance, Uncertainty Avoidance, and the Effects of Source Credibility on Health Risk Message Compliance [J]. *Health Communication*, 2017, Forthcoming (1-8).
- [23] Han P K J. Uncertainty and Ambiguity in Health Decisions [M]//DIEFENBACH M A, MILLER-HALEGOUA S, BOWEN D J. *Handbook of Health Decision Science*. New York, NY; Springer

- New York. 2016: 133-44.
- [24] Jayanti R K, Burns A C. The antecedents of preventive health care behavior: An empirical study [J]. *Journal of the Academy of Marketing Science*, 1998, 26(1): 6-15.
- [25] Lim S, Xue L, Yen C C, et al. A study on Singaporean women's acceptance of using mobile phones to seek health information [J]. *International journal of medical informatics*, 2011, 80(12): e189-e202.
- [26] Trumbo C W, McComas K A, Kannaovakun P. Cancer anxiety and the perception of risk in alarmed communities [J]. *Risk analysis*, 2007, 27(2): 337-50.
- [27] Anderson C L, Agarwal R. The digitization of healthcare: boundary risks, emotion, and consumer willingness to disclose personal health information [J]. *Information Systems Research*, 2011, 22(3): 469-90.
- [28] Druckman J N, McDermott R. Emotion and the framing of risky choice [J]. *Political Behavior*, 2008, 30(3): 297-321.
- [29] Fornell C, Larcker D F. Evaluating structural equation models with unobservable variables and measurement error [J]. *Journal of marketing research*, 1981, 39-50.
- [30] Chin W W. The partial least squares approach to structural equation modeling [J]. *Modern Methods for Business Research*, 1998, 295(2): 295-336.
- [31] Nunnally J. *Psychometric Methods* [M]. New York; McGraw-Hill. 1978.

The Motivation of Capital-giving in Crowdfunding Market: A Self-determination Theory Perspective

Xiang Yuan¹, Hongwei Wang^{2}, Yuan Meng³*

¹School of Economics and Management, Tongji University, Shanghai, 200092, China

²School of Economics and Management, Tongji University, Shanghai, 200092, China

³ School of Information Management, Shanghai Lixin University of Accounting and Finance, Shanghai, 201209, China

Abstract: How to promote crowd-funding results successfully are crucial to crowdfunding platforms and crowdfunding projects. The results of crowd-funding projects are determined by investors' subjective behavior, which is triggered by some certain motivations. However, for different investors, the motivation toward a speculative behavior may be different. Thus, it is very necessary to explore and analyze the composition of the motivations behind each investor's decision. In this paper, we identify different motivation modes mainly influenced by the project description, which will be beneficial to identify the investment intention of each investor. Based on the self-determination theory, we first create the corpus targeting different motives by means of the text mining method. Then, we classify the project description and project investment options. Last, we conduct an econometric model to examine the effect of investor's motives on crowd-funding results based on the real dataset from Indiegogo Platform.

Keywords: crowdfunding, motivation, investment, intention

1. RESEARCH QUESTION

In the Internet crowdfunding market, how to increase the success rate of Internet crowdfunding have become a hot topic. Existing research focuses on investment results and influencing factors of crowdfunding projects, lacking analysis of investors motivation difference. These different investment motivations affect investors' investment behavior, as well as the success rate of crowdfunding projects. According to the self-determination theory^[1], the physiological needs of the humans are classified into intrinsic motivations and extrinsic motivations. Extrinsic motivations refer to those that are generated because of the external awards or return. Intrinsic motivations include three psychological needs: 1. Autonomy refers to the self-control need of individual humans for behavior. 2. Competence refers to the need of individual humans to express the individual capacity. 3. Relatedness refers to the need that individual humans need to maintain correlation with others.

Based on the self-determination theory, we analyze which investors' investment motivations influence investors' investment behavior and the success rate of crowdfunding projects.

2. MAJOR RESEARCH FINDINGS

We establish an econometric model to evaluate the success rate of Internet crowdfunding and the effect of investors' investment motivations. See Formula (1).

$$\text{Rate}_i = \alpha + M_i' \beta + Z_i' \gamma + \varepsilon_i \quad (1)$$

Rate_i is a dependent variable and represents the completion rate of Internet crowdfunding projects. M_i' is the investment motivation vector, it represents the text effect of the project initiator on the investment motivations. Project description is the main channel for investors to understand the project, which impose a certain effect on the investors' behavior. Besides, different text contents may impose different investment

* Corresponding author. Email: hwwang@tongji.edu.cn (Hongwei Wang)

motivation effects on investors, which further affect the investment behavior of the investors. We utilize the TF-IDF value of the key word that represents the theme meaning in the project description as the independent variable of the investment motivations. Z'_i is the control variable set. Based on the existing researches [2], the following 6 variables are employed as the control variables of the model in this paper: fund-raising period, fund-raising target, description length, initiator, number of Facebook friends, updating times and comment times. Besides, α is the intercept of the model; β and γ respectively refers to the coefficients of the investment motivation vector and the control variable; ε_i represents the random disturbance factor. In general, $\varepsilon_i = N(0, \delta^2)$.

We classified all the crowdfunding projects into different categories based on fund-raising contents and analysed each category independently. Table 1 shows the analysis result. It can be seen that each project is affected by different investment motivation effect and displays different tendency.

The fund raiser usually takes advantage of words emphasizing the return or discount to stimulate the investors' investment motivation of obtaining the return. The result shows that the investors of art and food categories concern about the return of the projects, and the investors focusing on the local business and rights prefer the projects with little reward presentation. The relation words which stimulate the investors to enjoy the sense of relation imposes positive correlation effect on most kinds of crowdfunding project, these investors may hope to establish certain relation with the fund raiser or join the community. The

words conveying limit impose a positive effect on the extrinsic motivations of individuals and a negative effect on autonomy of individuals, the result shows that limit words have a positive correlation with the projects of several categories. The fund raiser usually employs words emphasizing gratitude or help for satisfying the competence need, we can find that three categories are related to the words concern for help and gratitude meaning.

3. CONCLUSIONS

This paper introduces the motivation theory in the area of Internet crowdfunding, providing a new perspective for analysis on Internet investor motivation and behavior. We have studied the effect of investment motivations of crowdfunding projects investors on the success rate of Internet crowdfunding in this paper. Through a list of key words generated by WordNet lexicon analysis, we use an econometric model to analyze the investment motives and effects of successful fund-raising of crowd-funding projects. In general, due to the diversity of project categories, the effect of the investment motivations on each category of crowdfunding projects is different, and the investors show different tendencies in investment motivations for different project

Table 1. Analysis results of the effect of the investment motivations on the project fund-raising results

Category	Help	Gratitude	Reward	Limit	Relation
Art	--	--	0.053*	--	0.085***
Game	--	--	--	--	--
Music	--	--	--	0.067***	0.037*
Film& Video	--	--	--	0.036***	0.032*
Journey	--	0.15*	--	--	0.169*
Podcasts& Shows	--	--	--	0.052*	--
Publishing	--	--	--	0.058*	--
Rights	--	--	-0.020*	--	0.064***
Education	--	--	--	0.055***	0.053***
Environment	--	--	--	0.069**	0.088**
Health	--	0.033*	--	--	0.042**
Culture	--	--	--	--	--
Food	--	--	0.075*	--	--
Local Business	--	--	-0.028*	--	--
Community	0.064*	--	--	--	0.053*
Design& Innovation	--	--	--	--	--

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$;

categories. The research results provide guidance for the fund raiser to prepare a more attractive project description; the crowd-funding platform can provide the project description guidance suitable for the fund raiser according to the different categories of crowd-funding projects, so as to increase the whole fund-raising success rate of the platform.

ACKNOWLEDGEMENT

This work is partly supported by the Natural Science Foundation of China [71771177, 71601119, 71601082].

REFERENCES

- [1] Deci E L, Ryan R M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4): 227-268
- [2] Wang W, Wei C, Zhu K, Wang H W. (2016). The Impact of Linguistic Persuasiveness on the Success Rate of Crowdfunding Campaigns: Evidence from Kickstarter. *Chinese Management World*, (5): 81-98

What Factors Will Determine Users' Knowledge Payment Decision?

An Theoretical and Empirical Research

Yang Zhao^{1,2}, Liqiang Wu¹, Yu Zhao¹, Xini Yuan¹*

¹School of Information Management, Wuhan University, Wuhan, 430072, China

²The Center for Electronic Commerce Research and Development, Wuhan University, Wuhan, 430072, China

Abstract: With the increase of peoples' eagerness for higher quality knowledge, paid Q&A is becoming a new tendency. However, what factors are helpful to drive potential users' payment decisions remains unknown. In this paper, we investigated the effects of expert attributes and reputation on users' payment decisions made on an online Q&A platform in China. We developed auto-parsing crawlers to collect online observational data and used the negative binomial panel regression method to estimate the effects of expert attributes and reputation on users' payment decision. The results show that expert attributes such as the number of paid questions, the number of times that answers are approved, whether the expert has a personal home page, whether the expert mentions his/her area of expertise, the number of followers, score of expert answers have significant effects, whereas the times that the expert shared knowledge free and whether the expert has a real name certification do not influence users' willingness to pay for an answer. The results help experts on paid Q&A platforms to improve their performance, perfect their personal information, and enhance users' trust, so as to promote the development of knowledge sharing economy.

Keywords: Paid Q&A, Sharing Economy, Users' Payment Decision, Reputation

1. RESEARCH QUESTION

The spring up of paid Q&A has facilitated the development of knowledge-sharing economy, which makes it possible for experts in different industries to do online transactions with their knowledge. However, unlike other C2C transactions, except for price, ordinary users (buyer) barely have no information about the knowledge commodity (answer) before the expert (seller) answers to the question. Trust problem is one of the greatest barriers between sellers and buyers who are not familiar with each other, and trust has been found as the fundamental determinant of consumer decisions when it comes to the establishment of transaction relationship. Therefore, we believe that studying users' payment decision in paid Q&A from the trust perspective is meaningful, and the research results may help lower users' perceived risk and improve the service quality provided by experts. Therefore, the two research questions addressed in this study are:

- What are the key factors that affect users' payment decision in online Q&A?
- How do these factors influence users' payment decision?

The establishment of the trust relationship between questioners and the respondents (answerers) is the prerequisite for users to pay for the professionally generated content (PGC). Therefore, this study builds a theoretical model based on trust theory, and explores the factors that affect users' decision on paying for PGC, from the dimensions of trusting beliefs and reputation respectively. Adapting viewpoints of Mayer and Schoorman, we measure the trustworthiness of experts from three perspectives: ability, benevolence and integrity. Based on the above theory, we proposed the theoretical framework of this study with eight hypotheses.

H1a: The number of paid questions the expert answered has a positive influence on users' payment decision.

H1b: Number of times that answers are approved has a positive influence on users' payment decision.

H2: Number of times that experts freely shared knowledge has a positive influence on users' payment

* Corresponding author. Email: yangzhao_0813@hotmail.com (Yang Zhao)

decision.

H3a: Whether the expert has a real name certification has a positive influence on users' payment decision.

H3b: Whether the expert has opened a personal home page has a positive influence on users' payment decision.

H4: Whether users are accessible to the information about the detailed area of expertise has a positive influence on users' payment decision.

H5a: Number of followers has a positive influence on users' payment decision.

H5b: Number of followers has a positive influence on users' payment decision.

2. MAJOR RESEARCH FINDINGS

In this paper, we did a data-driven study using real-world data from Zhihu.com to estimate the effects of expert attributes and reputation on users. Through the Homan test, we decided to use negative binomial panel regression method under the fixed effect model to estimate. The estimation results are reported in Table 1.

Table 1. Estimation result

Dependent Variable: the number of users that pay to question (in 10 days)				
	Coefficient	Standard Error	P-value	VIF
Ln(Price)	-0.292	0.011	0.000	1.14
Ln(Reply_Num+1)	0.497	0.032	0.000	2.80
Ln(Worth_Num+1)	0.207	0.015	0.003	3.74
Ln(Share_Num+1)	0.056	0.006	0.668	1.24
Has_Authenticated	0.005	0.003	0.213	2.67
Is_Infinity	0.310	0.002	0.000	1.74
Is_Proficient	0.192	0.052	0.000	2.19
(Follower ^{0.17} -1)/(-0.17)	0.212	0.049	0.009	1.12
(Score ^{6.88} -1)/(-6.88)	0.126	0.051	0.000	1.45
_Cons	-1.172	0.135	0.000	-

According to the result showed in Table 2, except H2 ($\beta_4=0.056$; $P=0.668$) and H3a ($\beta_5=0.005$; $P=0.213$), other hypotheses all passed the significance test. Through analysis, we think that the insignificance of “Share_Num” is closely related to the information overload that isolates users from high-quality knowledge.

Under the circumstances that anyone can easily express his opinion as well as nonsense, although “Share_Num” expresses the expert benevolence in sharing knowledge free, it cannot impress users who are eager to get high-quality knowledge. And the insignificance of “Has_Authenticated” indicates that, compared to other expert attributes, users are less concerned about whether the expert is real name authenticated, for most of the personal information has been displayed on the expert profile page.

3. CONCLUSIONS

In this paper, we did a data-driven study based on trust belief and reputation to estimate the effects of expert attributes on users purchase decision in paid Q&A. As a result, we find that, most of expert attributes pass the significance test, while the times that the expert shared knowledge free and whether the expert has a real name certification have no significant effect.

REFERENCES

- [1] Schoorman FD, Mayer RC, Davis JH. (1995). An integrative model of organizational trust: past, present, and future. *Academy of Management Review*, 20(3): 709-734.
- [2] D Harrison McKnight, Vivek Choudhury, Charles Kacmar. (2002). Developing and validating trust measures for e-commerce: an integrative typology. *Information Systems Research*, 13(3): 334-359.
- [3] Jeon YJ, Rieh SY. (2014). Answers from the crowd: how credible are strangers in social q&a? *Ischools*.

Network Structure Mining and Evolution Analysis - Based on BA Scale-Free Network Model

Rui chen¹, Ting He

School of Business Administration,
Zhongnan University of Economics and Law,
430073, Wuhan, China

Abstracts: The massive adoption of the Internet facilitates growth of online social networks, in which information can be exchanged in a more efficient way. Such as products, user accounts, web pages, there may be a variety of objects suitable to structure this kind of networks. As a result, this gives the networks complexity and dynamics. The work in this paper is aiming to studying the topological property of online social network structure from the aspect of dynamics, and make clear the evolution processes of the networks. This is done by a Mean-Field analysis of network growth based on BA Scale-Free network model. Data resources come from the Chinese online e-commerce platform you.163.com and graphs are modeled through commentator and mutual comments by calculating degree distribution of the networks. We build a growing random model for forecasting dynamics of degree evolution. Finally, we use data set on Sina Weibo to test the model and the results are satisfying.

Key words: social network; degree distribution; BA model

1. INTRODUCTION

Graphs provide a useful abstract for modeling various networks. Recently this approach has been employed intensively in online business researches. Subscribers, products and even web pages can be modeled as nodes where followship, comments and scores can be links ^[1]. Online social networks are obviously suitable for graph model and therefore the network structure characteristics can be studied. With the help of large scale computing platform and the appropriate data mining techniques, analysis from aspect of network science will give deeper insight into online social networks.

Node degrees may be one of the most basic and important topology properties for a graph ^[2]. Instead of counting static node's degrees, we observe the over-all distribution of the degrees. From the perspective of statistical probability, we denote $p(k)$ as the proportion of nodes which has degree k , and the p is exactly the distribution of nodes' degree we familiar with. The distribution of nodes' degree gives us a straightforward indication of how the whole network is linked. Furthermore, different distribution of degree indicates different network properties. The distribution of ER Random-Network-Model proposed by Erdős in 1959 obeys Gaussian distribution ^[3]. In 1999, Faloutsos et al found that the degree distribution of the Internet has a strong power law distribution ^[4]. In many recent studies, it was found that the distribution of the degree of social networks and that of time intervals of human behavior approximate obeys the power law distribution ^[5]. As the Internet data resources are increasing dramatically, the degree distribution of network shows more diversified distribution forms.

The research of this paper is based on BA Scale-Free network model ^[6]. BA Scale-Free network model is a

¹ Rui chen
School of Business Administration
Zhongnan University of Economics and Law
430073, Wuhan, China
Email: ruich@hotmail.com

network model based on node growth choice and power law distribution, which is an important random model outside the ER stochastic network. In the BA model, the edge probability is proportional to the size of the node itself, which reproduces the growth process of some networks [6]. The main feature of the BA Scale-Free model is that the probability of the node selecting the link varies between the two extremes, one is a new node uniform random selection of the original, the other the node is selected according to the current degree of the existing node, which called preferred attachment by Erdős and Albert [7]. We mainly consider the hybrid model and optimize it. Through empirical analysis, the structure of this model not only reflects the correlation of degree, but also shows the basic same type characteristics, which is also consistent with many social networks observed. When the process contains some form of preferred attachment, we also see that some large axis nodes appear in the network, which produces a smaller diameter than in the Poisson random network [8].

2. BA SCALE-FREE NETWORK MODEL AND ITS CHARACTERISTIC ANALYSIS

BA model, proposed by R. Albert and A. Barabasi, is a Scale-Free model, which contains two assumptions. i.e.,

Assumption 1: Networks are growing by time. Assuming a network initially has m_0 nodes, a new node will be added in each time unit, and it will be connected to m of original nodes ($m \leq m_0$). In this case we call it uniformly random choice link [8]. The final distribution of degree of nodes is:

$$F_t(d) = 1 - e^{-\frac{d-m}{m}} \quad (1)$$

Assumption 2: The probability of an existing node to be connected to a new node increases proportionally with the degree of it, that is to say, for an existing node i , the probability of connecting to the new born node is m times the degree of node i to the total degree of all existing nodes at time t . i.e.:

$$m \frac{d_i(t)}{\sum_{j=1}^t d_j(t)} \quad (2)$$

In this case we call it preferred attachment [9]. The final cumulative distribution function is:

$$F_t(d) = 1 - m^2 d^{-2} \quad (3)$$

Both of the above case are extreme cases. One is that a new born node uniformly chooses some of existing nodes to connect to. The other is that the new born node selects existing nodes to connect to according to current degree of the existing nodes. The main feature of the BA scale-free model is that the probability that a born node chooses existing node to link varies between the two extreme cases mentioned above. That is, a new node is connected by two different processes: a uniformly random link and a preferred connecting. Each new node forms m links, in which the probability it uniformly-randomly connect to original nodes is α , and the probability it connect to existing nodes with preferences is $1 - \alpha$. The probability distribution function for this mixed model is:

$$F_t(d) = 1 - \left(\frac{m + \frac{2\alpha m}{1-\alpha}}{d + \frac{2\alpha m}{1-\alpha}} \right)^{\frac{2}{1-\alpha}} \quad (4)$$

When $\alpha = 0$, the distribution function is $F_t(d) = 1 - m^2 d^{-2}$, which is completely under the condition of preferred selection, that is, the same as (3); When $\alpha \rightarrow 1$, its distribution function approximately equals to $1 - e^{-(d-m)/m}$, which is close to the probability distribution (1).

3. THE CALCULATION OF BA SCALE-FREE NETWORK HYBRID MODEL EXTENSION

The previous model usually has a single node at each time. If a fixed node is born in each period, the

characteristics of these systems are generally unchanged. However, if the number of newborn nodes grows with time, the degree distribution will change. Consider an extension of the mixed model such that the number of nodes born during each period grows over time. Assuming that the number of new nodes born at time t is gn_t , where n_t is the number of existing nodes at time t , and $g > 0$ is the growth rate. We estimate the distribution of degrees using continuous time approximations of degree distributions^[8,9].

3.1 Continuous time approximations of degree distributions.

Given nodes numbered with $i \in (1, 2, 3, \dots, t, \dots)$, considering a growing network, each node is denoted by its born time, then the degree of node i born at time t can be represented as:

$$d_i(t) = \varphi_t(i) \quad (5)$$

Where $\varphi^{-1}(d)$ represents the number of the node with degree d , and which is the inverse of $d_i(t)$. While the number is set strictly by $(1, 2, 3, \dots, t, \dots)$, that is also the number(count) of nodes of which the degrees are larger than d , so the proportion of nodes whose degree is less than d is:

$$\frac{t - \varphi^{-1}(d)}{t} = 1 - \frac{\varphi^{-1}(d)}{t} \quad (6)$$

That is, the distribution function is:

$$F_t(d) = 1 - \frac{\varphi^{-1}(d)}{t} \quad (7)$$

Considering the new born nodes added at time t is gn_t , these nodes randomly choose m of the t exist nodes to connect with probability α , thus, the initial condition of node i is $d_i(i) = m$ and for $t > i$ the change of degree over time is approximately:

$$\frac{dd_i(t)}{dt} = \frac{\alpha m g n_t}{n_t} \quad (8)$$

These nodes selected m of the t existing nodes with preference with probability $1 - \alpha$, then the initial condition of node i is $d_i(i) = m$ and for $t > i$ the change of degree over time is approximately:

$$\frac{dd_i(t)}{dt} = \frac{(1 - \alpha) m g n_t d_i(t)}{2 n_t m} \quad (9)$$

So the node's change of degree over time can be represented as:

$$\begin{aligned} \frac{dd_i(t)}{dt} &= \left[\frac{\alpha m}{n_t} + \frac{(1 - \alpha) m d_i(t)}{2 n_t m} \right] g n_t \\ \frac{dd_i(t)}{dt} &= \frac{2 \alpha m g + (1 - \alpha) g d_i(t)}{2} \end{aligned} \quad (10)$$

Where $d_i(i) = m$. The final result is:

$$d_i(t) = \frac{m(1 + \alpha) e^{\frac{g(1 - \alpha)(t - i)}{2}} - 2 \alpha m}{1 - \alpha} = \varphi_t(i) \quad (11)$$

That is:

$$\varphi_t^{-1}(d) = t - \frac{2 \ln \left[\frac{(1 - \alpha)d + 2 \alpha m}{m(1 + \alpha)} \right]}{g(1 - \alpha)} \quad (12)$$

And the distribution function is:

$$F_t(d) = 1 - \frac{t}{n_t} + \frac{2 \ln \left[\frac{(1 - \alpha)d + 2 \alpha m}{m(1 + \alpha)} \right]}{g t (1 - \alpha) n_t} \quad (13)$$

Where

$$n_t = m(1 + g)^t \quad (14)$$

And the final distribution of degrees is:

$$F_t(d) = 1 - \frac{t}{m(1 + g)^t} + \frac{2 \ln \left[\frac{(1 - \alpha)d + 2\alpha m}{m(1 + \alpha)} \right]}{mgt(1 - \alpha)(1 + g)^t} \quad (15)$$

4. EMPIRICAL ANALYSIS

4.1 Basic attribute characteristics of comment- interactive network

We have obtained the comments on the topic "black pig" of the microblogs of "you.163.com" from March 28, 2017 to April 2, 2017 (as shown in table 1.). Consider the degree distribution of the commenter network, where the link represents the response between the two commentators. The final distribution is shown in Table2, the final Network topology is shown in Figure1. In addition, we can also see from Figure1. that you.163.com network's degree distribution obeys power-law distribution, and the network shows strong heterogeneity.

Table1. you.163.com's basic properties of the network structure

Time (t)	3.28	3.29	3.3	3.31	4.1	4.2	4.3
Node(n)	468	775	939	969	990	996	999
Edge (e)	480	823	1049	1088	1109	1115	1119
Average degree (d)	2.051	2.124	2.234	2.243	2.24	2.239	2.24

Table2. Frequency distribution of the interactive network of reviewers

degree	1	2	3	4	5	6	7	8	9	11	12	17	25	29	393
amount	631	29	262	47	17	6	4	1	2	1	1	2	1	1	1

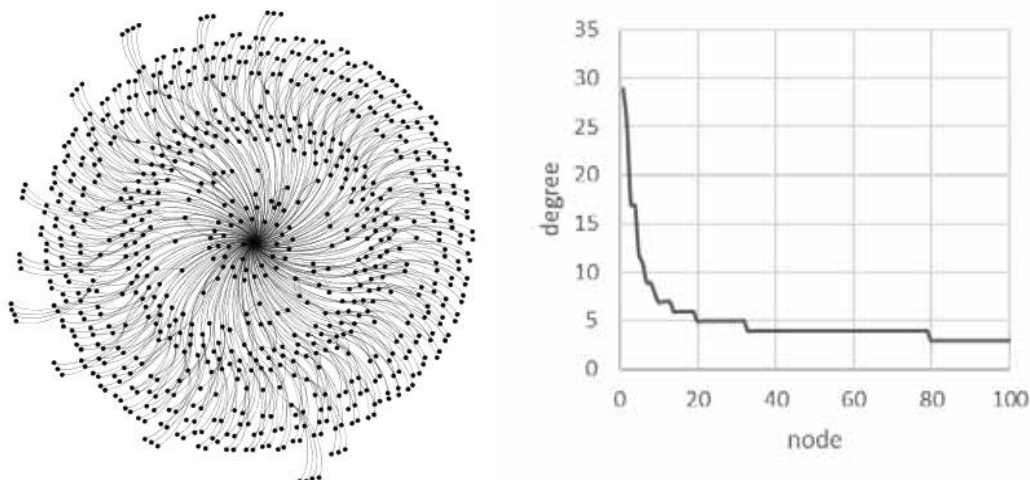


Figure 1. you.163.com's network topology graph and degree distribution

4.2 Degree correlation

Consider a growing hybrid random network formation process as described in Section3.1. Under the mean-field estimate, a node i 's degree is larger than a node j 's degree at time t after both are born if and only if i is older than j [10]. A number of social networks have positive correlation in their degree distribution [11], Figure 2. shows you.163.com's network evolution diagram in 3 units of time, from the figure we can clearly see the

degree of positive correlation. As time goes on, the degree of the original nodes obviously increases, which further reflects that older and higher nodes grow faster than young and lower nodes, and this richer and richer process results in Scale -Free distribution.

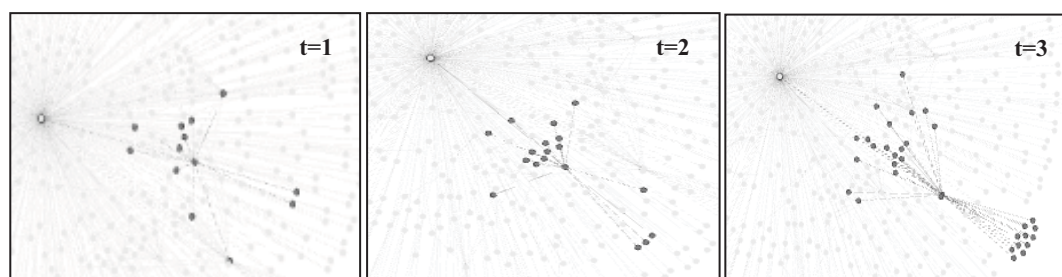


Figure 2. you.163.com's network evolution diagram

4.3 Fitting hybrid degree distributions to data

First, we calculate m directly, since m is the number of connections formed during each period, so it is half of the increase in each period. The total degree is $2tm$, so m is half of the average^[12]. The mean of the network is 2.24, so m is about 1.12, and according to table 1, g is calculated as 15.5%. The initial guess of α , α_0 is used as a starting point, the fitting of the estimation is investigated, and the fixed point of the process is investigated to estimate the parameter α ^[13]. We calculate α according to table 1 and table 2, as shown in table 3. In this case, the final estimated α is approximately 0.823.

Table3. Initial parameter estimate α_0 and final fit α_1

α_0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99	0.999
α_1	-0.63	-0.4	0.34	0.27	0.38	0.55	0.63	0.823	0.825	0.86	0.883

5. CONCLUSIONS

The process of random graph is often complicated, especially when nodes randomly enter into a link and have different degrees of distribution. Based on the BA model, this paper studied the growing random network model and illustrated some of its characteristics. What is important is that its result is more robust than Poisson random network, and in extreme cases, it provides an explanation for scale-free distribution. In order to overcome the challenges of social media data mining, this study provides a structured mechanism to extract values from the data. We can see that there is a significant characteristic of the teletype network which is the heterogeneity, the degree is positively correlated, and the degree distribution is the power law distribution. These empirical results are also consistent with the empirical results of some previous e-commerce networks. However, if you add more network topology considerations, the effect will be more significant. A deep understanding of network structure allows us to develop more effective strategies to serve market decisions.

REFERENCES

- [1] Jiang J, Wilson C, Wang X, et al. Understanding latent interactions in online social networks[J]. ACM Transactions on the Web (TWEB), 2013, 7(4):18.
- [2] Albert R. Barab'asi: Statistical mechanics of complex networks[J]. Rev.mod.phys, 2002, 74(1):2002.
- [3] Barabási A, Albert R. Emergence of Scaling in Random Networks[J]. Science, 1999, 286(5439):509.

- [4] Albert R, Jeong H. Internet: Diameter of the World-Wide Web[J]. *Nature*, 1999, 401(6):130-131.
- [5] M. Faloutsos, P. Faloutsos, C. Faloutsos. On power-law relationships of the internet topology[C]. *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, Cambridge, MA, USA, 1999, 251–262
- [6] Barabasi A L, Albert R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439):509.
- [7] Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks[J]. *Nature*, 2000, 407(6804):651-4.
- [8] Bernard H R, Johnsen E C, Killworth P D, et al. Comparing four different methods for measuring personal social networks [J]. *Social Networks*, 1990, 12(3):179-215.
- [9] Barabási A L, Albert R, Jeong H. Scale-free characteristics of random networks: the topology of the world-wide web[J]. *Physica A Statistical Mechanics & Its Applications*, 2000, 281(1–4):69-77.
- [10] Jackson M O, Rogers B W. Meeting Strangers and Friends of Friends: How Random Are Social Networks?[J]. *American Economic Review*, 2007, 97(3):890-915.
- [11] Littman M L, Kearns M, Singh S. An Efficient, Exact Algorithm for Solving Tree-Structured Graphical Games[J]. *Advances in Neural Information Processing Systems*, 2001:817-823.
- [12] Besag, J. E. (1974) “Spatial interaction and the statistical analysis of lattice systems (with discussion),” *Journal of the Royal Statistical Society, Series B*, 36:196-236.
- [13] Kakade S, Kearns M, Langford J, et al. Correlated equilibria in graphical games[C]// *ACM Conference on Electronic Commerce*. ACM, 2003:42-47.

Exploring Users' Intention to use QQ's Various Functions

based on Social Cognitive Theory

Xun Wang¹, Ying Wu², Chuan Luo³, Yalan Li⁴

^{1,3,4} School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, 611130, China

² School of Accounting, Southwestern University of Finance and Economics, Chengdu, 611130, China

Abstract: Based upon social cognitive theory, this study explores the effect of personal and environment factors on users' intention to use QQ's various functions. Online survey is used to collect data in China. The results show that relationship benefit, switching cost, compatibility and subjective norms can significantly affect users' intention to use QQ's various functions. Whereas image benefit, perceived advantage and popularity have no effect. Finally, we propose the theoretical contribution and practical implication of this study.

Keywords: QQ, social cognitive theory, personal factors, environment factors, intention to use.

1 INTRODUCTION

The thirty-ninth China^[1] Internet development situation report shows that Chinese Internet users have reached 731 million by December 2016, of which mobile users have reached 695 million, accounting for more than 95%. With the development of mobile Internet, the Social Networking Service (SNSs) have changed from PC to mobile terminal such as QQ, Facebook, Twitter, etc. Among them, QQ is one of the most popular SNS APP, it has penetrated into every aspect of people's life. Currently QQ is no longer a simple instant messenger, but a comprehensive APP that integrates various functions, such as QQ wallet, city services, QQ space group and others.

How can QQ continue to dominate the SNS market is a problem in a time, when other social network software is booming, such as WeChat, Alipay, MOMO, and Skype. We believe the development of multifunctional QQ can encourage people to continue to use QQ; consequently, it is an important topic to explore the important factors that can affect the willingness of users to use various functions of QQ. However, as far as we know, seldom of previous research has investigated on this research question, this study endeavor to fill in research gap. We plan to use Social Cognitive Theory (SCT) as the theoretical background, to explore the key influencing factors that can affect QQ users' willingness to use various functions of QQ. We believe this study will not only provide some important theoretical contributions; but also can help the enterprise improve its APP design, which finally will improve the QQ user's usage experience.

The rest parts of this paper organize as following. First we review the relevant literature and introduce the SCT as the theoretical background, second we built our research model based on SCT, then we design and issue questionnaire to collect data. We analyze the data to find out which factors have the significant impact; finally, based on the results of the data, we put forward some suggestions on the functional design of QQ to promote the better development of QQ.

2 LITERATURE REVIEW AND THEORETICAL BACKGROUND

2.1 Literature review

In a recent study, Brandtzaeg and Heim surveyed 1,200 users of SNS and found that the main reason

¹ Corresponding author. Email: wangxunlisle@foxmail.com (Xun Wang), luochuan@swufe.edu.cn (Chuan Luo)

people used SNS one was to establish contact with people^[2]. Ellison et al. and Park et al. also found that people use Facebook as a social network first and foremost because of social needs and environmental factors for the impact of self-positioning and the need to acknowledge the message, which based on a corresponding number of Facebook user surveys^[3]. Przepiorka, Błachnio and Díaz-Morales explored the discovery that people's procrastination was associated with the development of existing instant messaging tools through relevant psychological factors, social factors, investigators' own factor analysis, and modeling^[4]. Other related studies are shown in Table 1 on social networking sites and instant messaging tools:

Table 1. Literatures on SNSs and instant messaging tools.

Study	Area	Factors	Results
Joinson, A.N.(2008)	Social Network Sites (SNSs) use	social connection shared identities photographs content	Social relations, entertainment, information and social interaction, entertainment, information, and personal identity all are important motivation factors for users to choose a software.
Hong, S.Y., Oh, J.C., (2010)	SNSs use	expert search, communication, connection, content Sharing, identity	The study found that users prefer to the different functions about the different social media.
Sung-joon Yoon (2014)	SNSs	effort expectancy, performance expectancy social influence, facilitating conditions, behavioral intention, and so on;	Bridging capital only had significant impact on qualitative use while subjective well-being did not mediate the relationship between social capital and SNS use
Hui Lin Weiguo Fan Patrick Y.K Chau(2014)	Users Continuance of Social Networking Sites	appraisal factors: pleasure, awareness, connectedness, and system quality; emotional reaction: satisfaction and sense of belonging	Appraisal factors are strong determinants of emotional reaction while user satisfaction and sense of belonging together positively influenced continuance intention.
Edmund W.J. Lee, Shirley S. Ho, May O. Lwin(2017)	SNSs	adolescents' relationship with their parents, depression, loneliness, self-identity, habit strength	1. positive relationship with father adolescents' dependence on SNSs 2. identity formation deficient self-regulation and SNSs habit strength adolescents' time spent on SNSs.
Joe Phuaa, Seunga Venus Jinb, Jihoon (Jay) Kima,(2017)	SNSs		social networking sites (SNSs) use
Sang Woo Lee., Jiyoung Lee.(2017)	SNSs	relationship, convenience self-expression entertainment	Compared the different motive factors affecting Kakao Story and Facebook users, information only appeared as motive for Facebook users;

2.2 Social Cognitive Theory

SCT is a learning theory based on the idea that people learn by observing others. The core concepts of this theory can be explained by Bandura's schematization of triadic reciprocal causation. The schema shows how the environmental, behavioral and personal interact with each other^[5]. This theory indicates, various kinds of personal factors and environment factors can significant affect people's behaviors.

After a long period of improvement and development, SCT has been widely used to study the phenomenon. In 2001, Bandura applied SCT to mass communication to analyze how symbolic communication influenced human thought, emotion and action^[6]. Ming-Ten Tsai and Nai-Chang Cheng^[7] showed that knowledge sharing self-efficacy and outcome expectancy, as well as organizational climate, will affect individual intentions to share knowledge, thus affecting people's behavior. As mentioned above, after a long period of improvement and development, SCT has been widely used in the study of phenomena and problems about people's actions and motivations.

3 RESEARCH MODEL AND HYPOTHESIS

3.1 Research Model

Based on the SCT, this study analyzes the personal factors and environment factors' effect on QQ users'

intention to use various functions. We analyze the personal factors from three accepts: benefits, costs and habits. First, benefit is the users' basic needs and spiritual satisfaction of people when using QQ. In this research, we include image benefit, relationship benefit and perceived advantage into the research model. Second, in terms of cost, we believe that there will be a switching cost if the QQ's users abandon QQ to utilize other social software. Third, habit refers to whether the use of QQ consistent with people's real habits in their life, working and communication with others, we will use the compatibility to describe the habits. Environment is the external cause of individual behavior. In this study, we include popularity and subjective norms as the environment factors. With the above description, our research model is showed in Figure 1:

3.2 Hypothesis of Research

3.2.1 Personal factors

We describe the personal factors from benefits (image benefit, relationship benefit and perceived advantage), costs (switching cost) and habits (compatibility), based on SCT and previous research.

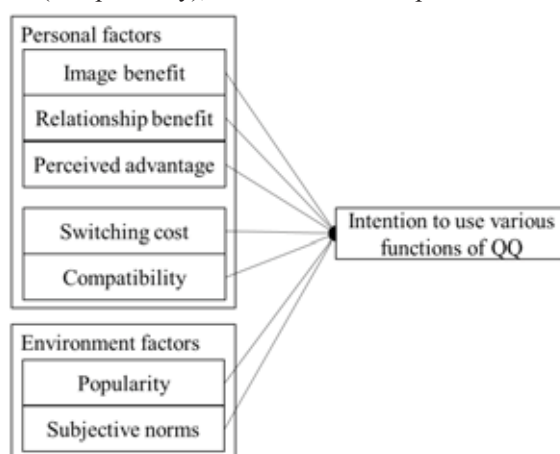


Figure 1: Model Framework

Image benefit refers to the users' personal expression in a social software, which can make others feel the users' personal image, and finally improving their personal recognition. According to the study of the Nadkarni and Hofmann, the two main motivation that people use Facebook are the sense of belonging and self-presentation^[8]. Schlenker believes that self-presentation is a personal attempt to influence others through the act of being perceived by others, and the purpose is to make a special impression on others^[9]. Following this viewpoints, we predict if QQ users can utilize QQ's various functions to fulfill their self-image recognition; they will be more willing to use these QQ functions.

H1: Image benefit positively affects users' intention to use various QQ functions.

Relationship benefit refers to the maintenance and expansion of interpersonal relationships gained through social software. Maintaining relationships has been recognized as an important reason for the use of a social software. Kuss et al. Proposed that social network is mainly used to keep in touch with the offline acquaintances^[10], Jih-Hsin Tang & Ming-Chun Chen believe people use Facebook to keep the existing relationships and expand their social network^[11]. QQ, as a social software prevailing in China, not only can help users keep contact with acquaintances, but also can broaden the scope of their social network. We predict if QQ users consider that the QQ's various functions can help them maintain personal relationship, expanding their social network and other aspects of the relationship, they will be more willing to use these QQ functions.

H2: Relationship benefit positively affects users' intention to use various QQ functions.

Perceived advantage refers to the degree to which individuals perceive their quality of life, job performance or work efficiency can be improved with the use of a social software. It has been suggested that the perceived

relative advantage would have a positive impact on knowledge contributing behavior and knowledge collecting behavior in professional virtual communities^[12]. Rui-Ting Huang^[13] found that perceived relative advantage has impact on the self-posting continuance intention in the social media. So we believe that if a system or platform can play an increasing role on job performance or work efficiency that can strengthen the users' intention to use system or platform. Finally we put forward the hypothesis of H3, as follows:

H3: Perceived advantage positively affects users' intention to use various QQ functions.

Switching costs refers to the real or perceived costs incurred when changing suppliers^[14]. Nakamura found that, the switching costs plays a significant role to affect people's attitude^[15]. More recently, Mikolaj Czajkowski found that switching costs contribute to the total magnitude of consumer lock-in in mobile telephony^[16]. If QQ users abandon QQ and turn to use other SNS APP, they need to bear various kinds of switching costs^[17]. Firstly, they need time and effort to manipulate the function of the new software, Secondly, there are so many sunk costs inserted in QQ, such as money, time, friend, effort. In conclusion, we predict if QQ users consider there is a high switching costs, they will be more willing to use QQ continuously resulting from its various functions.

H4: switching costs positively affects users' intention to use various QQ functions.

Compatibility refers to the extent to which the user's existing values, beliefs, previously introduced ideas, and technology are matched^[18]. In this article, we define compatibility as the degree to which QQ App fits the habits, lifestyle, work style and experiences of individuals. Wu&Wang found that high compatibility has a significant effect in preferable adoption of mobile systems. Furthermore, it also plays an important role in behavioral intention to use^[19]. Rui-Ting Huang^[13] found that compatibility had a positive impact on usefulness and playfulness, which significantly influenced selfie-posting continuance intention. As a conclusion, it is valid to consider that compatibility leads to continuous usage of QQ's various functions.

H5: Compatibility positively affects users' intention to use various QQ functions.

3.2.2 Environment factors

Popularity is defined as the "fact or condition of being well liked by the people".^[20] The higher one's popularity is, people around will be more competing to imitate his or her behavior. The high popularity of people in a group can not only affect the people around, and even can reflect a group of trends^[21]. For those who use social media, especially those who are considered "social media generation", the pursuit of high popularity has inspired social media users to seek more online friends and to disclose more personal information to them^[22]. Following these viewpoints, we can predict that, if QQ can help users get higher popularity, they will have a huge intention to use QQ's various functions. We thus construct the following hypothesis:

H6: Popularity positively affects users' intention to use various QQ functions.

Subjective norms can be defined as the degree to which a person thinks that "most of the people who are important to him or she think he or she should or should not perform the act"^[23]. As Chinese culture is a collectivism culture, individuals feel the pressure of the group is relatively high, therefore, individuals have a relatively high tendency to comply with subjective norms^[24]. Subjective norms have proven to be extremely important to one's acceptance and use of a new technology^[25]. Based on these viewpoints, we have a great chance to believe that if the use of QQ can be accepted by the relevant groups, the individual's willingness to use various functions of QQ will be much stronger. We thus construct the following hypothesis:

H7: Subjective norms positively affect users' intention to use various QQ functions.

4 METHODOLOGY

In this study, we utilized online survey to collect data. We used the mature scale of existing research to measure the variables, we made some changes to fit for the situation of QQ. The whole questionnaire is divided into two parts: The first part includes all the measurement items of the variables in the research model. All of

these items utilize seven point Likert scale, from "strongly disagree" to "strongly agree". The second part is to record the demographic information of the subjects, including gender, age, occupation, education and other demographic data. After 2 months data collection during March to April 2017, we got 450 respondents. Table 2 shows the demographic information of the subjects.

Table 2. QQ Sample statistical Description

index	options	frequency	proportion
Sample range	Prefecture-level city(288)	54	18.75%
	Province(34)	29	85.29%
Gender	Male	197	43.78%
	Female	253	56.22%
Age	Age under 20 years	108	24.0%
	21-30	197	43.78%
	31-40	69	15.33%
	40-50	63	14.0%
	Over 50	13	2.98%
Education	Primary	8	1.78%
	Secondary	61	13.56%
	Undergraduate	332	73.78%
	Master's and above	23	5.11%
	Others	26	5.78%
Occupation	Students	275	61.11%
	Businessmen	30	6.67%
	Farmers	8	1.78%
	Works	48	10.67%
	Civil servants	14	3.11%
	Others	75	16.67%

5 RESULTS

Confirmatory factor analysis (CFA) is utilized to test the measurement model. Fornell and Larcker ^[26] suggest that Cronbach's Alpha, composite reliability, and average variance extracted (AVE) can be utilized to assess the convergent validity. As shown in Table 3, the composite reliability of all of the constructs is above 0.9, Cronbach's Alpha is above 0.8, and the AVE is above 0.7, which are all beyond their corresponding thresholds. And the results shown in Table 4 that the factor loading of item is above 0.8, which means the item reliability is acceptable. Thus, the results confirmed the high convergent validity of our data. We further tested the discriminant validity of the constructs in our research model. From Table 4, the results showed that the AVE square roots of the constructed are much higher than cross-correlation, it approves high discriminant validity.

Table 3. Internal consistency of model constructs

	AVE	Composite Reliability	Cronbach's Alpha
Image (IMG)	0.905	0.966	0.947
Relationship benefit (RELA)	0.844	0.942	0.907
Perceived advantage(PAD)	0.760	0.927	0.895
Switching cost(SCST)	0.796	0.940	0.915
Compatibility(CMPA)	0.830	0.951	0.932
Popularity(PPUL)	0.811	0.928	0.883
Subjective norm(SBNR)	0.845	0.942	0.908
Intention to use(INT)	0.848	0.957	0.940

Table 4. Factor loadings and cross-loadings for all constructs.

	IMG	RELA	PAD	SCST	CMPA	PPUL	SBNR	INT
IMG1	0.948							
IMG2	0.951							
IMG3	0.956							
RELA1		0.889						
RELA2		0.937						
RELA3		0.929						
PAD1			0.847					
PAD2			0.892					
PAD3			0.862					
PAD4			0.887					
SCST1				0.872				

SCST2				0.905				
SCST3				0.878				
SCST4				0.912				
CMPA1					0.891			
CMPA2					0.936			
CMPA3					0.928			
CMPA4					0.888			
PPUL1						0.849		
PPUL2						0.920		
PPUL3						0.930		
SBNR1							0.931	
SBNR2							0.935	
SBNR3							0.891	
INT1								0.928
INT2								0.942
INT3								0.924
INT4								0.888

Table 5. Square root of AVE and cross-correlations

	IMG	RELA	PAD	SCST	CMPA	PPUL	SBNR	INT
IMG	0.951							
RELA	0.653	0.918						
PAD	0.547	0.507	0.872					
SCST	0.682	0.613	0.461	0.892				
CMPA	0.620	0.652	0.716	0.547	0.911			
PPUL	0.294	0.457	0.506	0.295	0.568	0.901		
SBNR	0.617	0.602	0.518	0.649	0.636	0.449	0.919	
INT	0.585	0.663	0.524	0.606	0.675	0.440	0.657	0.921

Smart PLS 3.0 was used to test the hypotheses in our research model. R^2 value is 0.599, indicating that the independent variables can explain the dependent variable 59.9% change, and the model fits well.

As Table 6 shows, we find that four factors can significantly affect intention to use, which are relationship benefit, switching cost, compatibility and subjective norm, with $\beta=0.24$, $t=3.84$; $\beta=0.15$, $t=2.58$; $\beta=0.27$, $t=3.41$; $\beta=0.27$, $t=4.27$. However, image, perceived advantage, and popularity have no significant effect on intention to use, thus H1, H3, and H6 are not supported.

Table 6 Main test.

	Original Sample (O)	T Statistics (O/STERR)	Hypotheses
IMG -> INT	0.012933	0.219940	H1 (not supported)
RELA -> INT	0.238916**	3.836844	H2 (supported)
PAD -> INT	0.005560	0.100754	H3 (not supported)
SCST -> INT	0.148272**	2.576592	H4 (supported)
CMPA -> INT	0.268401**	3.406418	H5 (supported)
PPUL -> INT	0.028295	0.550511	H6 (not supported)
SBNR -> INT	0.222732**	4.271889	H7 (supported)

Dependent variable : Intention to use(INT), $R^2 = 0.599$;

Note: ** $P < 0.01$

6 DISCUSSION

The findings of study help us understand the personal and environment factors' effect on QQ users' intention to use QQ's various functions. In this study, we analyze the personal factors from three accepts: benefits, costs and habits, we utilize image benefit, relationship benefit, and perceived advantage as the benefits of the QQ users, use switching cost to describe costs and use compatibility to describe habits strength. Then we include two environment factors, popularity and subjective norms into our research model.

The results showed that relationship benefit has positive impact on the intention to use, whereas image benefits and perceived advantage cannot affect the dependent variable. This highlights the importance of maintaining and expanding the relationship. Although QQ can maintain and enhance the personal image, improve people's life quality and work performance, but these two benefits (image & perceived advantage) are

not the main purpose for the users. In terms of switching costs, since there is high switching costs if QQ users change to use other SNS APPs, such as money, time, effort, which increase the exist barriers, so high switching costs will enhance QQ users' usage intention. Besides, If QQ users consider that their life and work habits fit QQ operating design and function layout, they will intend to use various kinds of QQ functions.

Also for environment factors, the results of the study show that subjective norm has significant influence on the willingness to use, but the popularity had no significant effect. That is to say, we find that the pressure of friends, relatives, classmates, colleagues and so on, have a significant impact on our behavior and decision-making. Whereas, without this pressure, just popularity do not have significant effect on people's usage intention.

7 THEORETICAL AND PRACTICAL IMPLICATIONS

This study has various theoretical contributions. First, this study utilizes SCT as the theoretical background, and propose various personal factors and environment factors which may affect users' intention to use QQ' various functions. This is one of the initial study that exert SCT in SNSs. Second, we further partition the personal factors into three groups: benefits (including image benefit, relationship benefit, and perceived advantage), costs (switching cost) and habits (compatibility). We find that only one kind of benefits (relationship benefit) can exert significant effects on SNS APP users' intention; besides, both the cost and habit factors take crucial roles on users' intention. Third, we identity the environment factors in this research from two directions, popularity indicates the users' observations toward other persons, whereas the subjective norms imply other persons' attitude toward the users. The statistical results that only subjective norms have significant effect whereas the popularity is ineffective in this research context.

This study has some practical implications. We find that relationship benefit, switching costs, compatibility and subjective norms have significant effect on users' intention to use various QQ functions. Firstly, we suggest some strategies to enhance the QQ users' relationship benefits. Current version of QQ has various kinds of designs relating to relationship maintenance, such as group activities, group operations, group voting, group check-in, etc., but some of these functions are not being used very frequently. Therefore, we propose the following suggestions: elevating the rank of the individual in the group based on the level of the activity and completion of the task, and then give special authority to individual based on the hierarchy, which can enhance the dependence of individual on groups. Increasing the anonymous comments function in QQ zone, which can enhance mutual intention by improving interaction each other. Secondly, QQ also needs to enhance the users' switching cost, QQ should design more online or offline activity to enhance users' time cost and relationship cost. QQ then also needs increase the switching cost of QQ account, the first step is improving the accounting security by improving the confirmation function of remote logging and QQ payment that is because it can enhance the user's stickiness to QQ which increase the users' switching cost. The second step is expanding the cooperation with other popular sites to enable QQ to log in more websites as a third party identity. Thirdly, in order to improve the compatibility, QQ may need to improve the usability by deleting some functions, for example, the 'nearby people' function is not used usually, and then QQ also needs increase the function of online preview and online alter of loaded files, which can keep in line with people's work habits, we also suggest that the QQ should include more functions of QQ's linked products, such as QQ music, QQ mail, QQ weather, which users can use those products all functions just by QQ service account without downloading those applications.

Finally, there are some limitations in our research. First, as an initial study to explore the antecedent factors of QQ users' intention to use various functions, we only includes seven variables in our research model. We suggest future research to explore and test more variables' effect on SNS users' usage intention. Second, we only

collect data from one SNS APP (QQ), it is not clear if our findings can apply to other SNS APPs.

ACKNOWLEDGEMENTS

This work is supported by the Fundamental Research Funds for the Central Universities (Grant No. JBK1802012, JBK120505), the Fund from Collaborative Innovation Center for the Innovation and Regulation of Internet-based-Finance (Grant No. JBK150407), and Fund from Laboratory for Financial Intelligence and Financial Engineering, Southwestern University of Finance and Economics.

REFERENCES

- [1] December 2016 <Statistics report on the thirty-ninth China Internet development situation> [EB/OL] .
http://www.cnnic.cn/hlwfzyj/hlwzxbg/hlwtjbg/201701/t20170122_66437.htm.
- [2]P.B. Brandtzæg, J. Heim. Why people use social networking sites. A.A. Ozok, P. Zaphiris (Eds.), *Online communities*, LNCS 5621, Springer-Verlag, Berlin Heidelberg (2009), pp. 143–152
- [3]N. Park, K.F. Kee, S. Valenzuela. Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *Cyber Psychology & Behavior*, 12 (2009), pp. 729–733.
- [4]Aneta Przepiorka, Agata Błachnio, Juan F. Díaz-Morales. Problematic Facebook use and procrastination. *Computers in Human Behavior*, Volume 65, December 2016, Pages 59-64
- [5] Bandura, A. (2002). Social cognitive theory of mass communication. In J. Bryant & M. B. Oliver (Eds.), *Media Effects: Advances in Theory and Research* (pp. 94-124). New York, NY: Routledge.
- [6] Bandura, A. (2001). "Social Cognitive Theory of Mass Communication" (PDF). *Media Psychology*. 3 (3): 265–299.
- [7] Ming-Ten, Tsai, Nai-Chang, Cheng. Programmer perceptions of knowledge-sharing behavior under social cognitive theory [J]. *Expert Systems with Applications*, 2010, (37): 8479-8485
- [8]A. Nadkarni, S.G. Hofmann. Why do people use Facebook? .*Personality and Individual Differences*, 52 (2012), pp. 243–249.
- [9]B.R. Schlenker. Self-presentation. M.R. Leary, J.P. Tangney (Eds.), *Handbook of self and identity*, Guilford, New York (2003), pp. 492–518.
- [10]D.J. Kuss, M.D. Griffiths, et al. Internet addiction in students: prevalence and risk factors. *Computers in Human Behavior*, 29 (3) (2013), pp. 959–966.
- [11]Jih-Hsin Tang, Ming-Chun Chen, Cheng-Ying Yang, Tsai-Yuan Chung, Yao-An Lee. Personality traits, interpersonal relationships, online social support, and Facebook addiction. *Telematics and Informatics*, 33 (2) (2016), pp. 102–108.
- [12] Chen, C.-J., & Hung, S.-W. (2010). To give or to receive? Factors influencing members' knowledge sharing and community promotion in professional virtual communities. *Information & Management*, 47(4), 226e236.
- [13] Rui-Ting Huang, 2018, What motivates people to continuously post selfies? The moderating role of perceived relative advantage, *Computers in Human Behavior*
- [14] A.J. Padilla, M. Williams, C. McSorley Switching costs. Part one: economic models and policy implications NERA (2003)
- [15] Nakamura, A., 2012. Estimating switching costs of changing social networking sites. In *Proceedings of the 19th ITS biennial conference, Bangkok 2012: moving forward with future technologies-opening a platform for all*, International Telecommunications Society (ITS).
- [16] Mikołaj Czajkowski, Maciej Sobolewski; 2015, How much do switching costs and local network effects contribute to consumer lock-in in mobile telephony, *Telecommunications Policy*
- [17] Sharpiro, C., H.R. Varian. *Information Rules*. Harvard Business School Press, Boston, MA, 1998
- [18] L. Chen, M.L. Gillenson, D.L. Sherrell Enticing online consumers: an extended technology acceptance perspective *Information & Management*, 39 (8) (2002), pp. 705-719

- [19] J.H. Wu, S.C. Wang What drives mobile commerce? An empirical evaluation of the revised technology acceptance model *Information & Management*, 42 (5) (2005), pp. 719-729
- [20] Etymology Online entry for Popular, 5 April 2009.
- [21] Becker M. Sociometric location and innovativeness: reformulation and extension of the diffusion model. *Am Sociological Rev* 1970; 35: 267–282.
- [22] Seidman, G. (2013). Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, 54(3), 402–407.
- [23] Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, Mass: Addison-Wesley.
- [24] Ajzen, Icek. The Theory of Planned Behavior [J]. *Organizational Behavior and Human Decision Processes*, 1991,50(2):179-211
- [25] Watjatrakul, B. (2013). Intention to use a free voluntary service. The effects of social influence, knowledge and perceptions. *Journal of Systems and Information Technology*, 15(2), 202–220.
- [26] (C.Fornell, D.F. Larcker. Evaluating structural equation models with unobservable variables and measurement error [J]. *J. Mark. Res.* 1981: 39-50.)

Appendix 1

Variable	Items
Image benefit	A. Using QQ can enhance my personal image in my friend circle. B. Using QQ can improve my recognition from the friends. C. Using QQ can make me win high reputation in my friend circle.
Relationship benefit	A. Using QQ can help me keep in good touch with my friends. B. Using QQ can help me make more friends. C. Using QQ can expand my friend circle.
Perceived advantage	A. Using QQ has improved my quality of life. B. Using QQ has improved my job performance. C. Using QQ has provided useful help for my life and work. D. Using QQ has improved my efficiency.
Switching costs	A. If I stop to use QQ, I will lose the reputation accumulated on the QQ. B. If I stop to use QQ, I will lose the friends on the QQ. C. If I stop to use QQ, I will lose the money, time, energy and other costs spent on the QQ.
Compatibility	A. Using QQ matches with my habit of communicating with others. B. Using QQ matches with my lifestyle. C. Using QQ matches with my current environment. D. Using QQ matches with my way of working.
Popularity	A. I have seen other people use QQ. B. In my company or school, many people are using QQ. C. People using QQ can be seen everywhere.
Subjective norms	A. People who have influence on me think that I should use QQ. B. People around me think that I should use QQ. C. The environment around me prompts me to use QQ.
Intention to use	A. I am willing to continue to use the various functions of QQ. B. In a long time I will continue to use the various functions of QQ. C. In the future, I will often use the various functions of QQ. D. When there is a need, I will continue to use the various functions of QQ later.

Study on Enterprises' Internet Public Opinion Area Hotspots

Based on Social Network Analysis

Yuanheng Li, Yan Tu, Xuefeng Li*

School of Information, Central University of Finance and Economics, Beijing, 100081, China

Abstract: With the rapid development of Web 2.0, online public opinion has become an issue in the companies' development process. With numerous user-generated contents about real-world events generated almost in real-time, monitoring, evolution and management of online public opinion play the critical role for the healthy development of enterprises. By collecting articles about public opinion on the corporate network from CNKI and using Citespace based on social network analysis, we have combed the context of current research in this area, analyzed the characteristics of the current research on this topic, excavated research rules in this field and summarized research results to provide references for further study.

Keywords: enterprise, online public opinion, bibliometric analysis, visualization, Social Network Analysis

INTRODUCTION

With the further popularization of the internet, many social web tools, such as BBS, blog, microblog and WeChat, gradually become the important access for internet users to communicate with each other, express views, elicits emotions and share tips. Consumers can express their opinions on all kinds of services and products through variety of social tools. And relevant topics will be disseminated at an extremely rapid rate in the network, thereby continuously attracting new groups to join, forming public opinions in eventually, which will bring unprecedented pressure to the development of enterprises. In addition, the media coverage will also form a strong public opinion that has strong influence on the internet. The opening, interactivity and vitality of the internet have promoted the public opinion events to have a wider influence, spread faster and have more destructive power. In the ever-changing network environment of public opinion, commenting on the online public opinion correctly, carrying out real-time monitoring of online public opinion, guiding the development of the public opinion crisis effectively and eventually deal with online opinion crises successfully, have a great significance of the reputation and health development of enterprises. Dellarocas (2007) argues that online remarks from online users are an important source of information for businesses in making decisions and he suggest that businesses can predict product sales by analyzing online reviews[1].

The study of public opinion in our country began in 2003, of which the research on internet public opinion began in 2008. At present, although scholars have conducted many explorations and researches on the spread of online public opinions, monitoring and early warning as well as guiding and coping with online public opinion, there is still a lack of depth in research and organic combination among various research subjects.

By utilizing Citespace, doing the bibliometric analysis and combing the current research of online public opinion, the paper can sum up the law and direction of the study, further grasp the research status, academic hot spots and the development trend of online public opinion, understand the research structure of the enterprise online public opinion deeply and then explore its research dimensions and theoretical basis systematically, which can provide some reference for the exploration of the theory and method of network virtual social management.

* Corresponding author: **Yan Tu**, Professor, Information School, Central University of Finance and Economics. Research Interest: The electronic commerce; Social network; Knowledge management Email: tuyancufe@163.com.

1. SAMPLE AND DATA COLLECTION

In the CNKI Chinese Academic Literature Network Publishing Database (99.9% acceptance rate), the paper use “enterprise” and “public opinion” as the key words and the periodical range is "all journals" so that 1021 articles are obtained. In the meanwhile, in order to ensure the accuracy and credibility of the research result, and to avoid the interference of unrelated documents, we first screen out some irrelevant documents such as "meeting reports", and then exclude the documents without keywords. Eventually, 453 articles meet the requirement.

2. MEASUREMENT

Co-Word Analysis is a content analysis technique^[2], which mainly analyzes the number of occurrences of a pair of words (enterprise, network public opinion, monitoring and early warning, guidance, etc.) in the same literature in the study of enterprise online public opinion. Based on this, doing the hierarchical clustering of these words can reveal the close relationship between these words and then analyze the structural changes of the subject and topic. Co-word analysis is an important method of information metrology. It is generally believed that the intensity of the relationship between the two topics is proportional to the number of times the representative topic's vocabulary appears in the same document. The higher the frequency, the closer the relationship between the two topics. What's more, scholars are more concerned with the research topic and the subject is usually the hot topics of the subject. The scientific knowledge map based on co-occurrence network analysis can not only explore the evolution of the field, but also visualize the research in this field^[3].

The map of scientific knowledge is an image that takes knowledge domain as the research object and shows the relationship between the development process and structure of a certain scientific knowledge. It has the double nature and characteristics of "graph" and "spectrum", showing the cross, interaction and evolution between knowledge elements and other complex relationships^[4].

3. RESULTS

3.1 Hot spot of research

The keywords can highly summarize papers, which can reflect the key contents of the research results. In this paper, we delete the non-standard key words and merge synonym in order to ensure the effect of common-word analysis. And ultimately, we get 32 keywords whose frequency is more than 7. Table 1 is the top 32 high-frequency keywords in the research field of corporate network public opinion. Through these keywords, we can characterize the hot spots in the research field of corporate network public opinion to a certain extent. The keyword in this field is “online public opinion”, whose frequency is 191. In addition, “public opinion information”, “new media age”, “crisis public relations”.

Table 1. High frequency keyword list

Serial number	Keyword	Frequency	Serial number	Keyword	Frequency
1	Online public opinion	191	17	First time	12
2	Public opinion information	39	18	Social public opinion	12
3	Public opinion	32	19	measure	11
4	New media age	31	20	New Media Environments	11
5	Public opinion analysis	30	21	Big data	10
6	Reputation risk	23	22	Food safety	10
7	Crisis public opinion	23	23	People's Network	10

8	Opinion leader	22	24	Public opinion monitoring	10
9	State-owned enterprises	21	25	Sinopec	8
10	Crisis management	19	26	Reputation management	8
11	self-media	18	27	Monitoring room	8
12	enterprise	17	28	Public opinion monitoring	7
13	Power Enterprise	15	29	Financial Institutions	7
14	new media	13	30	News spokesman	7
15	Corporate image	12	31	Negative information	7
16	hot-spot event	12	32	management	7

3.2 Key words co-occurrence knowledge map

In order to draw keywords co-occurrence map, the paper imports data into Citespace for format conversion and storage, and the node type is selected as "keyword". In addition, the top50 nodes in each time zone are selected, that is, the threshold it top50 and the time slice is set to 2 years. Last, the routing algorithm is used to draw keywords co-occurrence map. As shown in Figure 2, there are 189 nodes and 529 connections in the graph. Since "network public opinion" is the subject of retrieval, a total of 191 times occur. The node is too large to cover the entire network so that we delete the "online public opinion" to make the map clear and reasonable.



Figure 1. Key words co-occurrence knowledge map

By analyzing the frequency and occurrence time of high-frequency keywords, the time lines and hot spots in the field can be judged. The size and color change of the circle shows the importance of a node. The larger the circle is, the larger the number of times the keyword appears. The more vivid of the node color, the later it appears. When a node's intermediary center is large, a purple circle appears. If a node is at the core of the field, it will control the flow of information. The connection indicates the degree of association of the node with other nodes[5]. Integrated node frequency, centrality and relevance, the paper list 10 key nodes which have higher impact in the field of enterprise online public opinion: online public opinion (191,0.49), public opinion information (39,0.14), public opinion (32,0.16), new media age (31,0.07), public opinion analysis (30,0.01), reputation crisis (23,0.10), crisis public relationship (23,0.14), opinion leader (22,0.20), crisis management

(19,0.23), power enterprise (15,0.15). At the stage of generating atlas, there are interaction forces between nodes, and nodes with higher influence can attract the remaining nodes in the vicinity to connect with each other according to the degree of connection with each other, so as to form a common knowledge map of key words and visualize the research direction on enterprise online public opinion.

3.3 Hotspot topic classification



Figure 2. Key words cluster knowledge map

According to the key word co-occurrence knowledge map and cluster time chart analysis, we can divide the hot topics in the study of corporate network public opinion in China into “the online public opinion propagation and monitoring under the new media era (#2, #3, #9)”, “online public opinion guidance and crisis management (#1, #5, #6, #8)”, “financial industry online public opinion discussion (#4, #7)”, “food industry online public opinion discussion (#0)”.

3.3.1 Online public opinion propagation and monitoring under the new media era

In this part, the paper mainly analysis characteristics and modes of dissemination of corporate network public opinion so as to carry out public opinion monitoring, timely warning. This section includes cluster #2, #3, #9. In the era of web2.0, the dissemination of public opinion has become more extensive, public participation has become more popular, and the scope of dissemination has become broader. Under the new media environment, the rules of corporate crisis dissemination begin to appear "long tail effect". In addition, the stakeholders in the corporate crisis have deepened their mutual cooperation under the new media, resulting in the extension of corporate crisis communication process. To improve the ability to guide the public opinion crisis, enterprises need to have a comprehensive understanding of the mode and characteristics of the spread of online public opinion under the new channel of communication environment. For different industries, different public opinion monitoring and early warning systems can be set up: early warning system of network public opinion crisis based on viewpoint tree; design and application of on-line intelligent monitoring system of network public opinion network for power enterprises and research and implementation of public opinion mining system of listed companies based on Hadoop^[6]. In the daily time, what we need to do is just as list: ① Enhance the awareness of public opinion management and awareness of crisis; ②Real-time monitoring,

intelligent data collection and timely analysis of the results of the monitoring; ③Continuously improve the internal information database and internal public opinion monitoring system; ④Establish public opinion monitoring work process and system. Lin Qin and Guo Dongqiang (2017) did a research. Based on the characteristics of communication subjects, according to the principle of system dynamics, this research made sensitivity analysis and comparative analysis about the four factors to put forward the countermeasures to prevent and control the dissemination of enterprise network public opinion^[7]. Business emergencies are events that spread widely among the public and spread negative information associated with dangerous or defective products^[8]. Wang et al. (2014) found that diffusion period, immune period, brand reliance remodeling period, together with the structure of the public social network all had a great impact on the outburst threshold value, maximum of the diffusion, stability level, and fluctuation of public opinion^[9].

3.3.2 Guidance of enterprise online public opinion and management of public opinion crisis

In this part, this paper mainly analysis that how to effectively guide enterprises online public opinion and deal with public opinion crisis, minimize the public opinion on the business losses. This section includes clusters #1, #5, #6, #8. The essence of corporate crisis is not a matter of product attributes, nor is the incident itself, but the public's trust in the reputation of the company, related products, and the overall business^[10]. To effectively guide and control the network public opinion crisis in different industries, unnecessary losses can disappear intangible. What we need to do is: ①Improve the public opinion monitoring system, have an access to accurate sources of information, improve monitoring methods, judge public sentiment and emotional attitudes accurately so as to establish a more realistic monitoring system^[11]; ②Public opinion evaluation techniques. It need to integrate computer technology and literary knowledge to accurately find out the real situation of public opinion so as to predict future public opinion scientifically; ③We should properly handle the relationship between deterring rumors and freedom of opinion, and advocate the combination of online media literacy and legal supervision of public opinion so as to create an open and effective public opinion environment. ④We should do some in-depth study on generation, dissemination and guidance about online public opinion from multidisciplinary perspectives to promote the development of Chinese independent network security and train some high-level network supervisors to meet the government and public demand.

3.3.3 Online public Opinion on Financial Industry

In this part, the paper mainly discusses public opinion monitoring, prevention and control theory, reputation risk in financial industry. This section includes clusters # 4, # 7. The emergence and evolution of the related public opinion in the financial industry will have an important influence on the financial industry and even the macroeconomic operation. That's why financial should be distinguished with other industries deserving special attention. At present, the main objects of the research on internet public opinion in the financial industry include traditional commercial financial institutions and internet financial institutions, among which traditional commercial financial institutions include rural small and medium financial institutions, city commercial banks and large joint-stock banks; And internet financial institutions include P2P institutions, third-party payment Agencies and so on. In order to strengthen reputation risk management, we need to do: ①Emphasize the overall management of reputation risks; ②Do a good job in preventing public opinion risks; ③Improve the coping skills of financial institutions; ④Strengthen positive publicity efforts;

The characteristics of internet financial events include rapid spread of internet and confidential privacy of the financial industry features. Compared with the common public opinion, the public opinion of the Internet financial industry is especially difficult to control due to its intricacy. The public opinion in this area has the following characteristics: ①It's difficult to deal; ②It has many participants; ③Long duration. In view of the management of the public opinion crisis in Internet finance industry. Tan Yunming and Rao Xiao (2015) found that online financial public opinion has four reasons and they also believed that the network financial public

opinion process include information collection, quantitative analysis and early warning of crisis. And they put forward measures to strengthen the sense of crisis, strengthen the organization and leadership, be familiar with the formation and propagation of public opinion^[12].

3.3.4 Online public opinion on food industry

In this part, the paper mainly discusses the components, dissemination, response and management in the food industry. This section includes clusters #0. The development of online public opinion in food industry events are divided into four phases: sprout, sudden, development and blowout^[13], and it happens with public participation, mobilization of opinion leaders, media agenda setting, enterprise response strategy, government, Industry organizations involved. So the public opinion characteristics of the industry include: ①The industry receives a high degree of attention; ②Negative news usually starts in the social network; ③The time and scale of the outbreak are unpredictable; ④It has a long duration and a topic repeatedly discussed. Liu Bowei and Zeng Runxi (2017) pointed out that the components of public opinion on food safety network consists of some main elements like government, enterprises, internet users, food inspection agencies, non-governmental organizations (food experts etc.), the media (platforms such as traditional media, new media and social forums) and some object elements as well as public opinion. What's more, it also includes some elements like mobile phones, Weibo, opinion leaders, social networking sites, specialized forums^[14]. Gao Kuo and Zhu Guiguang (2017) pointed out that with the passage of time, the different phases of the public opinion communication in the food industry show different characteristics. Netizens' behaviors are also influenced by the spiral effect of silence, the Matthew Effect and present an integration trend.

3.4 Research frontier analysis

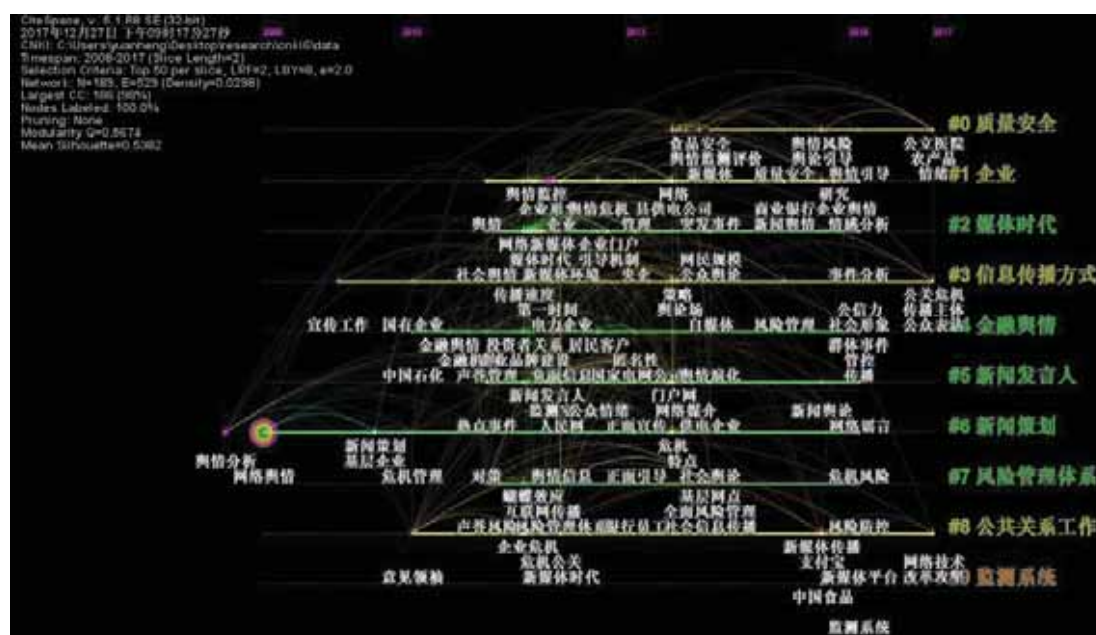


Figure 3. Corporate network public opinion research timeline map

As we can see from Figure 4, corporate network public opinion began in 2008. From 2008 to 2010, it is the initial stage of research on public opinion of enterprises. The research mainly focuses on state-owned enterprises and news planning. From 2011-2013, it is the outbreak time of online public opinion for the corporate. The research involves monitoring, dissemination, guidance, management, in addition to public sentiment, reputation risk. Research continued to be deepened from 2013 to 2016, adding more researches on

food safety, internet users' sentiment analysis and internet financial public opinion in the new media environment. The research directions include corporate public relationship crisis, farm products, and analysis of internet users' sentiment from 2016 to 2017.

Citespace can identify burst words by calculating the rate of change of frequency so that the research frontier can be explored. Burst words reflect the word frequency changes in a period of time greatly, we can find the forefront of research in this area. As shown in the figure below, the top three burst words detected by the burst term function in Citespace are crisis management (intensity 3.73,2010-2012), financial institutions (intensity 2.56,2011-2013), food safety (intensity 2.99,2015 -2017).

Top 3 Keywords with the Strongest Citation Bursts




Keywords	Year	Strength	Begin	End	2008 - 2017
危机管理	2008	3.7306	2010	2012	
金融机构	2008	2.5649	2011	2013	
食品安全	2008	2.9851	2015	2017	

Figure 4. Burst words and parameters of enterprise online public opinion

We can find that the hot spot research is “food safety” from the figure above. Public opinion on food safety network is both an important platform for the public to participate in the management of food safety and a risk factor for social stability. The industry and academia conduct research on food safety public opinion from different angles and construct a response path. Research in the field of food safety includes the evolution, monitoring and responses. Zeng Runxi(2017) collected and sorted the researches on internet public opinion of food safety according to the different types of internet public opinion and model; Then he analyzed and summarized some good ideas and strategies suggested in the related researches. Finally, he pointed out the problems of current researches, such as limited consideration of the stakeholders, life cycle research failing to reflect the particularity of food safety public opinion, etc.

4. DISCUSSION

In this paper, 453 essays obtained from CNKI are processed and analyzed by using Citespace, a scientific knowledge mapping tool. By means of visualization, combing the context of current research in this area, analyzing the characteristics of the current research on this topic, excavating research rules in this field so as to summarize research results and provide references for further study.

The study on public opinion of corporate network in our country began in 2008, and the high-frequency key words in the research field are internet public opinion, public opinion information, new media era, public opinion analysis, reputation risk, crisis public relations, opinion leaders, etc. which represent Scholar's research direction to a certain extent. Key words co-occurrence knowledge map and cluster diagram show that the hot topics in public opinion research include: dissemination and monitoring research of public opinion in the new media era, the guidance online public opinion and management of public opinion crisis, financial industry public opinion and food industry network public opinion. Through the study of the hot timeline and burst term, we can find how the research focus changes. At present, the public opinion of enterprises is in a steady development period, and the dissemination, monitoring and guidance of food safety issues under the new media environment have become the current research hotspots.

Although the study of public opinion of the corporate network has develop greatly, there are still some problems: ①At present, the research of the network public opinion mainly focuses on how to govern the public.

Although there are some researches on the monitoring and early warning of the public opinion on the network, it doesn't combine effectively with other aspects of research;②Under the new media environment, we don't apply the big data, deep learning and other scientific method into early warning and guidance of public opinion well; ③Internet finance is a newly emerging field in recent years. The online public opinion in this field is different from other industry. However, the research of theory and practice on the internet financial public opinion are still in a lack, which need further attention.

ACKNOWLEDGEMENT

This paper is supported by professor Tu and professor Li giving a lot of specific instructions and help. This paper is sponsored by National Social Science Fund of China “Research on Governance Mechanism of Network Society for Public Safety Incidents” (Project No. No. 14AZD045) and Humanities and Social Sciences of Ministry of Education Planning Fund “Research on the Decision-making Behavior of Investors and Borrowers and Risk Control Strategy in P2P online lending” (Project No. 15YJAZH066). Thank all of them very much.

REFERENCES

- [1] Dellarocas C, Zhang X, Awad NF. Exploring the value of online product reviews in forecasting sales: The case of motion pictures[J]. *Journal of Interactive Marketing*, 2007,21(4): 23-45
- [2] Qin H. Knowledge Discovery Through co-cord Analysis[J].*Library Trends*,1999,48(1) :133—159
- [3] An Analysis of Service Government in China based on Knowledge Mapping[J]. Jiang Xiaoping, Su Nan. *Journal of Sichuan University (Social Science Edition)*, 2014(2):98-109 in Chinese
- [4] Review on the Application of CiteSpce at Home and Abroad[J]. Hou Janhua, HuZhigang. *Journal of Modern Information*, 2013(4):99-103 in Chinese
- [5] Scientific Knowledge Mapping-based Visualized Analysis of Studies on Internet Public Opinion in China[J].Zhang Baosheng, Qi Xiaoting. *Information Research*. 2017(05) in Chinese
- [6] The Research and Implementation of Listed Company Public Opinion Mining System Based On Hadoop[D].Wu Jian. *University of Electronic Science and Technology of China*.2013 in Chinese
- [7] System Dynamics Simulation Research of Enterprise Network Public Opinion Dissemination[J]. Lin Qin, Guo Dongqiang. *Information Science*. 2017(04)
- [8] Dawar N, Pillutla M M.Impact of Product-harm Crises on Brand Equity [J]. *The Moderating Role of Consumer Expectations*, 2000, 37(2): 215—226
- [9] Diffusion Simulation of Enterprise Emergency Public Opinion Based on Social Network [J].Wang Delu, Liang Lixin, Sun Jianbin. *Journal of Intelligence*. 2014(10) in Chinese
- [10] Laufer D, Jung J M. Incorporating Regulatory Focus Theory in Product recall Communications to Increase Compliance with a Product Recall [J] .*Public Relations Review*, 2010, 36 (2): 147 -151
- [11] Situation Evaluation of Online Public Opinion on Enterprise Crisis Event[J]. Zhu Ge, Qi Jiayin. *Information Science* 2015(06)
- [12] Public Opinion Guidance Strategy of Network Financial[J]. Tan Yunming, Rao Xiao. *Chongqing Social Science*. 2015(07) in Chinese
- [13] Srivastava S, Lamadrid J G. Extracting an ontology from a document using singular value decomposition[R]. *Association of Computer and Information Science and Engineering Departments at Minority Institutions*, 2001
- [14] On Internet Public Opinion Research of Food Safety in China[J]. Liu Bowei, Zeng Runxi. *Journal of Intelligence*. 2017(06) in Chinese

A Bibliometric Review of Research Trends in Social CRM

Jiangping Wan^{1}, Leqi Xie²*

^{1,2} School of Business Administration, South China University of Technology, Guangzhou, China

Abstract: Social customer relationship management (social CRM) is a relatively new domain, numerous studies have begun to emerge along the border between CRM and social media. This study uses the method of bibliometrics to sort out and analyze the development of social CRM study, and conducts statistical analysis and co-citation network of the literature sample. There has been a significant growth in the social CRM literature with America, China and England being the main contributors. The newest social CRM hotspots focus on the value creation and technical approach of social CRM. The research ideas and conclusions of this paper are of some enlightenment to the academic research and practical application of social CRM.

Keywords: social customer relationship management (social CRM); bibliometrics; social media; co-citation network

1. INTRODUCTION

With the development and popularization of Internet, social networks have changed the traditional customer relationship management ecosystem. Today, many experts use social CRM as a new paradigm of enterprise and customer communication and exchanges, and many companies are adopting social CRM successively, or are putting social CRM into the agenda. Marketing scholars Greenberg's definition of social CRM in 2010 is generally accepted ^[1], social CRM as the integration of customer-facing activities, including processes, systems, and technologies, with emergent social media applications to engage customers in collaborative conversations and enhance customer relationships. While extant literature provides a firm footing upon which to base social CRM research, little research has yet reviewed the social CRM, including Yawised ^[2], Alalwan ^[3]. Yawised etc. (2015) reviewed the academic and practitioner literatures on social CRM, and make a broad comparison between the two literatures, but they do so without a clear picture to describe trends of academic publications on social CRM.

Bibliometrics could provide an answer by applying quantitative analysis and statistics to describe trends of academic publications and their citation counts. It allows us to map the overview of the specific literature, identify the most productive authors within the field and evaluate journal performances. It has been extensively employed to evaluate research trends in various fields ^[4]. Based on bibliometrics, this paper establishes an objective and systematic literature review framework to sort out and analyze the development of social CRM. The specific operation process is shown in Figure 1. This paper statistics the overall growth trend, research region, research domain of social CRM, and then constructs citation clustering with co-citation network, and analysis the research hotspot and the future research direction of social CRM. The conclusion shows that the research scale of social customer relationship management shows an exponential growth trend in recent years, and the research fields are gradually diversified. The latest hot topics in CRM research focus on the value creation and technical methods of social CRM.

2. STATISTICAL ANALYSIS OF SOCIAL CRM LITERATURE

The study was based on data provided by the Web of Science (WoS) Core Collection. The WoS Core Collection is an important database for access to global academic information. We searched the database for literatures using 'social* CRM*' or 'social* customer* relationship* management*' as the subject, and downloaded the relevant literature from 1997 to 2017, followed by a manual initial screening, finally obtained

* Corresponding author. Email: csjpw@scut.edu.cn (Jiangping Wan), xie_leqi@163.com (Leqi Xie)

1055 data records.

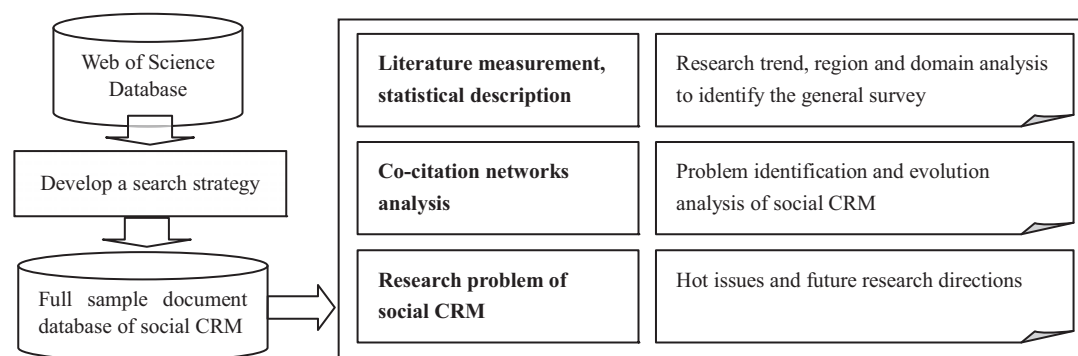


Figure 1. Literature review framework

2.1 The overall growth trend analysis

The growth of scientific knowledge and its rules are closely linked, and the changes in the number of paper directly reflect the changes in the amount of scientific knowledge. This section analyzes the overall growth trend of the research on social CRM by counting the number of published literatures and citation frequency in a period of time.

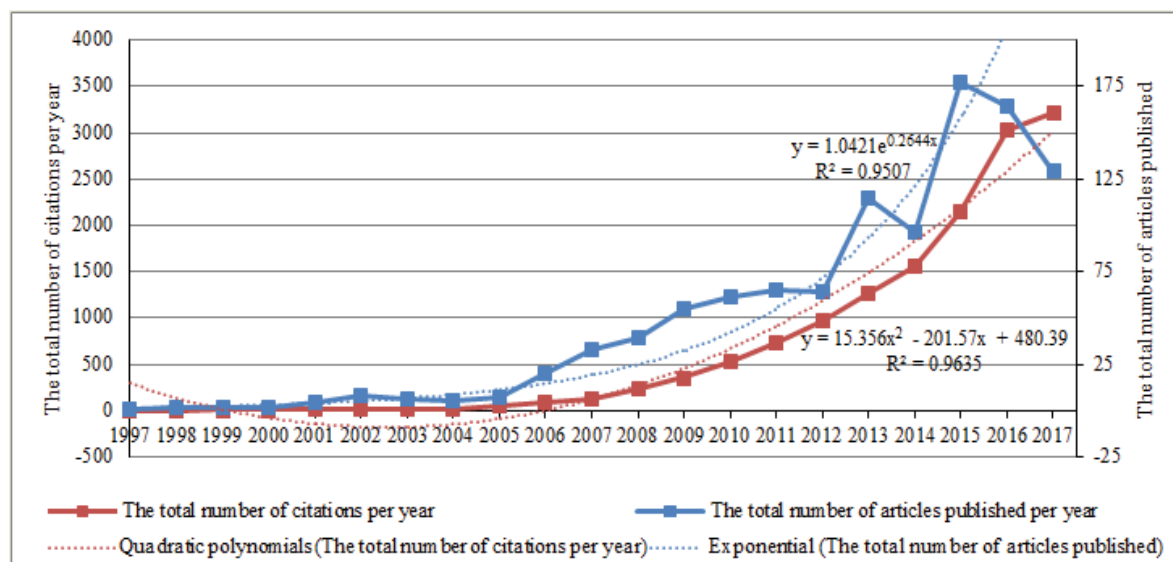


Figure 2. The number of papers published in each year and citation statistics

Figure 2 shows the statistics of the number of published papers and the total number of citations in each year. It can be seen that the large-scale growth of social CRM after 2006 is in line with the scholars' understanding of the craze of social CRM and the development of mobile Internet technology in the early 21st century. In 2015, the number of published papers reached a peak, including 177 articles, and reflecting the high attention of the academic community to social CRM in recent years. At the same time, the cited number of social CRM related literature is also rapidly increasing, reaching 3218 by 2017. In the sample of the literature, the most cited one is [5]. Kevin etc. [5] (2016) studied customer satisfaction partially mediates the relationship between corporate social responsibility and firm market value. As of 2017, this article has been cited 671 times, which is a foundational research in the field of social CRM. According to the curve fitting of the number of articles published and the number of citations each year, we found that the R-squares of the curve fitting are 0.951 and 0.964 respectively. The curve of the total number of articles published each year approximates the exponential function ($Y = ae^{\beta x}$), and the total number of citations per year accords with the trend of quadratic polynomials. This shows that the influence of social CRM in academia continues to grow in recent years.

The exponential growth trend of social CRM research indicates that social CRM research as a newly emerging research field, and gradually transiting from the birth of the discipline to the development of the discipline. The reason is that with the deepening of the global integration process and the popularization and development of mobile Internet technology, the rapid changes in the economic environment, enterprises need to understand and apply social CRM urgent. Scholars gradually realized that the transformation and innovation of enterprise information needs the support of social CRM theory, and the problem of enterprise practice drives the rapid development of social CRM research.

2.2 The region analysis involved research literatures

The regional distribution of the research papers can help researchers understand their international position in the field and find benchmark countries for more effective learning the study. This paper directly uses the retrieval results analysis tools in WoS for regional distribution analysis, and then summarizes the results as shown in Table 1. The results show that in the field of social CRM, China is second only to the United States in the amount of the published literature. Foreign scholars earlier conducted research on social CRM, which mainly focused on the origin and connotation of the research on social CRM, and the relatively mature achievements in the practical significance and application of social CRM [6]. In addition, quantitative research is also gradually enriched, including the content, type, and evaluation of social CRM, etc. China for social CRM research more focused on empirical research, including the application of precision marketing, the application in the automotive industry, the construction of management systems, etc [7]. Compared with other countries, there is still a big gap between theoretical research and practical application in China. In future research, China can learn from the research methods and scientific achievements of the United States in this field so as to enhance its social CRM Analysis and Application Capabilities.

Table 1. Regional distribution of papers related to social CRM in WoS

	Country / Region	Records	Percentage
1	USA	270	25.6%
2	PEOPLES R CHINA	163	15.5%
3	ENGLAND	105	10.0%
4	TAIWAN	68	6.4%
5	SPAIN	64	6.1%
6	AUSTRALIA	62	5.9%
7	GERMANY	55	5.2%
8	SOUTH KOREA	55	5.2%
9	FINLAND	33	3.1%
10	CANADA	32	3.0%

2.3 The disciplines analysis involved research papers

By analyzing the disciplines field of papers, we can effectively grasp the focus of research content and find core research content. In this paper, we analyze the research direction by using the search results analysis tools in WoS. The article selects the top ten results, as shown in Table 2. As the research direction of most of the literature crosses each other, the proportion of all research directions is more than 100% .It is easy to see that the research directions of social CRM are mainly in the areas of BUSINESS ECONOMICS, COMPUTER SCIENCE, SOCIAL SCIENCES OTHER TOPICS , ENGINEERING and other areas. Social CRM research shows significant multidisciplinary and disciplinary integration characteristics.

Table 2. Disciplines distribution of papers related to social CRM in WoS

	Research direction	Records	Percentage
1	BUSINESS ECONOMICS	643	60.9%
2	COMPUTER SCIENCE	194	18.4%
3	SOCIAL SCIENCES OTHER TOPICS	129	12.2%
4	ENGINEERING	115	10.9%
5	OPERATIONS RESEARCH MANAGEMENT SCIENCE	71	6.7%
6	PSYCHOLOGY	52	4.9%

7	INFORMATION SCIENCE LIBRARY SCIENCE	49	4.6%
8	ENVIRONMENTAL SCIENCES ECOLOGY	45	4.3%
9	COMMUNICATION	29	2.7%
10	SCIENCE TECHNOLOGY OTHER TOPICS	29	2.7%

Statistics on the highly cited social CRM literature publishers can further analyze the hot areas of social CRM research and design and the quality of research achievements. This paper uses impact factors to evaluate the combined impact of journals. Impact factor is a commonly used periodical evaluation index in the world. It refers to the total number of citations of papers published in the first two years of a journal divided by the total number of papers published by the journal in these two years. All journals in Table 3 have an impact factor greater than 2 and all journals are SSCI journals. In addition, most of the achievements in social CRM research are concentrated in journals such as JOURNAL OF BUSINESS ETHICS, JOURNAL OF MARKETING and other fields of enterprise economic management and social science theory, which shows that in recent years, the social CRM research has had a high quality research results that impact on the development of theoretical research in economic management.

Table 3. High cited literature published journal statistics

	Cited frequency	Number of published papers	Journal name	impact factor(2016)	H index
1	951	24	JOURNAL OF BUSINESS ETHICS	2.354	13
2	873	9	JOURNAL OF MARKETING	5.318	8
3	632	7	JOURNAL OF SERVICE RESEARCH	6.847	6
4	435	6	JOURNAL OF PRODUCT INNOVATION MANAGEMENT	3.759	4
5	403	7	DECISION SUPPORT SYSTEMS	3.222	7
6	380	8	JOURNAL OF THE ACADEMY OF MARKETING SCIENCE	5.888	6
7	365	8	INFORMATION SYSTEMS	2.777	7
8	362	19	JOURNAL OF BUSINESS RESEARCH	3.354	9
9	352	4	JOURNAL OF OPERATIONS MANAGEMENT	5.207	5
10	322	5	JOURNAL OF INTERACTIVE MARKETING	5.026	4

3. CO-CITATION ANALYSIS OF THE LITERATURE ON SOCIAL CRM

The statistical analysis of social CRM focuses on identifying the most influential disciplines and institutions in the field of social CRM, while co-citation analysis focuses on the contribution and role of specific literature. This section adopts Citespace to construct the citation network of social CRM, and conducts citation clustering analysis to identify the development process and hotspot issues of social CRM research. In order to reduce the number of unnecessary clusters and ensure the visualization effect, the paper was screened twice, retaining only the 10 related categories such as Business, Management, Computer Science Information System, Operations Research and Management Science in the WoS category. The number of subsamples obtained is 844, and the citation network and cluster analysis are performed on the subsamples.

3.1 Citation clustering analysis

Co-citation network based clustering analysis is a specific application of clustering analysis in co-cited domain, mainly refers to the co-cited strength as a basic unit of measure, and the quantitative processing techniques for classifying aggregates of a given set of citations or cited literatures. This technique can aggregate closely related essays into a cluster of literatures, and quantitatively gives the degree of connection between clusters and clusters according to relevant network indicators, and then generates a cluster analysis network of a discipline dissertation^[8]. Citespace first clustered the literatures in different time divisions, and then merged the sub-clusters to form a unified visualization view.

Figure 3 is the social CRM literature citations knowledge network and clustering results. The node size in the figure shows the citation rate. The higher the cited frequency is, the larger the nodes in the figure are. The connection to the two nodes represents that two literatures have been common reference. The colors of nodes and connections correspond to the timeline above Figure 3 (the right side of the timeline represents 2017 and the



Figure 3. Social CRM literature citation network and clustering

left side represents 1997). Figure 4 is a time-domain graph of literature clustering results. The clustering which label is 0,1,3,8 co-citation time is around 2003, and the clustering which label is 2,4,5,6,10 co-citation time in 2010 or so, representing the different stages of social CRM research knowledge frontier. Table 4 summarizes the basic situation of the citation clustering. Clustering labels are feature words extracted according to the title of highly cited literatures in each cluster. The research focuses in Table 5 are summarized based on the frequently occurring keywords in each cluster and the research direction of highly cited literatures. The result of citation clustering shows that there are 9 clusters in social CRM citation network, corresponding to different research hotspots in different periods.



Figure 4. Time-domain graph of literature clustering

Table 4 details the cluster ID, cluster label, size, mean of publication years and research priorities of the nine citation clusters, and the research priorities is based on the frequency of the key words. There are totally 90 literatures in the cluster labeled with 2, 4, 5, 6 and 10. The cited time of these literatures is about 2010, and the research direction is the value creation and brand innovation of social CRM, focusing on the environment and experience of the customer created by social media, and representing the latest research hotspots of social CRM in recent years. For example, Looy^[9] (2016) discussed the use of social CRM and how it can generate business value. Clustering label of 0, 1, 3, and 8 of the literature cited time are around 2003, with 94 articles in total. The research focuses on the business case analysis and service innovation of social CRM, which represent the early

research hot spots. By analyzing the members of the cluster, we can see that there are quite some inspiring achievements in the early social CRM research. For example, Richardson etc.^[10] (2003) focused on case analysis of CRM systems in call centres in the UK, and analyzed their relationship to knowledge management and their use in call centres.

Table 4. Details of citation network clustering to social CRM

Cluster ID	Cluster Label	Size	Mean(Year)	Research priorities
0	Consumer evaluation	32	2002	consumer review, brand experience, managing service innovation
1	Corporate social responsibility activities	25	2003	corporate social performance, social sustainability, business case
2	Supplier resource allocation	24	2008	b2b sales, service support, branding co-creation
3	Customer loyalty	23	2003	satisfaction, emotional value, service innovation
4	Social CRM	22	2009	customer networks, customer relationship management, social media strategy
5	Preferred customer status	17	2007	buyer behavior, customer attractiveness, customer knowledge development
6	Customer engagement	17	2010	online community, consumer brand engagement, SCRM adoption
8	Spectator sport	14	2004	cause-related marketing, stakeholders and market value, customer lifetime value
10	Large network	10	2008	social network analysis, mobile provider, diffusion process

3.2 Research hotspots on social CRM

This paper select the cluster which cluster ID is 4 and cluster label is social CRM, and calculate the highly cited literatures of this cluster by using scientific bibliometric method of citespace, as shown in Table 5. By analyzing these highly cited literatures, the study find that their authors have 4 to 5 papers, including Malthouse, Haenlein, Zhang, Xueming, Dirk, Wang, etc.^[11] These authors are also core authors in the study of social CRM, where there has been cooperation between Haenlein and Zhang, they examined how CRM needs to adapt to the rise of social media, and convergence of social media and CRM creates pitfalls and opportunities, which are explored. The research hotspots of social CRM focus on the value creation and technical approach of social CRM. (1) Value creation. Social CRM can create value for enterprise, such as better understanding of customer needs, more responsive to customer needs and improved customer experience, and more efficient knowledge sharing. Social CRM can give insights, among others, to evaluate and refine an organization's marketing efforts, to proactively build relationships with (potential) customers, and to avoid a crisis. Many scholars use structural equation modeling to study the social CRM, such as Trainor^[12], Hudson^[13] etc. Trainor etc. (2014) analyzed data from 308 organizations using a structural equation modeling approach, and find social media technology usage and customer-centric management systems contribute to a firm-level capability of social customer relationship management. (2) Technical approach. The technology for social CRM mainly can be divided into four categories. First, social listening, monitoring, and analytics technology (i.e., as a traditional helpdesk); second, fan marketing and profile management technology (i.e., as a fan database); third, social sales and marketing automation technology (i.e., to automatically send and manage deals); finally, community and collaboration technology. Kietzmann etc.^[14] (2011) presented a framework that defines social media by using seven functional building blocks: identity, conversations, sharing, presence, relationships, reputation, and groups. As different social media activities are defined by the extent to which they focus on some or all of these blocks, they explained the implications that each block can have for how firms should engage with social media.

Table 5. High cited literatures of social CRM

First author (year)	Literature title	Published journal	Cited frequency
Kaplan A M(2010) ^[6]	Users of the world, unite! The challenges and opportunities of Social Media	Business Horizons	44
Trainor K J(2014) ^[12]	Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM	Journal of Business Research	16

Baird C (2011) ^[15]	From social media to social customer relationship management	IEEE Engineering Management Review	15
Kietzmann J H(2011) ^[14]	Social media? Get serious! Understanding the functional building blocks of social media	Business Horizons	15
Skiera B (2010) ^[16]	The Impact of New Media on Customer Relationships	Journal of Service Research	13
Malthouse E C(2013) ^[11]	Managing Customer Relationships in the Social Media Era: Introducing the Social CRM House	Journal of Interactive Marketing	13

3.3 Future research directions on social CRM

This paper select the cluster which cluster ID is 4 and cluster label is social CRM, and calculate the frontier literatures of this cluster by using scientific bibliometrics method of citespace, as shown in Table 6. Literature for future research directions on social CRM could be categorised into six main themes: social CRM and advertising; social CRM and e-WOM; and adoption of social CRM; social CRM and customers' perception and behavior; social CRM and brand issues; social CRM from the firms' perspective. For instance, a number of studies concentrated on the effective use of social media for word-of-mouth and advertising activities [i.e. 13, 17], there is a need to consider different word-of-mouth and advertising activities could be different from one platform to another. Rosenberger etc. ^[18] (2015) explored 40 user behavior types are compared with actual features of ten social media sites, they found out that reducing the complexity of integrating multiple social CRM systems can improve the adoption of social CRM. Roberts etc. ^[19] (2013) pointed that utilizing input from social CRM can increase innovation project performance, as firms get access to novel market insights and innovative technical solutions.

Table 6. Frontier cited literatures of social CRM

First author (year)	Literature title	Published journal/book/ Conference
Hudson S (2015) ^[13]	The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees	Tourism Management
Rafiee V B (2013) ^[17]	Social Media Marketing: The Unavoidable Marketing Management Tool	21st IBIMA Conference on Vision 2020
Rosenberger M (2015) ^[18]	Conceptualising and Exploring User Activities in Social Media	Open and Big Data Management and Innovation
Dou Y (2013) ^[20]	Engineering Optimal Network Effects via Social Media Features and Seeding in Markets for Digital Goods and Services	Information Systems Research
Roberts D L(2013) ^[19]	Mapping the Impact of Social Media for Innovation: The Role of Social Media in Explaining Innovation Performance in the PDMA Comparative Performance Assessment Study	Journal of Product Innovation Management
Alalwan A A(2016) ^[5]	A Systematic Review of Extant Literature in Social Media in the Marketing Perspective	Social Media: The Good, the Bad, and the Ugly

4. CONCLUSIONS AND INSPIRATION

Bibliometrics information provides scholars with useful indications about the current status of the social CRM research field and its key journals. After analyzing a lot of literature, it could be concluded that research publications relevant to the social CRM have been expanding and gaining impact. Through the citation network and cluster analysis, we can see that social CRM is an interdisciplinary subject. The hot issues of SCRM research are concentrated in the two fields of value creation and technical methods. The literature on these two fields is mainly published after 2010. The future research directions on social CRM include six main themes: advertising, e-WOM, and adoption of social CRM, customers' perception and behavior, brand issues, and social CRM from the firms' perspective.

Through the bibliometrics analysis of this article, we can identify and grasp the development process and frontier hotspots of social CRM research in the past 20 years, provide some references for scholars and practitioners concerned with social CRM research. Due to the limitations of the search database, the sample of this study is mainly confined to the English literature, and the social CRM research on domestic scholars is seldom studied. Further research may compare the major achievements and frontier hotspots of social CRM

research both at home and abroad.

ACKNOWLEDGEMENT

This research was supported by Guangdong Tobacco Monopoly Bureau (company) science and technology projects (Guangdong tobacco science and technology projects 201708).

REFERENCES

- [1] Greenberg, P. (2010). The impact of CRM 2.0 on customer insight. *The Journal of Business and Industrial Marketing*, 25(6), 410–419.
- [2] Yawised K, Marshall P. (2015). Social CRM A Review of the Literature and the Identification of New Research Directions. *International Journal of Virtual Communities & Social Networking*, 7(1):14-20.
- [3] Alalwan A A, Rana N P, et al. (2016). A Systematic Review of Extant Literature in Social Media in the Marketing Perspective. *Social Media: The Good, the Bad, and the Ugly*. Swansea: Springer International Publishing, 79-89.
- [4] Prathap G. (2014). A bibliometric profile of Current Science. *Current Science*, 106(7):958-963.
- [5] So K K F, King C, Sparks B A, et al. (2016) Enhancing customer relationships with retail service brands: The role of customer engagement. *Journal of Service Management*, 27(2):170-193.
- [6] Kaplan A M, Haenlein M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1):59-68.
- [7] Zhang X, Kong X, Kong X. (2014). Model Design and Countermeasure Analysis of Social CRM Systems Based on China's Social Media WeChat. *Advances in Business and Finance*. Hong Kong: Singapore management & sports science, 213-218.
- [8] Chen C, Hu Z, Liu S, et al. (2012). Emerging trends in regenerative medicine: A scientometric analysis in CiteSpace. *Expert Opinion on Biological Therapy*, 12(5): 593–608.
- [9] Looy A V. (2016). *Social Media Management*. Springer Texts in Business & Economics, 87-111.
- [10] Richardson H J. (2003). CRM in Call Centres: The Logic of Practice. Working Conference on Information Systems Perspectives and Challenges in the Context of Globalization. Athens: DBLP, 69-83.
- [11] Malthouse E C, Haenlein M, Skiera B, et al. (2013). Managing Customer Relationships in the Social Media Era: Introducing the Social CRM House. *Journal of Interactive Marketing*, 27(4):270-280.
- [12] Trainor K J, Andzulis J, Rapp A, et al. (2014) Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM. *Journal of Business Research*, 67(6):1201-1208.
- [13] Hudson S, Roth M S, Madden T J, et al. (2015). The effects of social media on emotions, brand relationship quality, and word of mouth: an empirical study of music festival attendees. *Tourism Management*, 47:68-76.
- [14] Kietzmann J H, Hermkens K, McCarthy I P, et al. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, 54(3):241-251.
- [15] Baird C, Parasnis G. (2011). From social media to social customer relationship management. *IEEE Engineering Management Review*, 41(3):48-55.
- [16] Skiera B, Hennig-Thurau T, Malthouse E C, et al. (2010). The Impact of New Media on Customer Relationships. *Journal of Service Research*, 26(3):311-330.
- [17] Rafiee V B, Sarabdeen J. (2013). Social media marketing: the unavoidable marketing management tool. Vienna: 21st IBIMA Conference on Vision 2020, 933-942.
- [18] Rosenberger M, Lehmkuhl T, Jung R. (2015). *Conceptualising and Exploring User Activities in Social Media*. Open and Big Data Management and Innovation. Netherlands: Springer International Publishing, 107-118.
- [19] Roberts D L, Piller F T, Lüttgens D. (2016). Mapping the Impact of Social Media for Innovation: The Role of Social Media in Explaining Innovation Performance in the PDMA Comparative Performance Assessment Study. *Journal of Product Innovation Management*, 33(S1):117-135.
- [20] Dou Y, Niculescu M F, Wu D J. (2013). Engineering Optimal Network Effects via Social Media Features and Seeding in Markets for Digital Goods and Services. *Information Systems Research*, 24(1):164-185.

The Impacts of Environmental Cues and Browsing Experience on Impulse Buying on Social Shopping Website

Yanhong Chen¹, Jingwen Liu^{2}, Si Fan¹*

¹School of Management, Huazhong University of Science and Technology, Wuhan, 430074, China

²Huazhong University of Science and Technology Library, Wuhan, 430074, China

Abstract: Based on the stimulus-organism-response (S-O-R) paradigm, this study develops a model to theorize how environmental cues on social shopping website affect impulse buying behavior from the perspective of browsing experience. Social shopping websites bring a novel online browsing experience for customers. However, how browsing experience on this platform affect customers impulsively have received insufficient academic attention. The results reveal that the unique factors of social shopping website, namely the quality of user-generated content and social presence are critical for facilitating customers' browsing experiences. The results also show that both utilitarian browsing and hedonic browsing experience have a positive impact on customers' urge to buy impulsively. The theoretical and managerial implications are discussed.

Keywords: Social shopping website, Environmental cues, Utilitarian browsing experience, Hedonic browsing experience, Urge to buy impulsively

1. INTRODUCTION

Impulse buying is a sudden, powerful, and persistent urge to purchase something immediately when driven by stimuli encountered within the environment. Online impulse buying behavior has attracted considerable practical and academic attention because it occurs in about 40% of all online purchases^[1], that this proportion may have increased as the development of social commerce. Social commerce, a new trend in e-commerce in which e-commerce is integrated into social media, has enhanced social aspect of online shopping experience^[2]. Unlike tradition e-commerce, consumer can share product ratings, reviews, and recommendations and interact with others in social commerce context. These user-generated content and social interaction may stimulate consumers to buy product impulsively^[3, 4]. According to a report by Greenpeace, 72% of the respondents have a desire to buy the apparel and accessories when they see what others wear on social network sites. However, there are few studies have empirically investigated online impulse buying in social commerce context.

Previous researchers have shown that online browsing can spur online shoppers to purchase products they might not buy otherwise, that is, impulse buying^[1, 5-7]. As an important part of the shopping experience, online browsing is an act of looking around online websites with no intention to buy any specific product^[6, 7]. Social shopping websites (SSW), such as Xiaohongshu.com, Mogujie.com, Kaboodle.com, which are a result of integration of online shopping into a virtual community, are ideally suited for online browsing. On such sites, people can connect with like-minded individuals, exchange opinions on products and recommend the products they like to others. In addition, such sites support user-generated social shopping features, such as recommendation lists, ratings and styles to assist online shopping^[5, 8]. These sites help customers discover new products, brands and trends, which making online browsing more interesting for customers, as compared to simple sites where they purchase by simply clicking on a link^[5]. Given that SSW is much more suitable for browsing rather than directly buying and that may stimulate consumer online impulse buying, we examine customers' impulsive buying behavior on SSW from the perspective of browsing experience.

*Corresponding author. Email: liujw-518@163.com

Previous research on online browsing have identified factors that influence customer browsing experience, such as website atmosphere^[1], individual characteristics^[7]. However, these research did not consider the social factors on the website. SSW supports user generated content, facilitates social interaction among users and enables content sharing with peers, thus may improve customers' browsing experience on SSW. Therefore, we draw upon the stimulus–organism–response (S–O–R) model to theorize how environmental cues (social factors in particular) on SSW improve customers' browsing experience and how such experiences subsequently affect their impulse buying behavior. In doing so, this study offers insights into the role of browsing experience in impacting impulse buying on SSW. Also, this study makes practical contributions on how to enhance browsing experience and lead customers' impulse buying on SSW.

2. THEORETICAL BACKGROUND, RESEARCH MODEL, AND HYPOTHESES

2.1 Stimulus-Organism-Response paradigm

For the analysis of stimulus-driven consumer impulse buying behavior, the S-O-R paradigm has been applied to a number of online impulse buying studies^[9]. The S–O–R paradigm states that a stimulus (S) influences an individual's internal organism states(O), which in turn lead to approaching or avoidance responses (R).

Various environmental cues have been observed to affect a consumer's state that can positively or negatively influence the urge to buy impulsively, such as traditional store atmospherics (e.g. lighting, music), online environment (e.g. website quality, information fit-to-task)^[9]. According to previous research, the following two website environmental cues as stimulus that may influence consumers' online browsing experiences: information quality and social presence. Information quality here refers to the extent to which customers perceive that the information provided by others on a SSW is of high quality. Social presence refers to the feelings obtained by interacting with others in a SSW. Previous research report that information quality of the social content and social presence are essential attributes of a SSW^[10-12]. First, SSW starts with the notion of user-generated content, which means that the majority of information in a SSW is generated by the user. Thus information quality has emerged as an important issue in SSW^[3, 10]. Information quality in SSW influences consumers' responses and experiences as well as motivation to participate in SSW. Second, previous research has shown that customers are able to experience higher levels of social presence on SSWs as compared to e-commerce sites^[11]. Social presence, which is considered as the social aspects of website, is little mentioned as environmental cues in previous research, especially in impulse buying literatures. Therefore, we choose information quality and social presence as the environmental factors that may influence customers' browsing experience and their impulse buying on SSW.

The organisms pertain to internal processes and structures intervening between the stimuli and individuals' final responses, which consist of perceptions, experiences, and evaluations. In this paper, the utilitarian browsing and hedonic browsing experience be viewed as organisms. Park et al. considered two kinds of web browsing experience, namely, utilitarian browsing and hedonic browsing^[6]. Utilitarian browsing is defined as *'acquisition of products through the use of heuristics, goal-oriented behavior, risk reduction strategies, and achievement of information search goals'*^[6]. Compared to utilitarian browsing, hedonic browsing is recreational and is focused on experiencing fun and entertainment. Both utilitarian browsing and hedonic browsing reflects the perceived value derived from browsing experience^[13]. We use Park et al.'s classification to understand browsing behavior in the context of SSW.

The response refers to consumers' final decisions. According to Parboteeah et al., it is problematic to measure actual impulse buying in controlled settings^[9]. Because participants' responses or behaviors are often

biased when they are asked to recall their impulse buying behavior, previous research on impulse buying has used "the urge to buy impulsively" as a surrogate measure for the actual impulse buying behavior^[9]. The urge to buy impulsively is "a state of desire that is experienced upon encountering an object in the environment"^[14]. Therefore, we use the urge to buy impulsively as the response to the stimuli in this study.

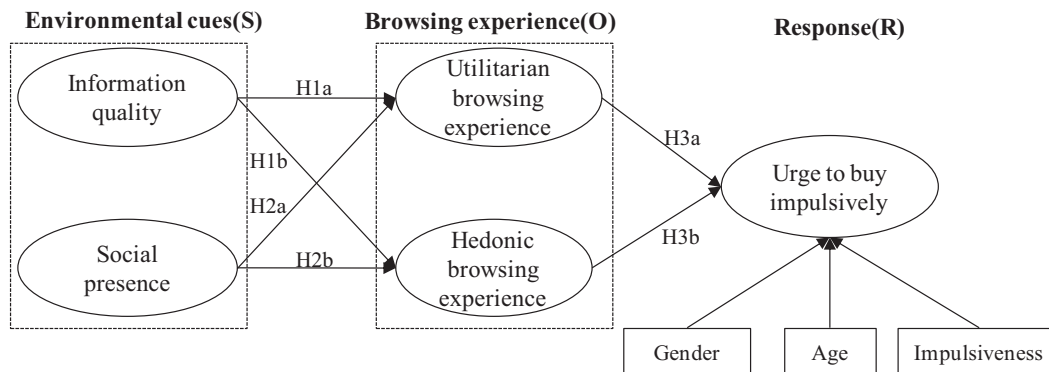


Figure 1. Research model

According to S-O-R paradigm, this paper puts up a theoretical analysis framework shown in Figure 1. The model indicates that information quality and social presence as the environmental cues have positive effects on browsing experience, which in turn lead to urge to buy impulsively. In addition, this study included some control variables that measured customers' characteristics, such as gender, age, and impulsivity trait. The impulsivity trait, which is the inherent impulsiveness of a consumer, defined as "both the tendencies (1) to experience spontaneous and sudden urges to make on-the-spot purchases and (2) to act on these felt urges with little deliberation or evaluation of consequence"^[14], can influence a consumer's propensity to engage in impulse buying. Individuals with high impulsiveness are more likely to experience a stronger urge to buy impulsively.

2.2 Information quality and browsing experience

As mentioned above, the essence of utilitarian browsing activities is obtaining the desired information, the quality of which can increase or decrease one's browsing experience on the website. Previous studies have found that information quality is related to goal-oriented decision making, which enhances users' understanding and decision-making^[10]. For example, Cheung et al. found that the relevance and comprehensiveness of electronic word-of-mouth influences information usefulness, and subsequently, the users' information adoption decisions in online communities^[15]. In the context of SSW, high quality of user generated contents in the website will help users gain a better understanding of social content, reduce the uncertainty associated with the decision making and facilitate goal-fulfillment. Given these arguments, we contend that the high-quality information may facilitate utilitarian browsing experience. Hence, we hypothesize:

H1a: Information quality is positively related to utilitarian browsing experience on social shopping website.

In addition, the quality of user generated content on the website is also entertaining for the users^[16]. Customers are likely to experience greater enjoyment if the quality of information in the website is high. For example, Ahn et al. reported that information quality enhances enjoyment among the users of online shopping sites^[16]. If the quality of information on SSW is low, browsers may feel uncomfortable and may not enjoy the same, whereas if the quality of information is good, users will enjoy surfing and interacting on the website. According to Hsu and Tsou, perceived reliability of blog content can enhance positive emotions in a readers' experiences of the blog^[17]. Hence, we hypothesize:

H1b: Information quality is positively related to hedonic browsing experience on social shopping website.

2.3 Social presence and browsing experience

The social presence of a medium influences perceived usefulness of online product assessments^[18], thus indicating that social presence is instrumental in shaping user's goal-oriented experience. Han et al. indicate that within corporate SNS accounts, social presence is characterized by a utilitarian dimension that has the potential to enhance SNS users' perceived effectiveness of information acquisition^[11]. In addition, previous research has indicated that the presence of a shopping companion in a retail environment significantly reduces the risk perceived by the shopper and enhances the shopping experience^[19]. Perceived social presence on the website may thus have an effect similar to the presence of others offline while shopping and hence would positively influence one's utilitarian web browsing experience in a SSW. Hence, we hypothesize:

H2a: Social presence is positively related to utilitarian browsing experience on social shopping website.

Perceived social presence results in greater enjoyment in the online shopping websites^[18]. Previous studies have demonstrated that companions play a prominent role in enhancing the feelings of fun and enjoyment while shopping^[20]. For example, Arnold and Reynolds suggested that shopping with others is enjoyable and a way to spend time^[20]. These findings may be the same in an online social shopping context. Accordingly, when customers perceive a high degree of social presence on a website, they will experience others as if they were co-present and socially engaged with them, which will in-turn make the shopping experience pleasurable. Therefore, while browsing a SSW, social presence enhances one's experience of fun and pleasure and makes them feel relaxed, thus having a positive effect on one's hedonic browsing experiences. Hence, we hypothesize:

H2b: Social presence is positively related to hedonic browsing experience on social shopping website.

2.4 Browsing experience and urge to buy impulsively

When customers engage in hedonic browsing, they might easily surrender to impulse buying^[1, 5-7]. In the context of e-stores, Floh and Madlberger argued that the enjoyable shopping experience makes customers spend more time browsing the website, which ultimately leads to impulse buying^[1]. This is because, if users browse the SSW for fun, they may be more receptive to environmental stimuli. While browsing they may encounter a trusted recommendation, shared experiences, or new products, which may in-turn trigger an impulse purchase^[3, 4]. Therefore, we propose that there is a positive relationship between hedonic browsing experience and impulsive buying.

H3a: Hedonic browsing experience is positively related to the urge to buy impulsively on social shopping website.

On the other hand, utilitarian browsing has been described by various scholars as rational, efficient and deliberate^[13]. Babin et al. states that hedonic factors impact unplanned shopping behavior, whereas utilitarian factors do not^[13]. In the context of online purchasing of apparel products, Park et al. confirmed that there is a positive relationship between hedonic browsing and impulse buying, while utilitarian browsing is negatively related to impulse buying^[6]. Given that utilitarian browsing serves the function of information acquisition, which is seen as more mission orientated, we propose that there is a negative relationship between utilitarian browsing and urge to buy impulsively. Hence, we hypothesize:

H3b: Utilitarian browsing experience is negatively related to the urge to buy impulsively on social shopping website.

3. METHODOLOGY

3.1 Measurements

A questionnaire was developed for the survey used in this study. Huang and Benyoucef pointed that information quality in social commerce website contains relevancy, accuracy, completeness and update^[12]. In

this study, following Huang and Benyoucef, we propose information quality as a formative construct that includes four dimensions to be examined in the SSW context^[10]: reliability, completeness, timeliness and relevancy. Measures of these four dimensions were adapted from Cheung et al.^[15]. In addition, the measures of social presence were adopted from Gefen and Straub^[21]. Measures of utilitarian browsing and hedonic browsing were adapted from Park^[6] and Grange and Benbasat^[8]. Measures of impulsiveness and urge to buy impulsively were adapted from Parboteeah et al.^[9]. All constructs were measured on a 7-point scale. All items for measuring the constructs were adapted from previous studies to suit the context of SSW, see Appendix A. The scale was originally created in English and translated into Chinese using back-translation technique as the survey respondents were Chinese.

3.2 Data collection

The data was collected using an online survey through a well-known online survey platform Sojump (www.sojump.com) in China. We selected those people who visited the SSW during the last 3 months prior to the study as respondents for our study. In addition, because browsing is an ongoing search, which is independent of any specific purchase^[7]. Therefore, those respondents who came to the websites with the specific intention to purchase immediately were deleted. Thus, in total 276 people participated in this study out of which 32 respondents were deleted, resulting in 244 effective responses. Because the websites in this study mainly focus on apparel and accessory products, the main target customers are female. Of all the respondents, 75.4% were women, and only 24.6% were men. The majority of respondents were below 25 years old (55.3%), and most of them with a relatively high educational level. Table 1 shows the profile of respondents.

Table 1. Profile of respondents

Measure	Items	Frequency	Percentage
Gender	Male	60	24.6%
	Female	184	75.4%
Age	Below25	135	55.3%
	26-30	73	29.9%
	31-40	34	13.9%
	Above40	2	0.8%
Education	High school or below	8	3.3%
	Junior college	53	21.7%
	University	178	73%
	Master or above	5	2%
SSW (respondents may visit multiple SSWs)	xiaoshongshu	138	56.6%
	aiguangjie	90	36.9%
	meilishuo	72	29.5%
	mogujie	80	32.8%
	huabanwang	34	13.9%

We followed Podsakoff and Organ's single factor test to examine the likelihood of common method bias in our study^[22]. The results of exploratory factor analysis showed that the variance explained by the largest factor was 39.8%, which is below the critical value of 50%. The results indicated that common method bias was unlikely to be an issue in our study.

4. DATA ANALYSIS AND RESULTS

Because our research model includes second-order formative constructs, we employed Partial Least Squares (PLS) to analyze the data. The PLS-based SEM technique is better suitable for theory development, and it is suggested to test models that include formative constructs and hence was used in this study.

4.1 Measurement model

First, we examined the reflective constructs for reliability, convergent validity and discriminant validity. As shown in Table 2, Cronbach's alpha ranged from 0.83 to 0.93, thus meeting the recommended minimum of 0.7

and demonstrating reliability of the measures. The average variance extracted (AVE) values for all constructs were larger than the suggested threshold value of 0.5, thus supporting the convergent validity of the measures.

Table 2 also presents the discriminant validity of the constructs. The AVE for each construct should be greater than the corresponding squared correlations between all latent constructs. As can be seen from Table 2, all AVE values meet this criterion, thus demonstrating discriminant validity.

Table 2. Reliability and validities correlation coefficient matrix and roots of the AVEs

Construct	AVE	Composite reliability	Cronbach's alpha	RL	CM	TM	RE	SP	UB	HB	UI	IM
RL	0.83	0.94	0.90	0.91								
CM	0.79	0.92	0.87	0.69	0.89							
TM	0.74	0.90	0.83	0.61	0.64	0.86						
RE	0.79	0.92	0.87	0.66	0.62	0.62	0.89					
SP	0.78	0.95	0.93	0.60	0.60	0.48	0.63	0.88				
UB	0.68	0.91	0.88	0.50	0.50	0.58	0.67	0.55	0.82			
HB	0.78	0.93	0.91	0.51	0.50	0.50	0.63	0.62	0.69	0.88		
UI	0.77	0.91	0.85	0.36	0.39	0.34	0.52	0.44	0.58	0.61	0.88	
IM	0.83	0.95	0.93	0.41	0.38	0.23	0.39	0.42	0.17	0.32	0.38	0.91

Second, we assessed the construct validity and the reliability of information quality, a second-order construct with formative indicators. As shown in Table 3, the weights of the four of the first-order constructs for information quality were significant, thus demonstrating construct validity. From Table 3, The variance inflation factors (VIF) for the first-order constructs were all below the recommended 10 threshold. Therefore, the formative constructs of information quality are reliable and valid.

Table 3. Weights of first-orders on information quality second-order construct

Second-order construct	First-order construct	Weight*	T value	VIF
Information quality	Reliability	0.30	24.99	2.32
	Completeness	0.28	27.71	2.03
	Timeliness	0.27	22.04	2.11
	Relevance	0.32	26.48	2.35

Note: *All weights were significant at $p < 0.001$.

4.2 Structural model

The resulting estimations of structural model testing are shown in Figure 2. Most of the hypotheses were supported excepted H3a, and most of the paths were positively significant at 0.001 level. Meanwhile, the results of control variable show that only impulsiveness have positive effect on browsers' urge to buy impulsively.

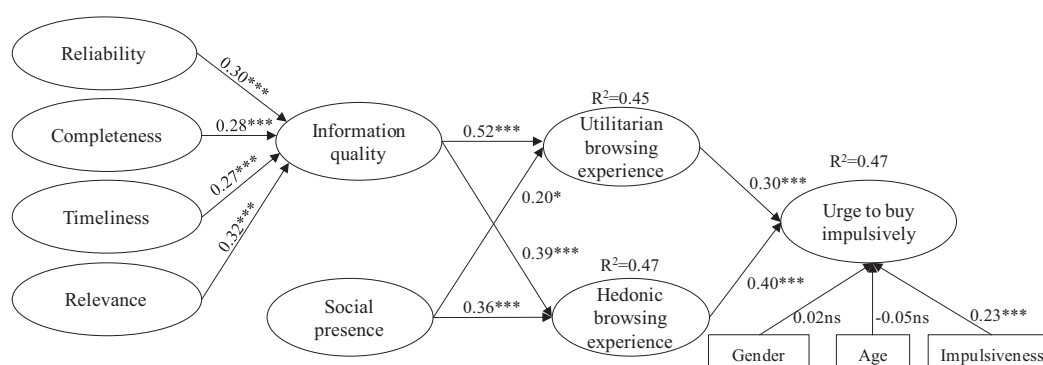


Figure 2. PLS results (Notes: * $p < 0.001$; ** $p < 0.01$; * $p < 0.05$ (two-tailed))**

5. DISCUSSION, CONTRIBUTIONS AND LIMITATIONS

Our findings indicate that a SSW with high-quality user generated content enables users to access superior browsing experience. Information that is reliable, complete, timely and relevant facilitates information seeking and makes browsing more appealing. Apart from the factors of information quality, the results reveal that social presence also influences utilitarian browsing and hedonic browsing experience. As an important feature of SSW, social presence-the feeling of 'being there and warmth' - not only makes information seeking more efficient, but also stimulates and enhances the feeling of fun, pleasure, and enjoyment on a SSW. These findings are consistent with that of previous studies in offline environments that found significant influence on one's overall shopping experience due to other's presence^[19].

Our findings present that hedonic browsing experience positively influence browsers' urge to buy impulsively. This is consistent with previous studies which indicate that hedonic browsing experience has a positive influence on impulse buying behavior both in offline and online environment^[6]. However, contrary to our expectations, our study shows that utilitarian browsing experience is also positively related to the urge to buy impulsively. This may be attributed to social influence in the SSW. According to Xiang et al., due to social interactions, customer are prone to impulse buying behavior on the SSW^[4]. Chen et al. also found that the number of "likes" a post has could increase customers' urge to buy impulsively^[3]. This research shows that the social influence in a SSW can affect customers impulse buying behavior. Although previous studies have shown that utilitarian browsing experience is negatively related to the urge to buy impulsively^[6, 13], they are conducted in traditional e-commerce context and do not consider the influence of a socially interactive environment. Users' decision making is different in the context of social shopping^[2]. The circumstances of SSW are full of stimulus which are not only from website itself, but also from the social interaction and social content^[5]. Browsing on these websites, no matter for information or relaxation, will increase the opportunities for exposure to social stimuli, which may increases the possibility of impulse buying. Clearly, this would be an interesting area to examine in future research.

This study makes contributions for academia and practitioners. From the theoretical perspective, this study enriches the extant literature by providing insights into consumer browsing behavior in a new context, namely social commerce. The SSW presents unique characteristics(e.g. information quality of user-generated content and social presence) that are different from an e-commerce website and provides a customer better browsing experience, which lead to customers' impulsive buying behavior. Unlike Park et al.^[6], this study found that both utilitarian browsing and hedonic browsing experience are positively related to the urge to buy impulsively on SSW. The interesting finding of the positive significant effects of utilitarian browsing experience on impulse buying provide a better understanding and insight to researchers in the field of impulse purchase. The results of this study also provide guidelines for designers and managers of SSW. The designers and managers should give special attention to the value of browsing activity in a SSW. In China, when looking for a specific product, many customers may defer to major e-commerce sites such as Taobao and JD.com rather than the SSWs such as aiguangjie (guang.taobao.com). Thus, in order to attract customer to the website, a SSW can provide different browsing experience to their customers and pay special attention converting web browsers into impulse purchasers. Browsing activities on SSW increase the opportunities for exposure to social stimuli, which in-turn increases the possibility of impulse buying. In addition, website managers should focus on improving the quality of user generated content and pay more attention to social presence. The managers need to provide incentives (e.g. coupons and reputation) to encourage high quality user generated content and set up controlling and screening mechanisms to ensure the quality of content. Moreover, practitioners should take full advantage of social media to increase the feeling of social presence on the website.

The present study has several limitations and requires further examination in the future. First, this study

was conducted in China. Future study should further test the findings in different contexts and cultural environments to cross-validate our results. Second, the samples were primarily female, which may generate gender bias. In order to improve the generalizability, future studies should extend this study to other shopping websites and include gender neutral data sample for analysis. Finally, we only considered the most obvious characteristics of SSW. There may be other factors that would influence user's browsing experience [7]. In this regard, future research may include more factors (such as visual appeal, consumer characteristics etc.) in the research model in order to enrich our understanding of browsing behavior on a SSW.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71332001.

REFERENCES

- [1] Floh A., M. Madlberger, The role of atmospheric cues in online impulse-buying behavior, *Electronic Commerce Research and Applications*, 12 (2013) 425-439.
- [2] Wang C., P. Zhang, The evolution of social commerce: The people, management, technology, and information dimensions, *Communications of The AIS*, 31 (2012) 5.
- [3] Chen J.V., B.-c. Su, A.E. Widjaja, Facebook C2C social commerce: A study of online impulse buying, *Decision Support Systems*, 83 (2016) 57-69.
- [4] Xiang L., X. Zheng, M.K.O. Lee, D. Zhao, Exploring consumers' impulse buying behavior on social commerce platform: The role of parasocial interaction, *International Journal of Information Management*, 36 (2016) 333-347.
- [5] Olbrich R., C. Holsing, Modeling consumer purchasing behavior in social shopping communities with clickstream data, *International Journal of Electronic Commerce*, 16 (2011) 15-40.
- [6] Park E.J., E.Y. Kim, V.M. Funches, W. Foxx, Apparel product attributes, web browsing, and e-impulse buying on shopping websites, *Journal of Business Research*, 65 (2012) 1583-1589.
- [7] Rowley J., 'Window' shopping and browsing opportunities in cyberspace, *Journal of Consumer Behaviour*, 1 (2002) 369-378.
- [8] Grange C., I. Benbasat, 2010, Online Social Shopping: The Functions and Symbols of Design Artifacts, in: 43rd Hawaii International Conference on System Sciences(HICSS), pp. 1-10
- [9] Parboteeah D.V., J.S. Valacich, J.D. Wells, The Influence of Website Characteristics on a Consumer's Urge to Buy Impulsively, *Information Systems Research*, 20 (2009) 60-78.
- [10] Zheng Y., K. Zhao, A. Stylianou, The impacts of information quality and system quality on users' continuance intention in information-exchange virtual communities: An empirical investigation, *Decision Support Systems*, 56 (2013) 513-524.
- [11] Han S., J. Min, H. Lee, Building relationships within corporate SNS accounts through social presence formation, *International Journal of Information Management*, 36 (2016) 945-962.
- [12] Huang Z., M. Benyoucef, From e-commerce to social commerce: A close look at design features, *Electronic Commerce Research and Applications*, 12 (2013) 246-259.
- [13] Babin B.J., W.R. Darden, M. Griffin, Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value, *Journal of Consumer Research*, 20 (1994) 644-656.
- [14] Beatty S.E., M.E. Ferrell, Impulse buying: Modeling its precursors, *Journal of Retailing*, 74 (1998) 161-167.
- [15] Cheung C.M., M.K. Lee, N. Rabjohn, The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities, *Internet Research*, 18 (2008) 229-247.
- [16] Ahn T., S. Ryu, I. Han, The impact of Web quality and playfulness on user acceptance of online retailing, *Information & Management*, 44 (2007) 263-275.

- [17] Hsu H.Y., H.-T. Tsou, Understanding customer experiences in online blog environments, *International Journal of Information Management*, 31 (2011) 510-523.
- [18] Hassanein K., M. Head, Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping, *International Journal of Human-Computer Studies*, 65 (2007) 689-708.
- [19] Lindsey-Mullikin J., J. Munger, Companion Shoppers and the Consumer Shopping Experience, *Journal of Relationship Marketing*, 10 (2011) 7-27.
- [20] Arnold M.J., K.E. Reynolds, Hedonic shopping motivations, *Journal of retailing*, 79 (2003) 77-95.
- [21] Gefen D., D.W. Straub, Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services, *Omega*, 32 (2004) 407-424.
- [22] Podsakoff P.M., D.W. Organ, Self-Reports in Organizational Research: Problems and Prospects, *Journal of Management*, 12 (1986) 531-544.

Role of Danmu Function in User Experience and Engagement:

A Double-edged Sword

Si Fan¹, Jingwen Liu^{2}, Ting Zhu¹*

¹College of Management, Huazhong University of Science and Technology, Wuhan, 430074, China

²Huazhong University of Science and Technology Library, Wuhan, 430074, China

Abstract: Digital video sites currently seek new ways to build a platform to encourage user engagement and enable users to connect with one another through multiple means and channels. The widespread introduction of danmu technology has aroused concerns of academia. However, comparatively little work has been done to explicate the role of danmu function and research progress has been falling behind practical interest. In this study, we seek to investigate user experience and engagement to understand the double-side effect of danmu technology. Using S-O-R framework, we propose that danmu features, namely proximity and coherence, play essential roles in creating augmentation experience of para-social interaction and degradation experience of goal impediment and information overload. Additionally, we investigate user experience to understand user engagement in the danmu-enabled situation. We suggest that para-social interaction will elicit increased user engagement; yet goal impediment and information overload will lead to less user engagement.

Keywords: danmu features, para-social interaction, goal impediment, information overload, user engagement

1. INTRODUCTION

With the evolution of mobile technology and portable devices, digital videos have become the ideal way to satisfy the information and entertainment needs of individuals. More than 62 percent of the global Internet users have watched streaming or downloaded video content in 2017 ^[1]. According to a report of China Internet Network Information Center (CNNIC), the number of digital video viewers has achieved 565 million in China, representing three quarters of the Internet users ^[2]. In addition, it has been estimated that nearly 299 million people in China will watch digital videos regularly via a subscription streaming service in 2018 ^[3]. In the process of booming development of video market, strategies to attract potential users and retain old users have become overwhelmingly important for platforms. The widespread implementation of Danmu function to enhance user experience on digital video platforms has aroused general concern.

Danmu has been defined as an augmented type of review with two unique characteristics ^[4]. In contrast to traditional review function, danmu comments are integrated with the reviewed objects (such as video, pictures, text, etc.) and organized by the attributes or elements of the objects rather than the order of input or popularity. In the case of video watching, danmu comments are projected onto the video screen, somewhat similar in appearance to the film subtitles. However, not quite identical with subtitles, text of danmu can slide across or float above all parts of the video screen. In addition, danmu comments are projected according to the playback time of the video that these comments were input during viewing. Currently, the application for danmu has not been limited to digital videos. Figure 1 presents three screenshots of danmu-enabled objects (i.e., a video, a novel app and a map app). Despite danmu's prevalence, some platforms just blindly follow the trend of the introduction of this technology, hardly knowing the actual role of danmu function in user experience and behaviors.

* Corresponding author. E-mail: liujw-518@163.com(Jingwen Liu) , fansi@hust.edu.cn(Si Fan) , tzh@hust.edu.cn(Ting Zhu)



Figure 1. Screenshots of danmu-enabled objects

Danmu function has been referred to as part of the viewing experience^[5]. Ostensibly, the directly projection of user comments onto the reviewed objects (i.e., video, text, picture etc.) might be regarded as an intrusion of viewing experience and could cause the problem of information overload. However, previous research has indicated that danmu creates a feeling of ‘virtual liveness’^[5] as well as an experience of co-viewing^[4]. Furthermore, Liu et al. proposed a conceptual model to explore the impacts of danmu technology on user experience and they draw a conclusion that apart from being a play experience danmu function and its content can also be a distraction problem^[6]. By and large, danmu technology does indeed exert a two-side impact on user experience and behaviors; however, so far, there are not so much existing researches examining the dual roles of danmu to clarify the relationships among danmu functions, user experience and behaviors.

Accordingly, this study aims to investigate user experience and engagement to understand the double-side effect of danmu technology. There are two theoretical significances of the present paper: first, we propose two concepts to describe danmu function, i.e., proximity and coherence. The concept of proximity is used to capture the closeness that one feels toward other users, both temporally and spatially, which reflects the ability of danmu to create a sense of ‘virtual liveness’. In addition, we use the concept of coherence to capture the perceptions of users with regard to the orderliness and consistency of contents and structures of danmu-enabled objects, which reflects the degree of integration of danmu and the reviewed object. Second, we explore the role of danmu function in user experience and engagement from two perspectives, i.e., augmentation and degradation, and provide some following studies in the theoretical basis and the managerial implication. A Stimuli-Organism-Response(S-O-R) model is used as a framework for this study to examine the relationships among danmu function, user experience and engagement (as shown in Figure 2).

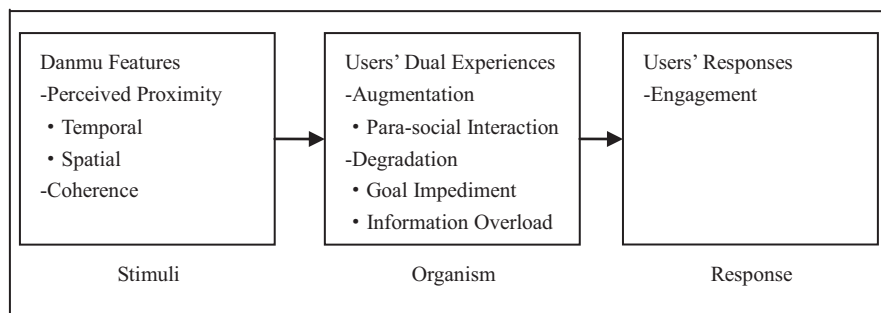


Figure 2. Theoretical framework

2. THEORETICAL BACKGROUND

2.1 Proximity and coherence as stimuli

The viewing environment afforded by danmu has an influence on users' visual experiences^[4]. There are two prominent aspects of danmu features^{[4], [5]}: one is the synchronization of danmu comments to the targeted content and the other is the direct projection of danmu comments onto the targeted content. The posting of danmu comments seems to be happening 'live' during a video viewing experience because every time the video clip is playing, danmu comments appear in the video display screen at the same point of playback time that they were submitted during viewing by their authors. Danmu comments are projected in this way so as to address the video content or respond to other comments, which makes it possible to facilitate embedded interactions among viewers^[5]. This form of commenting lends a sense that viewers can be engaged in social interactions simultaneously during viewing despite 'real' differences in time among users. Based on the above discussion of danmu, we introduce two constructs to conceptualize these unique characteristics of danmu, i.e., perceived proximity and coherence.

Perceived proximity refers to the interpersonal closeness that one feels toward others^[7]. In this article, we continue along this trail but conceptualize proximity as two-dimensional: temporal proximity and spatial proximity. In the social realm, different things may be proximally related in time or space^[8]. Temporal proximity refers here to the temporal closeness that one feels toward other users during viewing. It facilitates viewers' social experience with a sense of simultaneity and an illusion of interaction with other viewers. And the other dimension of proximity captures one person's perception of how spatially close others are. To sum up, danmu commenting can invoke a sensation that one is consuming the danmu-enabled objects (such as video, pictures, text, etc) alongside other users, which is referred as a feeling of co-experience^[4]. As more and more users contribute to the danmu comment feed, a feeling of participation may be reinforced. In a social constructivist way of thinking, perceived proximity is a perceptions product which derives from and may be strengthened by social interactions^[7]. In this study, we suggest that danmu can create for one person a sense of proximity to generate an intense perception of social visual experience.

Coherence used to be referred as the capability of virtual space to provide orderly and consistent structures, contents, and multimedia components in the space^[9]. Accordingly, we define coherence as the ease of understanding of the organization of danmu-enabled objects and postulate that coherence refers here to one person's perception of the orderliness and consistency of contents and structures of danmu and danmu-enabled objects. Previous studies found that a virtual space with well connected and unifying design components rendered better comprehension among users and made them feel more control^[9].

2.2 Para-social interaction, goal impediment and information overload as organism

The relationship between environment stimuli and users' responses is mediated by users' experience. Study has indicated that the sense of virtual liveness of watching a danmu video composes of multiple experiences^[5]. The property of danmu might be seen as the temporal congruence between the point in time of the viewing of the danmu comment by subsequent users and that of it being submitted. More specifically, adding user comments to content, i.e., videos, pictures, text, etc., enhances user experience by creating a sense of 'here and now' of viewing alongside others. Furthermore, we can think of danmu comments as supplementary information provided by other users for the danmu-enabled object, indexical not only of the object itself but also of the submitters of danmu. However, since this form of commenting presents the idea of a crowd of viewers, the text of danmu might be crowded as well and turn to be the potential for obstructing the object and preventing one person from viewing the object clearly. Additionally, we propose that, in certain circumstances, individuals

might enter an information-overload situation with too many comments presented with the danmu-enabled object.

Given the virtual liveness^[5] and pseudo-synchronicity^[4] induced by danmu, social viewing or reading experience through danmu-enabled object can be understood as para-social interaction (PSI). PSI is also called imaginary social interaction, pseudo-interaction and pseudo-friendship, which makes users not mere passive viewers but also involved and active parts of the viewing process^[10]. A PSI experience has been defined as an immediate feeling or impression derived from users' automatic mindreading activities and can provide the users with a sensation of being engaged in a real, lively social interaction with others^[11]. The present paper proposes that user experience of danmu-enabled object can be augmented through PSI.

The crowded and moving text of danmu comments, in the other way, might be seen as information noise, interrupting users' activities and causing cognitive overload. Intrusion of an interruption into a viewer's task leads to psychological reactance and results in user avoidance^[12]. In the case of danmu, the more closeness one person feels towards other users, the more likely he or she feels intrusion of other users' danmu communication into his or her viewing or reading activities. Additionally, the more one person feels the orderliness and consistency of contents and structures of danmu-enabled objects, the more likely he or she could not distinguish desired content from the unwanted, which can also result in a sense of intrusion. We identify two elements contributing to the degradation of user experience, i.e., perceived goal impediment and information overload.

Perceived goal impediment refers here to a hindrance in achieving users' goals (i.e. viewing or reading). In the field of advertising research, perceived goal impediment plays an important role in negative attitudes, aggravation, and ad avoidance^[13]. When danmu comments are regarded as a significant source of nuisance or noise, hindering user efforts to view or read the accompanying object, they can disrupt users' viewing or reading activity, distract users from the object's editorial integrity, and interfere with users' access for desired information. For instance, individuals might feel that the learning process is difficult when they are going to watch danmu-enabled computer science education videos because danmu comments distract them from the video content.

Information overload describes the state that information is afforded beyond one person's need and results in the perception of being overwhelmed^[14]. Danmu comments provide users with the illusion of viewing or reading together with others and help build an artificial atmosphere of lively communication. In this case, users are required to process the information of both the object content and danmu simultaneously, which may make users feel stressed and result in avoidance from the source of the interference.

2.3 User engagement as response

The concept "engagement" has caused a tremendous amount of disagreement and debate across industry and academia. Engagement has been referred to scholars as the state of being interested in something, involved, occupied, or as mental models, attitudes, goals, and users' activities^[15]. This study adopts a behavioral engagement perspective, building on the definition of engagement as behavioral manifestations resulting from motivational drivers^[16]. In the field of online social platform research, behavioral engagement has been referred to as active participation and word-of-mouth referral^[17]. Accordingly, we use engagement construct to capture perceptions of users with respect to active participation as well as word-of-mouth referral of danmu technology.

3. RESEARCH MODEL AND HYPOTHESES

Figure. 3 depicts the research model of this study. It shows that user engagement is determined by para-social interaction, perceived goal impediment and information overload. Proximity and coherence exert impacts on user experience. The interrelationships of the constructs are addressed as follows.

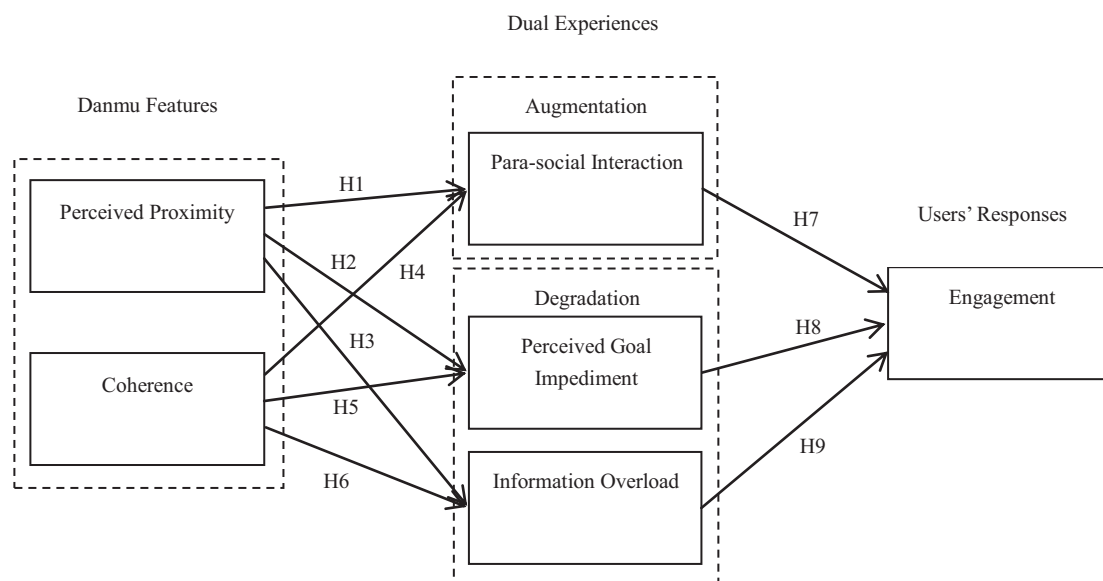


Figure 3. Research model

3.1 Effects of proximity

Previous research found that proximity plays an important role in computer-based communication due to its impacts on individuals' feeling of being observed or not by others^[18]. In addition, proximity assists in identifying para-social relationships, the degree of which also affects the development of para-social relationships^[19]. Immediacy learning course scholars have found that temporal proximity led to better learning experience resulting in more participants' engagement^[20]. Prior study also found that perceived spatial proximity contributed to coordination, shared awareness and mutual understanding^[21]. A para-social experience is accompanied by a sense of mutual awareness with the target object (e.g., a TV performer)^[11]. Thus, we suggest that perceived proximity has a positive impact on para-social interaction.

(H1) Perceived proximity is positively correlated with para-social interaction.

Proximity is used to reflect the actual physical or conceptual closeness^[19] and has been found to affect individuals' feelings and behaviors^[18]. When individuals do not feel the proximity to other users, they may consider danmu comments as a pure interruption: not only preventing them from goal attainment but also creating an information-overload experience. Once individuals have an artificial feeling that other users are in close proximity, they may start to experience para-social interactions. Prior study has claimed that social interactions provide useful information and facilitate goal attainment^[22]. Para-social interaction acts in a similar way with real social interaction despite of the non-reciprocal relationships among 'interacting parties'^[11]. Therefore, we propose that proximity would reduce user goal impediment and information overload. However, beyond a certain point, social interactions might cause information overload problem and began to deter individuals from their goal attainment^[22]. By and large, in the case of danmu-mediated communication, we suggest that proximity negatively correlate with one person's goal impediment and information overload yet with an increased marginal effect. Hence, we hypothesize:

(H2) There is a non-linear relationship between perceived proximity and perceived goal impediment, such that (a) appropriate level of proximity can mitigate perceived goal impediment (i.e., negative linear term), but (b) undesirably high level of proximity has an increased marginal effect on perceived goal impediment (i.e., positive squared term).

(H3) There is a non-linear relationship between perceived proximity and information overload, such that (a) appropriate level of proximity can mitigate information overload (i.e., negative linear term), but (b) undesirably high level of proximity has an increased marginal effect on information overload (i.e., positive squared term).

3.2 Effects of coherence

Coherence reflects one person's perceptions of the orderliness and consistency of danmu and the accompanying objects in terms of content and structure. A virtual space with well connected and unifying design components rendered better comprehension among users and made them feel more control^[9]. Coherence of danmu and the reviewed object may also facilitate better comprehensions among users and lead to shared understanding. Thus, we suggest that coherence positively affect users' perceptions of para-social interaction.

(H4) Coherence is positively correlated with para-social interaction.

On the one hand, coherence makes individuals feel control^[9], which reduces the perceptions of goal impediment and information overload. But, on the other hand, high coherence implies the high level of integration of danmu and the reviewed objects. In this situation, individuals may be overwhelmed because they have to process the wealth of information obtained from object content (i.e. videos, pictures, text, etc.) and danmu simultaneously, which may prevent them from goal attainment and exceed their ability to view or read. Thus, we hypothesize:

(H5) There is a non-linear relationship between coherence and perceived goal impediment, such that (a) appropriate level of coherence can mitigate perceived goal impediment (i.e., negative linear term), but (b) undesirably high level of coherence has an increased marginal effect on perceived goal impediment (i.e., positive squared term).

(H6) There is a non-linear relationship between coherence and information overload, such that (a) appropriate level of coherence can mitigate information overload (i.e., negative linear term), but (b) undesirably high level of coherence has an increased marginal effect on information overload (i.e., positive squared term).

3.3 Effects of dual experiences

Previous studies have found that stronger para-social interaction led to higher social norms and greater enjoyment^[11]. Accordingly, we suggest that viewers experience para-social interaction will generate more engagement. Both perceived goal impediment and information overload measure the degradation aspect of user experience. When danmu interrupts one person's goal, it may result in negative attitudes and lessen their engagement. In addition, individuals may feel stressed because the information provided by danmu-enabled objects is far beyond their abilities to view or read which could result in a retreat from the stressor (i.e., the danmu-enabled objects). Therefore, we expect:

H7: Para-social interaction is positively correlated with engagement.

H8: Perceived goal impediment is negatively correlated with engagement.

H9: Information overload is negatively correlated with engagement.

4. RESEARCH METHODOLOGY

We will conduct an online survey to test the hypotheses. Only those who had danmu-mediated experience are targeted as respondents. We plan to recruit 400 participants and they will be given monetary rewards for their participations.

REFERENCES

- [1] eMarketer. (2017). Worldwide Digital Video Viewers: eMarket's Estimates for 2016-2020. <https://www.emarketer.com/Report/Worldwide-Digital-Video-Viewers-eMarketers-Estimates-20162020/2001983>.

- [2] CNNIC. (2017). The 40th Statistical Report on Internet Development in China. In (pp. 45).
- [3] eMarketer. (2018). China's subscription OTT video market is blooming. www.emarketer.com/content/the-subscription-video-on-demand-market-in-china-is-booming.
- [4] Fan S, Lu Y, Zhao L, Pan Z. (2017). You Are Not Alone: the Impacts of Danmu Technological Features And Co-experience On Consumer Video Watching Behavior. PACIS 2017.
- [5] Johnson D. (2013). Polyphonic/Pseudo-synchronic: Animated Writing in the Comment Feed of Nicovideo. *Japanese Studies*, 33(3): 297-313.
- [6] Liu L, Suh A, Wagner C. (2017). Who is with You? Integrating a Play Experience into Online Video Watching via Danmaku Technology. In: Kurosu M, ed. *Human-Computer Interaction. Interaction Contexts: 19th International Conference, HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part II*, Cham: Springer International Publishing, 63-73.
- [7] Boyer O'leary M, Wilson J M, Metiu A. (2014). BEYOND BEING THERE: THE SYMBOLIC ROLE OF COMMUNICATION AND IDENTIFICATION IN PERCEPTIONS OF PROXIMITY TO GEOGRAPHICALLY DISPERSED COLLEAGUES. *MIS Quarterly*, 38(4): 1219-1243.
- [8] Faulkner P, Runde J. (2013). TECHNOLOGICAL OBJECTS, SOCIAL POSITIONS, AND THE TRANSFORMATIONAL MODEL OF SOCIAL ACTIVITY. *MIS Quarterly*, 37(3): 803-818.
- [9] Lee Y, Chen A N K. (2011). Usability Design and Psychological Ownership of a Virtual World. *Journal of Management Information Systems*, 28(3): 269-307.
- [10] Shin D-H. (2016). Do Users Experience Real Sociability Through Social TV? Analyzing Parasocial Behavior in Relation to Social TV. *Journal of Broadcasting & Electronic Media*, 60(1): 140-159.
- [11] Hartmann T, Goldhoorn C. (2011). Horton and Wohl Revisited: Exploring Viewers' Experience of Parasocial Interaction. *Journal of communication*, 61(6): 1104-1121.
- [12] Zanjani S H A, Diamond W D, Chan K. (2011). Does Ad-Context Congruity Help Surfers and Information Seekers Remember Ads in Cluttered E-magazines?. *Journal of Advertising*, 40(4): 67-84.
- [13] Cho C-H, Cheon H J. (2004). Why do people avoid advertising on the internet?. *Journal of Advertising*, 33(4): 89-97.
- [14] Cenfetelli R T, Schwarz A. (2011). Identifying and testing the inhibitors of technology usage intentions[J]. *Information Systems Research*, 22(4): 808-823.
- [15] Pagani M, Mirabello A. (2011). The influence of personal and social-interactive engagement in social TV web sites. *International Journal of Electronic Commerce*, 16(2): 41-68.
- [16] Van Doorn J, Lemon K N, Mittal V, Nass S, Pick D, Pirner P, Verhoef P C. (2010). Customer Engagement Behavior: Theoretical Foundations and Research Directions. *Journal of Service Research*, 13(3): 253-266.
- [17] Fang J, Zhao Z, Wen C, Wang R. (2017) Design and performance attributes driving mobile travel application engagement. *International Journal of Information Management*, 37(4): 269-283.
- [18] Pinsonneault A, Heppel N. (1997). Anonymity in Group Support Systems Research: A New Conceptualization, Measure, and Contingency Framework. *Journal of Management Information Systems*, 14(3): 89-108.
- [19] Yuan C L, Kim J, Kim S J. (2016). Parasocial relationship effects on customer equity in the social media context. *Journal of Business Research*, 69(9): 3795-3803.
- [20] Haythornthwaite C, Kazmer M M, Robins J, Shoemaker S. (2000) Community development among distance learners: Temporal and technological dimensions. *Journal of Computer - Mediated Communication*, 6(1).
- [21] Shadiev R, Hwang W-Y, Huang Y-M, Yang Y-S. (2015). Study of using a multi-touch tabletop technology to facilitate collaboration, interaction, and awareness in co-located environment. *Behaviour & information technology*, 34(10): 952-963.
- [22] Zhang C, Phang C W, Wu Q, Luo X. (2017). Nonlinear Effects of Social Connections and Interactions on Individual Goal Attainment and Spending: Evidences from Online Gaming Markets. *Journal of Marketing*, 81(6): 132-155.

Self-disclosure in Social Network Sites: An Integration of Stimulus-Organism-Response Paradigm and Privacy Calculus Model

Ning Zhang¹, Jiaqi Pang¹, Jinlin Wan^{1*}

¹School of Information, Central University of Finance and Economics, Beijing, 100081, China

Abstract: This study aims to research the influencing factors of self-disclosure in social network sites (SNSs). Although much has been reported on the issue of influence factors of self-disclosure, very few have been discussed from the aspect of environment stimulus. Drawing upon the Privacy Calculus model and the Stimulus-Organism-Response paradigm, we developed an integrated model to investigate the impacts of SNSs features as environment stimuli on users' perceived value and subsequently their self-disclosure behavior. The results indicate that self-disclosure in SNSs is determined by the tradeoff between perceived benefits (relationship management and enjoyment) and costs (privacy risk). And users' perceptions of SNSs features (interactivity, personalization and privacy control) have significant impacts on perceived value and self-disclosure behavior.

Keywords: social network sites, self-disclosure, stimulus-organism-response paradigm, privacy calculus model

1. INTRODUCTION

Over the past few years, there has been an exponential growth in the use of social networking sites (SNSs). The development of SNSs partly depends on the user-generated content, while self-disclosure is the foundation of content generation, so it's important to figure out what are the most important factors that influence users' self-disclosure behavior.

Generally, researchers paid more attention to users themselves. For example, some researchers explored how users' perception value influence self-disclosure based on privacy calculus theory and found that the perceived benefits could encourage users self-disclosure intentions, while perceived costs may become the obstacles to their self-disclosure, and the self-disclosure is the result of users' cost-benefit tradeoff^[1]. Besides, age, gender, personality are also important factors to influence users' self-disclosure. On the other hand, in order to avoid becoming outdated without ongoing content creation and sharing, SNSs providers need to keep up updating SNSs to stimulate users involving in self-disclosure, which means, SNSs features may be considered as an important influencing factor on self-disclosure. Although a considerable number of researches have been done to discuss the relationship between website features and the user behavior, such as purchase intention and participation intention, very few have been done to explore effects of SNSs features on self-disclosure. Relative work was done by Wang et al.^[2] who researched the effects of SNSs characteristics on self-disclosure, concluded that application compatibility and reputation positively influences the self-disclosure.

This study aims to explore the influence of SNSs features and perceived value on self-disclosure. Based on the stimulus-organism-response (S-O-R) paradigm and privacy calculus model, we developed an integrated model to examine the effects of SNSs features on users' perceived value, following their intention to self-disclosure. This study makes significant contributions to the existent literature. Firstly, the study focusses on how SNSs features influence self-disclosure indirectly, which was rarely studied. Secondly, this study enriches the research framework of self-disclosure with the S-O-R paradigm.

* Corresponding author. Email: zhangning@cufe.edu.cn (Ning Zhang), pangjiaqi15@163.com (Jiaqi Pang)

2. THEORETICAL BACKGROUND

2.1 Privacy calculus model

The privacy calculus model states that people will disclose personal information when perceived benefits exceed potential costs, which means that online self-disclosures are based on a cost-benefit tradeoff^[1]. There are a large of benefits users can get when self-disclosing in SNSs, such as enjoyment, friendship. At the same time, there are also some risks to be considered, such as privacy risks. So, users have to consider the merits and potential negative consequences with respect to the current interaction as well as future situations. Privacy calculus has been widely used to explain self-disclosure behaviors in SNSs. For example, Krasnova et al.^[3] found that enjoyment can promote users' self-disclosure, while privacy concerns had a negative influence on self-disclosure. Xu et al.^[4] combined the privacy calculus model with the theory of planned behavior to explain privacy disclosure behaviors, and found that privacy risks predicted privacy concerns, which in turn determined self-disclosure.

2.2 The S-O-R paradigm

The S-O-R paradigm states that environmental stimuli can affect people's internal state (organism), which directs their behavioral response. Environmental stimuli refer to the environmental factors that users are exposed in the process of interaction with the environment. The internal states refer to users' emotional or cognitive states, including their perceptions, experiences and evaluations. And the response refers to behavior, such as self-disclosing behavior, purchase behavior. The use of S-O-R paradigm as a theoretical foundation for this study is advantageous for three reasons: First, the S-O-R paradigm is widely used to explore how environmental factors influence users' behavior. For example, Islam and Rahman^[5] explored the impacts of online brand community characteristics on customer engagement behavior using S-O-R paradigm. Second, under the online self-disclosure research, previous studies have indicated that environmental factors such as policy and culture differences, as well as features of social network could influence the process of self-disclosure^[2]. Third, it's also convenient to exam the cost-benefit tradeoff that users perceived while self-disclosing in SNSs.

3. RESEARCH MODEL AND HYPOTHESES

3.1 Model building

Based on the S-O-R paradigm and privacy calculus model, an integrated model is built to explore the influence factors of self-disclosure in SNSs. First, this study takes SNSs features as stimuli, which refer to the properties of SNSs. This study focuses on these three crucial SNSs features, that are, interactivity, personalization and privacy control, which reflect various aspects of users' interactions with SNSs. Interactivity and personalization reflect the features that support interactions between users and SNSs, whereas privacy control reflects the feature that facilitates safer and better interactions. Second, this study takes the perceived value as organism. According to the privacy calculus theory, user's perceived value determines the self-disclosure behavior, while the perceived value is based on environment stimuli represents user's internal state. This study uses perceived value as the mediator of SNSs features and self-disclosure. Perceived value consists of perceived benefits and perceived costs, whereas the behavior is determined by the result of the tradeoff. Studies on self-disclosure confirm the important roles of relationship management^[4], self-presentation and enjoyment^[6] for perceived benefits and perceived privacy risks^[7] for perceived costs. Finally, self-disclosure is taken as the response, which refers that one voluntarily and intentionally reveal about themselves to others in this study.

3.2 Research hypotheses

3.2.1 SNSs features and perceived value

Interactivity is the degree to which one can control the medium in modifying its form and content in real time^[8], which is defined as the interactions between users and technology in this study. As a key design feature, interactivity has been found to influence users' response to an online medium^[9]. First, users can present themselves, create a public profile and make connections with the others. Therefore, the SNSs with high perceived interactivity provide a convenient venue for content contribution and relationship management. Second, interactivity feature in SNSs enables a user to portray a desired self-image. According to the self-presentation theory, people desire to project a social self-image among other users and are motivated to use this medium to fulfill the desire. Third, users interact with the SNSs through content creation and sharing, such as boasting about what they read, think and experience in their daily lives. The social networking sites with high interactivity provide users more enjoyable and relaxing environment. Thus, we hypothesize:

H1a: interactivity is positively related to the relationship management.

H1b: interactivity is positively related to the self-presentation.

H1c: interactivity is positively related to the enjoyment.

Personalization is defined here as the adaptability of the site to meet their preferences^[10]. In SNSs, the design of user profiles and recommendation systems closely match users' preferences and needs. Relevant items are classified under a distinct category which labels a list of friends who have similar tastes or someone that you may know, which create an encouraging environment for users to build new relationships and maintain relationships with others^[11]. When users find social recommendations in a medium, they will develop a stronger sense of social identification and familiarity with other users. Thus, increased identification and familiarity with others may generate a sense of social presence and enhance the individual's well-being among them. Besides, due to information overload in social networking sites, personalization will reduce customers' information screening cost and improve their information-interaction quality, and thus contribute to a better experience and be more likely to gain enjoyment. Thus, we hypothesize:

H2a: Personalization is positively related to the relationship management.

H2b: Personalization is positively related to the self-presentation.

H2c: Personalization is positively related to the enjoyment.

Privacy control means the ability to control users' personal information to avoid being leaked without their permission. For example, the ability to control who can see their information and whether they can be found by strangers in SNSs. Some research results have shown that, the stronger the ability of privacy control, the lower the perceived privacy risks^[12]. Thus, we hypothesize:

H3: Privacy control is negatively relating to users' perceived privacy risk.

3.2.2 Perceived value and self-disclosure

Relationship management in SNSs refers to users who are able to build new connections and maintain existing relationships with others^[3]. When a user is willing to disclose more personal information, it sends out desired signals and helps to establish a common group with potential friends. On the contrary, out of relationship management need, users will be more positive to disclose information. Thus, we hypothesize:

H4: Users' beliefs regarding the convenience of managing relationships is positively related to their self-disclosure in social networking sites.

Self-presentation refers that individuals intentionally manage the impression they wish to convey to others^[13]. Through presenting desirable information about themselves in SNSs, users could reveal their identity and formulate the impression they wish to produce^[14]. Driven by the desire to self-present, users make use of available functionality to showcase their achievements and experiences by sharing photos and taking part in

group in groups they deem appealing. Thus, we hypothesize:

H5: Users' belief regarding self-presentation benefits is positively related to their self-disclosure in social networking sites.

Except for the basic communication function, SNSs also offer many entertaining features, such as games, attention-grabbing articles, or videos, which encourage users to participate and disclose their personal information. Prior studies have empirically demonstrated that perceived enjoyment is an influential factor that drives users to self-disclosure^[6]. Following from this, we hypothesize:

H6: Users' enjoyment of the platform use is positively related to their self-disclosure in SNSs.

Perceived privacy risk refers to the degree to which an individual believes that a high potential for loss is associated with the release of personal information in SNSs. SNSs involve particular risks associated with the public accessibility of users' information secret sharing, collection and sharing of information by third parties, identity theft or use of the information for phishing^[15]. And users will control information disclosure according to the degree of the perception of privacy risk. Thus, we hypothesize:

H7: Users' Perceived privacy risk is negatively related to self-disclosure in social networking sites.

According to the theoretical analysis above, an integrated model shown in Figure 1 was proposed to explain the factors influencing the self-disclosure behavior in SNSs.

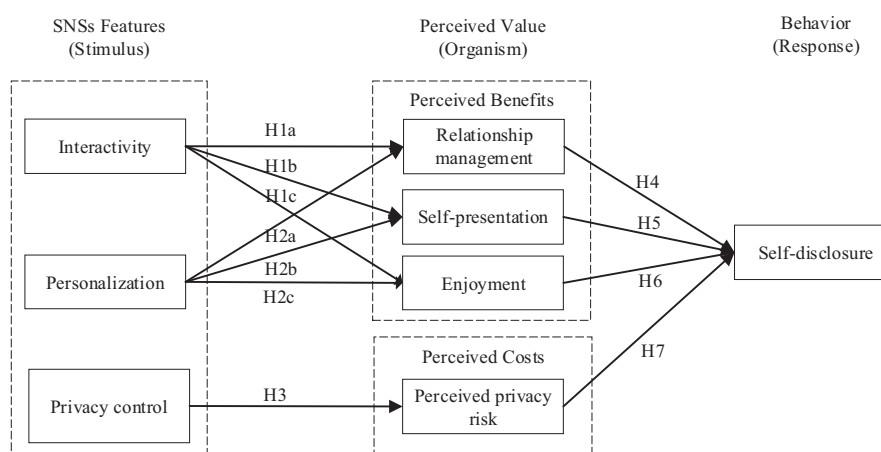


Figure 1. Research model

4. METHODOLOGY

4.1 Instrument

To ensure content validity, all measurement items were adopted from previous literature, with appropriate modifications to suit the context of this study. Each item was taken on a seven-point Likert scale with 1 = “not agree at all”, 4 = “neutral”, and 7 = “absolutely agree”. As the respondents participating in the survey are all Chinese, the questionnaire was translated to Chinese in order to measure precisely.

4.2 Pilot study

Before the formal investigation, we performed a pilot test with 86 valid questionnaires we collected from the classmates to refine the instrument and improve its validity. Exploratory factor analysis (EFA) was conducted to test the convergent and discriminant validity of the instrument. Table 1 reports the principal component analysis results with varimax rotation using SPSS20. According to previous research, the factor loading of the item on the intended construct should be greater than 0.5 to satisfy convergent validity, and on the unintended construct should be less than 0.4 for discriminant validity. As we can see from Table 1, our

instrument has good convergent validity and discriminant validity, and can be used for formal investigation.

4.3 Data collection

An online survey was conducted to collect data from the users of a variety of SNSs in China, such as WeChat, Sina Weibo. As a result, a total of 286 responses were received, however, only 210 usable and valid ones were left. Generally, Males composed 46% of the sample, it was a little lower than female. And that most of the respondents were aged between 18 and 25 years old. Almost 96% of respondents engaged in SNSs at least one year, and 61% had the SNSs experience over than 5 years.

Table 1. Factor loading table for EFA of pilot data

Constructs	Items	Components							
		1	2	3	5	6	7	4	8
Interactivity	INT1	0.857	-0.012	0.05	0.011	-0.044	0.111	-0.139	-0.093
	INT2	0.75	0.123	0.119	0.289	0.118	0.013	0.171	0.12
	INT3	0.826	0.023	0.052	0.018	0.032	0.153	-0.112	0.021
Personality	PER1	0.307	0.62	-0.062	-0.134	-0.09	0.053	0.199	0.232
	PER2	0.056	0.769	0.082	0.126	0.087	0.052	-0.094	0.17
	PER3	-0.199	0.721	0.239	0.273	0.117	0.146	-0.162	0.083
Privacy policy	PP1	0.031	0.049	0.863	0.083	-0.151	-0.059	0.145	-0.025
	PP2	0.084	0.188	0.819	-0.084	-0.003	0.217	-0.198	0.206
	PP3	0.054	0.009	0.89	0.087	0.018	0.052	0.041	0.195
Relationship	RM1	0.061	0.363	-0.015	0.728	0.005	0.076	0.049	0.102
Management	RM2	0.155	0.197	0.139	0.768	0.1	0.077	-0.021	0.245
	RM3	0.05	-0.169	0.003	0.84	0.111	0.086	0.022	0.049
Self-presentation	SP1	0.145	0.117	-0.101	0.143	0.821	-0.049	-0.067	0.074
	SP2	0.191	-0.055	-0.047	0.263	0.75	0.228	0.264	0.17
	SP3	-0.245	0.02	-0.002	-0.089	0.678	0.175	-0.062	-0.015
Enjoyment	EN1	0.018	-0.038	0.077	0.197	0.113	0.777	-0.003	0.051
	EN2	0.05	0.351	0.007	-0.015	0.072	0.718	0.285	0.088
	EN3	0.082	0.043	0.069	0.029	0.09	0.814	0.12	0.248
Perceived privacy risks	PPR1	-0.075	-0.077	-0.065	-0.031	-0.05	-0.016	0.883	-0.024
	PPR2	0.067	-0.015	0.086	0.049	-0.027	0.07	0.889	-0.126
	PPR3	0.021	0.028	0.021	0.034	0.109	0.269	0.838	-0.091
Self-disclosure	SD1	-0.065	0.378	0.164	0.204	-0.099	-0.016	-0.058	0.714
	SD2	-0.079	0.141	0.109	0.099	0.032	0.06	-0.078	0.877
	SD3	0.028	-0.013	0.066	0.145	0.047	0.174	-0.098	0.886
	SD4	0.127	0.124	0.076	0.005	0.199	0.171	-0.038	0.779
Eigenvalue		5.248	3.075	2.327	1.862	1.588	1.444	1.387	1.154
Variance %		26.868	12.813	9.697	7.758	6.616	6.016	5.781	4.809
Cumulative		26.166	34.682	44.379	52.137	58.753	64.769	70.550	75.359
Variance %									

4.4 Data analysis and results

4.4.1 Reliability and validity testing

In order to ensure the quality of the scales, we performed confirmatory factor analysis(CFA) to assess the reliability and validity. CFA is the conventional statistical method used to test reliability and validity for the measurement model. For construct reliability, composite factor reliability (CFR) and Cronbach's alpha (α) are two key indicators. Generally, the constructs have good reliability if CFR and α are greater than 0.7, which are

satisfied according to table 2, suggesting a good reliability in our study.

To ensure sufficient convergent validity, the average variance extracted (AVE) by the latent factor should be greater than 0.5, that is, a construct should explain more than 50% of the item variance. In addition, the standard loadings should be greater than 0.7. Table 2 shows that AVE and standard loading values were higher than the suggested thresholds. Therefore, the constructs have good convergent validity.

Discriminant validity is the degree to which the measurement of a variable is not a reflection of other variables. To ensure sufficient discriminant validity, the square root of AVE for any latent variable should be higher than the correlations between that latent variable and other latent variables. The data shown in Table 3 satisfies this requirement, indicating that each reflective construct is more strongly related to its own measures than to the rest of the constructs, thus demonstrating discriminant validity.

Table 2. The convergent validity of the measurement model

Construct	Item	Standard Loading	T value	AVE	CFR	α	Construct	Item	Standard Loading	T value	AVE	CFR	α				
Interactivity	INT1	.867	13.7	0.76	0.90	0.90	Self-presentation	SP1	.844	11.9	0.66	0.85	0.83				
	INT2	.819	14.1					SP2	.787	12.8							
	INT3	.934	20.8					SP3	.807	13.3							
Personality	PER1	.716	7.80	0.53	0.77	0.71	Enjoyment	EN1	.805	10.7	0.60	0.82	0.81				
	PER2	.784	11.9					EN2	.723	12.9							
	PER3	.697	11.5					EN3	.802	13.4							
Privacy control	PC1	.859	13.9	0.74	0.89	0.88	Perceived privacy risk	PPR1	.891	14.0	0.78	0.91	0.89				
	PC2	.863	16.6					PPR2	.895	15.4							
	PC3	.870	12.7					PPR3	.867	14.9							
Relationship management	RM1	.761	9.16	0.52	0.76	0.70	Self-disclosure	SD1	.723	13.0	0.62	0.87	0.86				
	RM2	.769	11.5					SD2	.785	13.9							
	RM3	.639	8.17					SD3	.867	13.3							
								SD4	.784	12.1							

Table 3. Descriptive statistics and factor correlation

	Mean	SD	INT	PER	PP	RM	SP	EN	PPR	SD
INT	5.75	1.16	0.87							
PER	4.71	1.54	0.26	0.73						
PC	4.61	1.61	0.15	0.53	0.86					
RM	5.25	1.22	0.33	0.46	0.24	0.72				
SP	5.39	1.18	0.35	0.37	0.2	0.23	0.81			
EN	5.38	1.30	0.33	0.53	0.28	0.29	0.25	0.78		
PPR	5.64	1.28	-0.03	-0.09	-0.17	-0.04	-0.03	-0.05	0.88	
SD	4.51	1.62	0.27	0.4	0.24	0.44	0.26	0.51	-0.25	0.79

4.4.2 Hypotheses testing

Because our measurement model was satisfactory, we proceeded to hypothesis testing. Using LISREL 8, we created a structural equation model to specify both item-construct correspondence and construct-construct causal relationship. The results summarized in Figure 2.

Before testing the hypotheses, we must ensure that the model must fit the data well. There were a few model fitting indices to be employed here, including Chi-square/degree of freedom ($\chi^2/df \leq 3$), goodness of

fit index ($GFI \geq 0.90$), normed fit index ($NFI \geq 0.90$), comparative fit index ($CFI \geq 0.90$), and root mean square error of approximation ($RMSEA \leq 0.08$), non-normed fit index ($NNFI \geq 0.90$). According to Figure 2, our results indicated satisfactory model fit.

The results shown in Figure 2 indicate that all our hypotheses were supported except the effects of self-presentation on self-disclosure (H5). Both of perceived interactivity and perceived personalization have positive impacts on perceived benefits (relationship management, self-presentation and enjoyment), while privacy control has a negative impact on perceived privacy risk, thus hypotheses H1a-H1c, H2a-H2c and H3 are supported. Meanwhile, perceived benefits (relationship management and enjoyment) have positive and significant impacts on self-disclosure, and perceived privacy risk also has a negative impact on self-disclosure, thus, H4, H6 and H7 are supported.

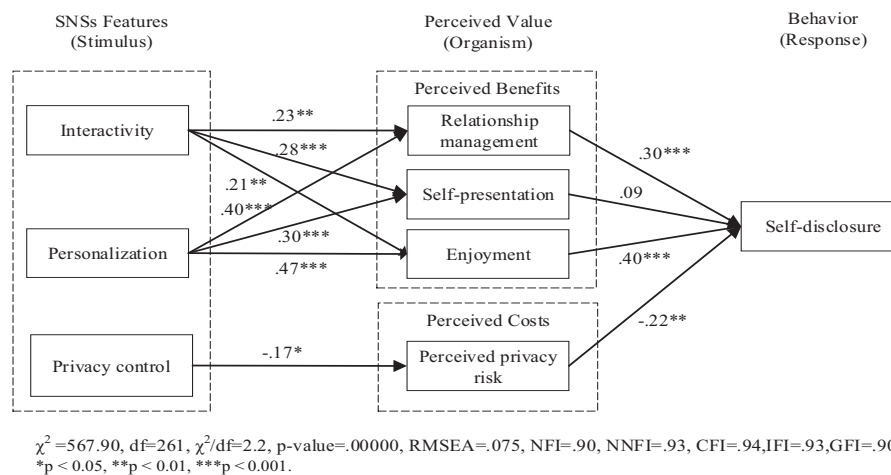


Figure 2. Empirical results

5. DISCUSSION

5.1 Discussion of findings

In summary, ten out of eleven hypotheses are supported. The results show that it is feasible to use the S-O-R paradigm to explain the self-disclosure behavior in SNSs, and the features of SNSs significantly influence their perceived value, and in turn users' cost-benefit tradeoff influence their self-disclosure behavior. On the one hand, the technical features, including interactivity, personalization and privacy control, are proved to have an indirect influence on self-disclosure. In other words, the technical features could stimulate users' internal states and affect users' behaviors. On the other hand, this study also replicates previous research of using privacy calculus to explain self-disclosure. However, we did not find support for a positive relationship between self-presentation and self-disclosure. One plausible explanation is that users will only want to present in breadth and depth their positive side and refrain from presenting their negative side in order to project a good image, which means that users may in fact withhold information or only selectively release certain information^[16]. If this is the case, self-presentation may be unrelated to self-disclosure.

5.2 Implications

This study has both theoretical and practical significance. First, this study extends the applicability of the S-O-R paradigm to the context of SNSs, which offers a new perspective to the research on the influencing factors of self-disclosure in SNSs. Second, the SNSs features are proved to be important influencing factors of self-disclosure in SNSs, which extends the scope of influencing factors of self-disclosure. Third, in the light of

the importance of SNSs features, service providers are suggested to increase interactivity and personalization by allowing real-time and less-restrained interactions, and also improve users' virtual experiences to provide more enjoyable socializing interactions. In addition, the privacy control function should be well established and enforced.

5.3 Limitations

There are some limitations in this study, which provide opportunities for further research. First, there are still some other stimulus factors that may affect users' self-disclosure behavior which are not contained in the integrated model, such as awards, social support, culture. Second, as the sample we collected in this study is from China's SNSs, the results may not be applicable universally. So, further verification in different contexts of SNSs, such as Facebook, Twitter, is needed in the future.

6. CONCLUSIONS

This study applied the S-O-R paradigm in the context of SNSs and developed an integrated model combined with the privacy calculus model to explore whether SNSs features influence users' self-disclosure. The results show that SNSs features play a crucial role in stimulating self-disclosure behavior and suggest that SNSs service providers should increase interactivity, provide personalization experience and enforce privacy protection to enhance the users' self-disclosure, and in turn, to improve the development of SNSs.

ACKNOWLEDGEMENT

This research was supported by the National Social Science Foundation of China under Grant 13AXW010 and National Key R&D Program of China under Grant 2017YFB1400701.

REFERENCES

- [1] Dienlin T, Metzger M J. (2016). An Extended Privacy Calculus Model for SNSs: Analyzing Self-Disclosure and Self-Withdrawal in a Representative US Sample. *Journal of Computer-Mediated Communication*, 21(5): 368-383.
- [2] Wang Le, Yan Jie, Lin Jun, et al. (2017). Let the users tell the truth: Self-disclosure intention and self-disclosure honesty in mobile social networking. *International Journal of Information Management*, 37(1): 1428-1440.
- [3] Krasnova H, Spiekermann S, Koroleva K, et al. (2010). Online social networks: why we disclose. *Journal of Information Technology*, 25(2): 109-125.
- [4] Xu F, Michael K, Chen X. (2013). Factors affecting privacy disclosure on social network sites: an integrated model. *Electronic Commerce Research*, 13(2): 151-168.
- [5] Jamid U I, Zillur R. (2017). The impact of online brand community characteristics on customer engagement: An application of Stimulus-Organism Response paradigm. *Telematics and Informatics*, 34(4): 96-109.
- [6] Cheung C, Lee Z W Y, Chan T K H. (2015). Self-disclosure in social networking sites: The role of perceived cost, perceived benefits and social influence. *Internet Research*, 25(2): 279-299.
- [7] Wang L, Yan J, Lin J, et al. (2017). Let the users tell the truth: Self-disclosure intention and self-disclosure honesty in mobile social networking. *International Journal of Information Management*, 37(1): 1428-1440.
- [8] Jonathan S. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, 42(4): 73-93.
- [9] Jiang L C, Bazarova N N, Hancock J T. (2013). From Perception to Behavior: Disclosure Reciprocity and the Intensification of Intimacy in Computer-Mediated Communication. *Communication Research*, 40(1): 125-143.
- [10] Tam K Y, Ho S Y. (2005). Web Personalization as a Persuasion Strategy: An Elaboration Likelihood Model Perspective. *Information Systems Research*, 16(3): 271-291.

- [11] Tam K Y, Ho S Y. (2006). Understanding the Impact of Web Personalization on User Information Processing and Decision Outcomes. *Mis Quarterly*, 30(4): 865-890.
- [12] Benson V, Saridakis G, Tennakoon H.(2015). Information disclosure of social media users Does control over personal information, user awareness and security notices matter?. *Information Technology & People*, 28(3): 426-441.
- [13] Boyd D M, Ellison N B. (2007). Social Network Sites: Definition, History, and Scholarship. *IEEE Engineering Management Review*, 13(1): 16-31.
- [14] Ellison N, Steinfield C, Lampe C. (2007). The Benefits of Facebook “Friends:” Social Capital and College Students’ Use of Online Social Network Sites. *Journal of Computer-mediated Communication*, 12(4):1143-1168.
- [15] Hogben G. (2007). Security issues and recommendations for online social networks. *Track Security & Privacy*, 15(2): 81-90.
- [16] Liu Z L, Min Q F, Zhai Q G, et al. (2016). Self-disclosure in Chinese micro-blogging: A social exchange theory perspective. *Information & Management*, 53(1): 53-63.

Assessing Customer-Based Brand Equity of E-Commerce

Platform Based on Promotion Benefits

Ting Zhu^a, Jingwen Liu^{b}, Yanhong Chen^a*

a. College of Management, Huazhong University of Science and Technology, Wuhan, 430074, China

b. Huazhong University of Science and Technology Library, Wuhan, 430074, China

Abstract: The prosperity of e-commerce in China leads to large amount of businesses venturing into this marketplace. Sales promotion has been widely used to improve their advantage, while platform-based sales promotion has been confirmed more successful. This research examines the impacts of the benefits of sales promotions based on electronic commerce platforms on customer-based brand equity (platform brand awareness and platform brand association) and how their relationships are moderated by the promotion type (pre- and post-purchase promotion). Based on the two functions of sales promotions (stimulation vs. maintaining), we propose a five-benefit framework including exploration, convenience, savings, social bonds and structural bonds. We suggested that both the two types of sales promotion could provide the five benefits of sales promotions and the positive effects of the benefits on consumer-based brand equity might be significant, while the differences between pre- and post-purchase sales promotions could be also significant.

Keywords: Platform-based sales promotion, multiple-benefit framework, consumer-based brand equity

1. INTRODUCTION

Since its inception, e-commerce has experienced fast growth in China. Because of the prosperity of e-commerce, both large and small businesses have ventured into this marketplace, leading to fierce competition. Multiple e-commerce platforms exist and provide businesses and their customers with functions such as product catalog, sales promotion, communication, financial transaction, and customer service. How to increase e-commerce platforms' competitiveness has become a critical issue. Since its launch in 2009 by Tmall.com, the business-to-consumer (B2C) branch of Taobao.com in China, a large-scale sales promotion dubbed "Double 11" has taken place annually on November 11. On "Double 11," consumers can purchase heavily discounted merchandise, similar to the Black Friday promotion in the US. More and more sellers on Tmall or even other online shopping platforms in China have participated in this activity. Table 1 summarizes the exponential growth of the one-day sales amount from 2009 to 2015. The overwhelming success of Tmall and its platform-wide sales promotion has drawn the attention of both business practitioners and academic researchers. In the marketing literature, brand equity can improve businesses' competitiveness by reducing consumers' uncertainty and decision cost^[1]. Because customers perceive higher risks in e-commerce, brand equity is even more important. Although the success of e-commerce in China owes to the initial rise of consumer-to-consumer (C2C) e-commerce, B2C e-commerce has become mainstream in online shopping, suggesting that more and more consumers take brand into consideration when they shop online.

As brand equity's importance increases, it has attracted researchers' attention^[2]. Multiple perspectives about the concept of brand equity exist^[1]. Among them, customer-based brand equity (CBBE) is an important view that focuses on customers' reactions to marketing tactics implemented by the brand compared to their reactions to the same marketing tactics implemented by others^[3]. Many researchers have examined the relationship between sales promotion and brand equity. Some argue that sales promotion increases customers'

*Corresponding author. E-mail addresses: tzh@hust.edu.cn (T. Zhu), liujw-518@163.com (J. Liu), yachne6688@sina.cn (Y.H. Chen)

price sensitivity and lowers quality perception, thus decreasing brand equity in the long-term^[4], while others find no negative effects^[5]. However, no research has investigated this relationship in the e-commerce context in China. This study aims to examine the differential effects of sales promotion on CBBE in the Chinese e-commerce industry. In doing this, we introduce two types of sales promotion – monetary promotion versus non-monetary promotion^[6] – and the timing of the promotion – pre-purchase versus post-purchase sales promotion^[7]. Pre-purchase sales promotions such as advertisement, sale prices and coupons are given to consumers prior to their purchases and are used to stimulate consumption. In contrast, post-purchase promotions are given to customers after their purchases and are used to maintain the relationship between the consumers and the shopping platform. Examples include Taobao coins (virtual coins that can be used on Taobao and Tmall toward future purchases with sellers who accept them), sweepstakes and shipping cost guarantee (to reimburse the customers their shipping charges when they return their purchases). We investigate the effects of sales promotions on the e-commerce platform's CBBE to better understand consumer behavior and identify effective sales promotion tactics that can positively affect CBBE. Hence, we have the following two research questions:

RQ 1. How do sales promotions of the e-commerce platform affect its CBBE?

RQ 2. Are the effects of pre-purchase and post-purchase sales promotions on CBBE different?

To understand the effects of e-commerce platform sales promotions on CBBE, we identify the benefits consumers receive from sales promotions. Instead of using the benefits framework provided by Chandon et al.^[6], we develop a new five-benefits framework including exploration, convenience, savings, social bonds and structural bonds to explain how sales promotions affect CBBE. This paper extends the study on sales promotion and brand equity to the e-commerce platform context and reveals the positive effects of sales promotions on CBBE. Moreover, we are the first to examine the timing of pre-purchase and post-purchase promotions.

2. THEORETICAL BACKGROUND

2.1 Sales promotion.

Promotion has been widely studied by academic researchers^[5]. Among such research on promotion, sales promotion, which comprises of various motivational tactics to stimulate consumers' consumption^[8], receives the most attention^{[4][6]}. Sales promotion includes multiple tactics such as sales (e.g., markdowns and clearance), promotion (e.g., gifts, cash rebates, or two for one pricing), and purchase ideas (such as one-stop shopping)). However, relationship marketing tools such as store specific currency rebates (e.g., Taobao coins), free returns and payment financing service (a customer can buy a product on the e-commerce platform by paying in installments with no money down for the first few months) have also been used in sales promotion to enable relationship maintenance between buyers and sellers.

The extant literature on sales promotion has examined a variety of topics. Some studies focus on performance^[4], some focus on the types of promotion^[9], and others focus on the benefits^[6]. Though price discounts are thought to be the main benefit sales promotion provides^[10], other hedonic or utilitarian benefits are also important value generated by sales promotion^[6]. Thus, sales promotion can be divided into monetary promotion such as coupons, sale prices, rebates and non-monetary promotion such as contests and sweepstakes^[9]. Different types of promotion have different effects on consumers' purchase decision^[11] and brand equity^[9]. Monetary promotions negatively affect brand image^[12], while non-monetary promotions have positive effects on brand equity^[13]. Though researchers have studied the effect of sales promotion on brand equity, none has examined the relationship from the benefits perspective.

According to Keller^[3], benefits of sales promotion are the perceived value attached to consumers' experience, which includes exposure (e.g. seeing promotional information) and usage (e.g. buying a product at a

discounted price). Chandon et al.^[6] identify savings, convenience, and exploration as the most important benefits people obtain from pre-purchase sales promotions. relationship bonds are the most important benefits obtain from post-purchase sales promotion, which are usually classified into two categories: structural and social bonding^[14].

Benefit	Definition	Sources
Exploration	Customers' ability to find new shopping ideas when they are exposed to promotional information.	(Kahn, 1995; Baumgartner and Steenkamp, 1996)
Convenience	Customers can be reminded of their needs and their shopping experience can be more efficient through sales promotions that reduce their search and decision costs.	(Wansink et al., 1998; Chandon et al., 2000)
Savings/Financial bonds	Customers' ability to pay a lower price for a product of the same quality.	(Blattberg and Neslin, 1990; Berry, 1995)
Structural bonds	Special services or privilege customers can obtain from sales promotions based on the relationship; Cannot be taken away if customers end the relationship.	(Berry, 1995; Gwinner et al., 1998; Lin et al., 2001)
Social bonds	Personal affection customers develop about the seller through sales promotions.	(Berry, 1995; Gwinner et al., 1998)

According to Cao and Gruca^[7], services provided by e-tailers can be classified into pre-purchase and post-purchase services based on two basic functions of the service provider: information and delivery. Pre-purchase services are used to stimulate customers into making purchases, while post-purchase services are provided to enhance value to retain customers. Sales promotions also provide two functions. First, they provide information on the promoted products and reduce prices to appeal to prospective customers. Second, coupons or promotional services can be provided after a purchase to stimulate future purchases. Thus, we define pre-purchase sales promotion as promotional tactics that can be viewed prior to a purchase and post-purchase sales promotion as promotional tactics provided based on a purchase. For example, information about product delivery made available before a purchase is a pre-purchase sales promotion, while coupons or discounts for the next purchase provided after a purchase are post-purchase sales promotions. Post-purchase sales promotion focuses more on maintaining the relationship with the customers than pre-purchase sales promotion.

2.2 Consumer-based brand equity.

Brand equity is defined as the added value to a product or service by being linked with a brand name^[3]. It helps a customer make a choice among multiple alternative products and increases her confidence in the purchase decision and satisfaction with the product^[1]. Thus, it is important for businesses to build a strong brand. Aaker^[1] identifies five components of brand equity: brand awareness, brand loyalty, perceived quality, brand associations, and other proprietary assets. Farquhar^[15] suggests CBBE is the added equity a consumer perceives of the brand.

CBBE has been investigated by academics and practitioners since it was coined in the early 1980s^[16]. The concept of CBBE and its dimensions defined in this study are based on Keller^[3]. CBBE refers to a customer's reaction to the marketing tactics implemented by a brand compared to her reaction to the same marketing tactics implemented by others brands. It is based on the customer's knowledge about the brand, which comprises of a

variety of associations in memory. CBBE consists of two components in this study: brand awareness and brand association. Brand awareness refers to the ability of a customer to recognize or recall the name of the brand when given a signal. It is positively affected by the presentation of the brand^[17]. Brand association is also called brand image^[18] or brand meaning^[17] and refers to a customer's subjective perception of the brand, which is positively affected by the customer's experience. In the context of this study, brand awareness means the extent to which an individual recalls the name of an e-commerce platform when she wants to purchase online. Further, brand association refers to the dominant trait of the e-commerce platform that comes to a customer's mind when she is reminded of the e-commerce platform. For example, low prices and variety of products are the main appeals of Taobao.com, while high quality 3C products (computer, communication and consumer electronics) are the main appeal of JD.com.

3. WHAT IS A SYSTEMS APPROACH?

Based on the five benefits of sales promotions, we next developed our theoretical model on how these benefits affect the CBBE and how their relationships are moderated by the promotion type. Figure 1 summarizes our research model.

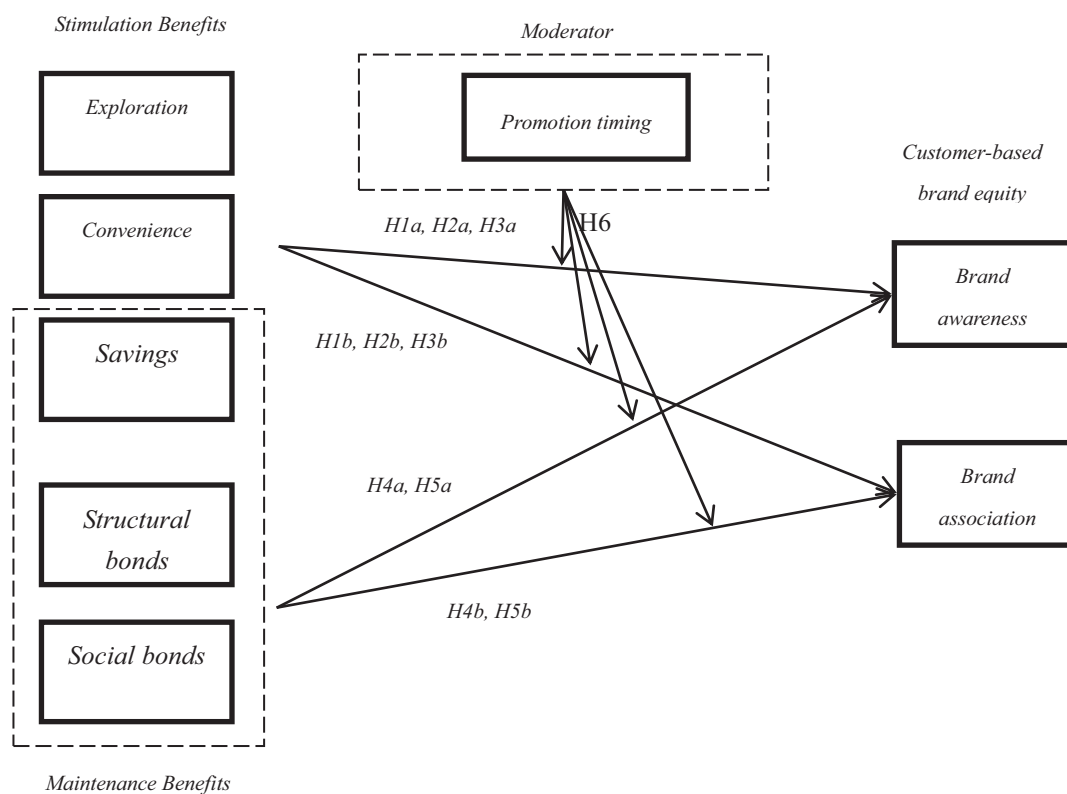


Figure 1. The web of system performance

Exploration refers to the ability of consumers to identify new shopping ideas when exposed to promotional information^[6]. Compared to consumers who receive no promotional information of an e-commerce platform, those frequently exposed to promotional information of the e-commerce platform may recognize and recall the name of the e-commerce platform readily. Moreover, more product information is usually provided to consumers together with the promotional information. According to Ballantine^[19], the amount of information contributes to customers' satisfaction in online shopping, thus leading to improved brand awareness and brand association.

H1(a,b): Exploration is positively associated with the e-commerce platform's brand awareness/ brand

association.

Convenience as a result of sales promotion refers to the reduction of search and decision costs through the availability of products and the advertisement of sales^[6]. When a platform-wide sales promotion occurs, many sellers participate in the promotion and advertisement for the promotion with a link to the participating sellers and products is placed on the front page of the e-commerce platform. This makes it easy for the customers to locate the sellers and products that participate in the sales promotion. It also gives customers a variety of products to choose from. Hence, customers' search and decision costs are reduced significantly. The lower search cost a customer perceives from an e-commerce platform-based sales promotion may lead to high brand associations with strength and ease to use, because consumers develop brand association between a brand and the perceived benefits of the brand. Many studies have examined the effect of recommender system on brand loyalty^[20] and showed that recommendations remind consumers of what they want and enhance their perception of the brand. In other words, personalized recommendations will positively affect CBBE, thus convenience benefits contribute to CBBE. Overall, the ease of the shopping process owing to convenience benefits will lead to a positive shopping experience, which contributes to CBBE^[17]. Thus, we propose:

H2(a, b): Convenience is positively associated with the e-commerce platform's brand awareness/ brand association.

Chandon, et al.^[6] define savings as the monetary savings sales promotions provide. Both pre-purchase and post-purchase sales promotions provide consumers with savings. Low prices can help the e-commerce platform develop a competitive brand association and enhance the platform's appeal^[17].

H3(a, b): Savings are positively associated with brand awareness.

Structural bonds refer to situation that the benefits customers obtain from the relationship cannot be taken away when the relationship ends^[14]. Hence, structural bonding between a brand and a customer is limited to the brand. In other words, a mutual commitment is created between the customer and the brand by structural bonds^[14]. This commitment significantly relates to brand identification, which contributes to enhanced CBBE. Thus, we have:

H4(a, b): Structural bonds are positively associated with brand awareness.

Social bonds refer to the interpersonal relationships customers have with a brand^[14]. Social bonds may develop through personalized services delivered to consumers and leading to the consumers' preference of the brand. They represent the degree to which customers develop an affect for the brand. Aaker^[1] argues that emotional benefit is very important in building strong brands.

H5(a, b): Social bonds are positively associated with brand awareness.

Due to the two different functions of the five benefits including stimulation and maintenance^[7], their effects on brand awareness and brand association may be different. The stimulation function of sales promotion is to arouse consumers' consumption needs and purchase intention on the e-commerce platform rather than on other platforms^[8]. Hence, the benefits associated with this function including exploration, convenience, and savings are provided mainly to distinguish the e-commerce platform from other platforms and make consumers more aware of the brand^[6]. In contrast, the maintenance function of sales promotion is to maintain a strong relationship with the customer so as to facilitate repeated purchases. Thus those benefits associated with this function including structural and social bonds mainly aim to increase the importance of the brand to the customers. It follows that the effects of sales promotion benefits on brand association are related to the extent to which the maintenance benefits are provided. Similarly, the effects of sales promotion benefits on brand awareness are related to the extent to which the stimulation benefits are provided. Thus, we propose that the effects of the benefits on CBBE between pre- and post-purchase promotions are different:

H6: The relationships between the benefits of sales promotion and CBBE are moderated by the promotion

type.

4. METHODOLOGY.

We used a questionnaire survey to examine the benefits of e-commerce platform-based sales promotions and test our research model. While sales promotions can stimulate consumers' consumption, they can also be used for relationship marketing. Many studies have investigated the stimulation function of sales promotions^[8], but the buyer and seller relationship maintenance function has received little attention in academic research. Hence, our empirical analyses were performed in two steps. First, we conducted an exploratory study to confirm the stimulation and maintenance functions of the five benefits and how these five benefits affect the evaluations of pre- and post-purchase sales promotions. Next, we analyzed the data to test our research model.

4.1 Instrument.

While developing the questionnaire for the research, we need interview some consumers on e-commerce platforms to explicitly distinguish between pre- and post-purchase sales promotion tactics. Based on their feedback, we adapt the measurement scales from the extant literature. The items on savings, convenience, and exploration are based on Chandon et al.^[6], the items for structural bonding and social bonding are adapted from Berry^[17], and three items measuring the evaluation of sales promotions are also adapted from Chandon et al.^[6]. CBBE consists of brand awareness and brand association^[3]. The three items on brand awareness and three items on brand association are adapted from Wang and Finn^[2]. All items are measured on seven-point Likert scales. Back translation will be used to ensure the consistency of the scales between the English and Chinese versions.

ACKNOWLEDGEMENT

This research was supported by a grant from the National Natural Science Foundation of China (71332001, 71201062, 71602063).

REFERENCES

- [1] Aaker, D. (1991). "Brand equity." *La gestione del valore della marca*.
- [2] Wang, L. and A. Finn (2014). "A consumer-based brand equity study for small market share brands." *Market & Social Research* 22(2): 6-14.
- [3] Keller, K. L. (1993). "Conceptualizing, measuring, and managing customer-based brand equity." *Journal of Marketing* 57: 1-22.
- [4] Blattberg, R. C. and S. A. Neslin (1990). "Sales promotion: Concepts, methods, and strategies." Englewood Cliffs, NJ: Prentice Hall,
- [5] Davis, S., J. J. Inman and L. McAslister (1992). "Promotion has a negative effect on brand evaluations: Or does it? Additional disconfirming evidence." *Journal of marketing research* 29(1): 143-148.
- [6] Chandon, P., B. Wansink and G. Laurent (2000). "A benefit congruency framework of sales promotion effectiveness." *Journal of Marketing*.
- [7] Cao, Y. and T. S. Gruca (2004). "The influence of pre - and post - purchase service on prices in the online book market." *Journal of Interactive Marketing* 18(4): 51-62.
- [8] Kotler, P. (2012). "Marketing Management." Harlow, England; New York: Pearson.
- [9] Buil, I., L. d. Chernatony and E. Martinez (2013). "Examining the role of advertising and sales promotions in brand equity creation." *Journal of Business Research* 66: 115-122.

- [10] Inman, J. J., L. McAlister and W. D. Hoyer (1990). "Promotion Signal: Proxy for a Price Cut?" *Journal of Consumer Research* 17(June): 74-81.
- [11] Santini, F. O., C. H. Sampaio, M. G. Perin, L. B. Espartel and W. J. Ladeira (2015). "Moderating effects of sales promotion types." *BAR-Brazilian Administration Review* 12(2): 169-189.
- [12] Montaner, T. and J.-M. Pina (2008). "The effect of promotion type and benefit congruency on brand image." *Journal of Applied Business Research* 24(3).
- [13] Palazón-Vidal, M. and E. Delgado-Ballester (2005). "Sales promotions effects on consumer-based brand equity[J]." *International Journal of Market Research* 47(2): 179-204.
- [14] Wilson, D. T. (1995). "An integrated model of buyer-seller relationships. ." *Journal of the Academy of Marketing Science* 23: 335-345.
- [15] Farquhar, P. (1989). "Managing brand equity " *Marketing Research*(September): 24-33.
- [16] Barwise, P. (1993). "Brand equity: snark or boojum?" *International Journal of Research in Marketing* 10(1): 93-104.
- [17] Berry, L. (2000). "Cultivating service brand equity." *Journal of the Academy of Marketing Science* 28(1): 128-137.
- [18] Sharp, B. (1995). "Brand equity and market-based assets of professional service firms." *Journal of Professional Services Marketing* 13(1): 3-13.
- [19] Ballantine, P. W. (2005). "Effects of interactivity and product information on consumer satisfaction in an online retail setting." *International Journal of Retail & Distribution Management* 33(6): 461-471.
- [20] Schafer, J. B., J. A. Konstan and J. Riedl (2001). "E-commerce recommendation applications." *Applications of Data Mining to Electronic Commerce*: 115-153.

Network Structure and Employee Creativity in a Heterogeneous Team

Jianping Peng^{1,2}, Taiye Luo², Jing Quan^{3*}, Peiwen Guo²

¹Sun Yat-Sen Business School, Sun Yat-Sen University, Guangzhou, China 510275

²Sun Yat-Sen Xinhua College, Guangzhou, China 510275

³Franklin P. Perdue School of Business, Salisbury University, Salisbury, MD 21801, USA

Abstract: This paper investigates the relationships between individual creativity, social network position, and knowledge sharing in a heterogeneous team with personnel from four different companies to complete a software outsourcing project. Our research model is based on a theoretical construction attained from the analysis of previous frameworks proposed in the literature. Based on the survey data of the team members, we construct the advice social network, extract employee network centrality, and obtain employee behavioural variables. We find that a positive and strong relationship between knowledge sharing and creativity. The relationship between network centrality and creativity is also positive, but the relationship is moderated by knowledge sharing. Based on the findings, we propose measures and strategies for the hosting company to effectively coordinate the heterogeneous team.

Keywords: creativity, knowledge sharing, employee social networks, heterogeneous team

1. INTRODUCTION

As the information technology (IT) outsourcing market grows, a new outsourcing approach has emerged for software developers to implement large- and medium-sized software outsourcing projects. Different from the existing model in which a project is divided into parts that are then outsourced, this new approach allows a firm to build a temporary need-based human resource development plan to attract software developers from companies that have a surplus of developers, and embed those developers into the firm's own software project teams. Once the project is finished, the developers return to the companies where they are originally employed. In this model, software developers from the entire IT industry can be considered a pool of potential human resources for software developers. Companies are able to acquire suitable software developers from this pool based on the specific needs of a project. Compared to the traditional method of software outsourcing, this new heterogeneous team model approach can better control costs, progress, and quality of the project.

Because team members on the heterogeneous team are not familiar with each other, questions arise about effective team management. One of the key research questions is how the creativity of individual employees is affected in this context. As new members join an existing team, how would the relational network structure evolve? Would the creativity be enhanced or hindered as a result? What is the role of knowledge sharing among the new and old team members?

The answers to these questions are inconsistent in the research community, reflecting, in a broader sense, the debate between the “pessimistic” and “optimistic” views of how the demographic diversity of a team affects its performance [11]. In the “pessimistic” view, demographic diversity is considered problematic, because it introduces social divisions that hinder effective teamwork, due to tensions between members of different organizational cohorts [10]. Hence, it is easier to coordinate the activities of homogenous groups than diverse teams and, as a result, the performance of such teams is expected to be better ([7], [9], [14]).

In contrast, the “optimistic” view argues that diverse teams improve team productivity and performance, because members from different parts of the organization know different sets of people and often have different technical skills [1]. Researchers (e.g., [2]) who hold this view contend that the relative redundancy of

* Corresponding author. Email: jxquan@salisbury.edu

homogenous team members' perspectives, information, and resources limits the team performance. Teams with greater demographic diversity, however, entail relationships among people with different sets of contacts, skills, information, and experiences, thus enhance creativity for problem solving ([3], [4], [5], [6], [8], [12], [13]).

Our research extends the existing understanding of team diversity and performance from two different perspectives. First, instead of team heterogeneity within an organization, our focal point is on diversity of team members from different organizations. In our study, a research and development (R&D) team in the host software company was formed by temporarily hiring additional workers from three other software companies. Second, instead of focusing on the team productivity, our focal point is the individual team members' creativity. Although research on how team diversity relates to team creativity is extensive, little research has been done to explore the relationship between team diversity and individual creativity. Similarly, research calls for the investigation of the individual-level outcomes of team diversity, because team diversity may not have a similar effect on individual creativity as it does on team-level creativity. Other research suggests that team creativity is a function of individual creativity, as well as other team dynamics including structure, climate and leaders' abilities. Further study proposes an interactionist model in which individual creativity is a result of complex person-situation interaction influenced by events of both past and current situations. Creativity is a function of antecedent conditions, cognitive style and ability, personality factors, relevant knowledge, motivation, social influences, and contextual influences.

To investigate the relationship between team diversity and individual creativity under the aforementioned new perspectives, we construct an empirical model with two main variables, i.e., network centrality and knowledge sharing. The reasons for these variables are as follows. First, the basis for the relationship between social networks of the heterogeneous team and the creativity of individual team members is the interplay of network centrality, risk cognition and creativity. Research has established that nodes in the center of a given network are more capable of acquiring information and knowledge through communication than those in the peripheral positions. The opportunity and ability to acquire knowledge would further give individuals the confidence and judgment to take risks. Studies assert that risk taking is integral to creativity. Risk is associated with the uncertainty about outcomes of decisions, given creative efforts. Creativity is risky because the action-outcome link is highly uncertain and lengthy in time. For example, creativity required for inventing innovative products and/or proposing a novel and effective solution to existing problems often involves a high degree of risk taking. When employees are willing to take risks, it is conducive to creativity.

Second, for the role knowledge sharing, current research suggests two possible linkages: social network to knowledge sharing and knowledge sharing to creativity. The first linkage can be explained from the perspective of the Social Capital Theory. Social capital refers to the relationships in a social network and the associated set of embedded resources. The Social Capital Theory believes that social capital strongly influences the extent to which knowledge sharing between the network nodes occurs. Empirical research has found significant relationships between employee social networks and knowledge sharing. Study finds that social interaction ties are positively associated with individuals' knowledge sharing in virtual communities. For the second linkage, research has shown that collaboration among knowledge workers is an important factor for successful R&D projects. Knowledge sharing and knowledge reconstruction within a team is important for promoting creativity, because team members can more effectively process information and generate creative ideas by sharing information. Researchers find that knowledge sharing facilitates the generation of new ideas, and that higher levels of knowledge sharing are significantly correlated with higher levels of creativity in work projects.

Specifically, this study attempts to answer the following three research questions:

RQ1: How is network centrality related to the individual creativity of a heterogeneous team?

RQ2: How is knowledge sharing related to the individual creativity of a heterogeneous team?

RQ3: Does knowledge sharing moderate the relationship between network centrality and creativity?

2. THEORETICAL BACKGROUND AND RESEARCH MODEL

2.1 Team Diversity and Creativity

Creativity can be defined as the generation of novel and useful ideas concerning products, services, processes, and procedures by an employee. Researchers agree that a prerequisite for creativity is the ability to think differently, to see things from unique perspectives, and to sense the possibility of turning previously unrelated processes, products, or materials into something new and better. Although individual cognition and motivation are the main sources for creative ideas, the interactions within teams foster their generation [8]. Researchers find that team creativity is iterative and interactive in nature and requires an individual's willingness to engage in individual level creativity.

Diversity can be defined as the distribution of differences among the members of a unit with respect to a common attribute such as tenure and ethnicity. The diversity literature suggests the value of team diversity and performance can be seen from the “structural holes” conception of social capital that emerges from social network theory [11]. Structural holes are the gaps between nodes in a social network [2]. The value derived from bridging the gaps lies in the boundary spanning that generates “information benefits”, because information tends to be relatively “redundant” within a given group ([2] p. 13-16). Hence, members who span team boundaries to create ties with other teams can gain access to a wider range of ideas and resources than those who are restricted to a single team [4]. This reasoning is also applicable to the value of team diversity. According to research, heterogeneous teams with members from diverse pools of talents can develop ties between people with different skills, knowledge, perspectives, and experiences. Such links bridge structural holes within the team [11], and can be very valuable sources for team creativity. The heterogeneous team structure can also be conducive for individual creativity because it allows team members to draw from individuals with differences in backgrounds, knowledge, skills, thinking styles, perspectives and experiences. Such benefits can be even greater with a heterogeneous team that draws members from different organizations, because those differences are more pronounced.

2.2 Network Position and Creativity

Centrality reflects the number of ties between a member and all other members within a given network. A node with a higher degree of centrality implies the actor has many ties. The high number of ties allows them to have alternative ways to satisfy needs, and to gain access to more of the resources of the network as a whole. Within a social network, individuals in the center are more capable of acquiring information and knowledge through communication, are more likely to be perceived as having higher status, and are more likely to know what is going on within the network. A central individual, with this access and status, should have more favorable perspectives and outlooks, and enjoy freedom and power. Researchers argue that the advantages should provide the confidence and personal judgement for risk cognition and risk taking. Research has identified positive links between risk taking and creativity, because creativity often implies either doing something new or using an unusual approach for an existing problem. The fact that central individuals are more likely to be exposed to the various disparate social circles within the network allows them to see different perspectives and to think creatively. Because of the connections, they may be more open-minded when considering different approaches or ways of thinking. This line of reasoning leads researchers to propose that centrality is positively related to individual creativity.

In our context, the triad of network position, risk cognition and creativity is even more pronounced.

Because the heterogeneous team was comprised of members from different companies, such a high degree of the diversity requires some sort of persona to maintain the connections between members. The person in the center can draw from many individuals including those with different backgrounds, knowledge, skills, and experiences. Because the person knows what is taking place within the network, and enjoys freedom and power, the individual will have sound judgment for risk cognition and confidence for taking risks. As a result, this person is more open-minded to different approaches or ways of thinking, leading to be more creative.

Hypothesis 1: The network centrality is positively related to individual creativity.

2.3 Knowledge sharing and Creativity

Knowledge sharing is the dissemination of knowledge. Knowledge sharing takes place between owners of knowledge and recipients of knowledge, so that the emphasis is on the exchange and the relationship that exists between the participants. Knowledge sharing behavior is a selective interpersonal process under specific circumstances. Knowledge givers not only choose with whom to share their knowledge, but they decide what knowledge to share.

When team members share new and relevant ideas with each other, it boosts their creativity. Knowledge sharing takes place when a person is eager to engage in knowledge collection and knowledge dissimulation for rising novel ideas. One of the positive outcomes of knowledge sharing is that new knowledge can be generated. Creativity needs new knowledge because knowledge can enhance an individual's learning, problem solving, and decision making [7]. Research finds that the intensifying role of knowledge sharing helps shape employee creativity. Given that creativity is the result of knowledge formation [13], employees with a higher level of knowledge sharing are more likely also creative. Recent studies confirm that employees with a high level of knowledge sharing are associated with tuning their creative potentials into creative outcomes.

Hypothesis 2: Knowledge sharing is positively related to individual creativity.

2.4 Knowledge Sharing as a Moderator

Given the importance of knowledge sharing, a great amount of research has been undertaken to explain the motivational factors for the transfer of knowledge between providers and recipients. One of such efforts relies on the social network theory to help explain why people share knowledge. In particular, the centrality of an employee's network position is one of the most prominent social network structural characteristics used by researchers to demonstrate that possessing a central network position predicts knowledge sharing in positive ways [6]. Central employees with more numerous network ties have more relationships to draw on to accumulate more knowledge. The knowledge accumulation positively affects their knowledge sharing with colleagues. Their accumulated knowledge and large networks as valuable resources help others to perceive them as attractive knowledge sharing partners. Because credibility is attached to their positions, employees are more likely to seek feedback from them. This further increases their engagement in knowledge sharing. Consequently, many opportunities exist for them to engage in a high extent of knowledge sharing with colleagues. In an empirical study, Research [6] shows that a central network position and the amount and diversity of the knowledge acquired by an employee are positively related. In addition, researchers find that centrally positioned organizational units in company's networks engage in more knowledge sharing, and are more innovative than those in the peripherals of the network.

We hypothesize the moderation role of knowledge sharing by reviewing its relationships with the network position and creativity, respectively. Our reasoning is that although the network position is important for creativity, the knowledge of the central individual still needs to be effectively shared with other team members. Research [7] finds employee creativity is strengthened through extensive knowledge sharing. Consistent

with prior research, we argue that a creative individual requires a strong sense of knowledge sharing. Specifically, the correlation between network central position and creativity is high only when an individual's belief that knowledge sharing can produce creative outcomes is high [7]. Hence, we propose our final hypothesis:

Hypothesis 3: Knowledge sharing moderates the relationship between centrality and individual creativity.

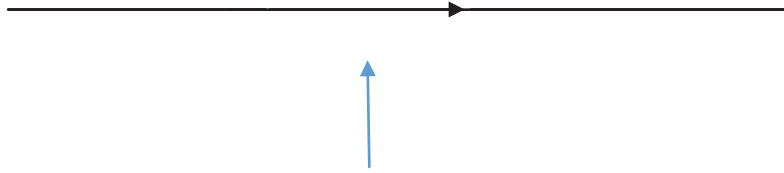


Figure 1 Research Model

3. RESEARCH DESIGN

For this research, we used a case from a software supplier's R&D project in a telecommunication company in China. In order to meet the project's demand, the parent company temporarily employed R&D workers from three other software companies. The temporary R&D team of the project is made up of 68 employees from four companies: 32 from the parent company, and 20, 8, and 8 from the three other companies, respectively.

3.1 Instrument development

We designed our instrument based on established constructs from the literature. The measure of individual creativity was based on an established scale, modified slightly to reflect the Chinese working environment. To measure knowledge sharing, we used measurement models based on a framework from [16]. We performed confirmatory factor analysis (CFA) on individual creativity and knowledge sharing using AMOS 7.0. The models provided a generally good fit to the data. For individual creativity ($\alpha = 0.899$), $\chi^2 = 17.6$, $df = 13$, $\chi^2/df = 0.74$, $RMSEA = 0.073$, $NFI = 0.939$ and $CFI = 0.934$. They all satisfied the suggested critical value requirements of $\chi^2/df \leq 2.5$, $RMSEA \leq 0.08$, $NFI \geq 0.9$ and $CFI \geq 0.9$. For knowledge sharing ($\alpha = 0.924$), $\chi^2 = 69.0$, $df = 34$, $\chi^2/df = 2.03$, $RMSEA = 0.11$, $NFI = 0.85$ and $CFI = 0.84$. Although $RMSEA$, NFI and CFI are slightly below the suggested values, we believed that the model was still acceptable given the overall goodness of fit χ^2/df is below 2.5. As for the network centrality, we used UCINET6 to calculate the work advice social network based on the survey data.

3.2 Data collection

Before we distributed the questionnaire, we first contacted the hosting company's Human Resources Department and the project manager. We stated the academic purpose of our research with the promise of keeping the data strictly confidential and sharing the research results with them. After obtaining their consents, we communicated with the project managers from the three partner companies and obtained their consents. The electronic version of the questionnaire was sent to the project managers via e-mail, who then distributed it to project team members. A total of 68 questionnaires were distributed. All respondents emailed their responses to the project managers, who then forwarded the responses to us. After receiving the responses, we first performed a visual screening to check for completeness and accuracy. With the support from the Human Resources Department, we obtained the basic information for each person, and reconciled the information with the completed questionnaires. After performing the steps above, we concluded that all 68 returned questionnaires were usable, yielding an effective response rate of 100%.

The demographics of our data show gender: male (72%) and female (28%); age: <30 (63%), 30-35 (24%), 36-40 (9%), and 41-45 (4%); tenure (years): <3 (75%), 3-5 (13%), and >5 (12%); and education: associate degree (24%), bachelor's degree (68%), master's degree (non-MBA) (6%), and MBA (3%).

4. RESULTS

Our dependent variable (y) is individual creativity. The independent variable is network centrality, with knowledge sharing as a moderator. As for the control variables, we control for gender, length of service, and education level because they might be associated with creativity through either gender induced cognitive flexibility differential, share identity, or task domain expertise [8]. We use the multiple regression and the results of the standardized coefficients are given in Table 1.

Table 1. Regression Analysis

	β
Network Centrality	0.225**
Knowledge Sharing	0.525***
Network Centrality* Knowledge Sharing	0.210**
Education	-.031
Gender	.284***
Tenure	-.091

** $p \leq .05$, *** $p \leq .01$

The results show that the regression coefficient of network centrality is 0.225 and significant at the 5% level, and that of knowledge sharing is 0.525 and significant at the 1% level. These findings demonstrate that both network centrality and knowledge sharing are positively correlated with creativity. The difference in their magnitudes shows that knowledge sharing has a stronger correlation than network centrality. Hence, both H_1 and H_2 are supported. More interestingly, the interaction term between knowledge sharing and centrality is significant and positive, indicating the moderation effect of knowledge sharing on the relationship between network centrality and individual creativity. This shows that a higher level of knowledge sharing strengthens the relationship between network centrality and creativity, while with a low level of knowledge sharing, the relationship between network centrality and creativity is relatively flat (see Figure 2). This result shows that H_3 is supported.

Among the control variables, only gender is significant, indicating that in this study the males were more creative than the females in the specific heterogeneous team, holding other variables constant. This may be attributable to the relatively small number of females (only 28%) on the team. While the findings of insignificant education and tenure are inconsistent with those in other research, we think the nature of the heterogeneous team may offer some explanation. Since the members on the team were from different companies, the tenure or seniority does not matter as much as when all the members had come from single firm. Similarly, the education level may play a role in promotions in the home company but not so much on a temporary team.

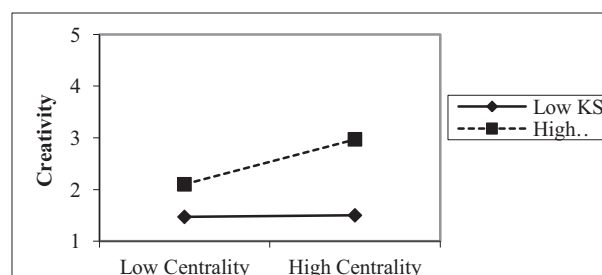


Figure 2. Cross level interaction plot

5. DISCUSSION

Our findings that network centrality and knowledge sharing are positively correlated with creativity. They are in agreement with the existing literature. The cross level interaction plot shows that a team member's knowledge sharing moderates the relationship between network centrality and individual creativity. That is, network centrality is more positively related to individual team member creativity when the level of knowledge sharing is high. When the knowledge sharing level is low, the relationship is somewhat flat. This finding is consistent with [7]. In addition, although our results show a positive correlation for creativity with both network centrality and knowledge sharing, the relationship is stronger with knowledge sharing than with network centrality. There are a couple of implication about this finding. First, in agreement with the existing literature, our study confirms that knowledge sharing plays important roles in helping individuals capitalize on the potential benefits of team diversity for their creativity. Second, the finding implies that network centrality by itself is not sufficient for a high level of individual creativity, as Figure 2 shows that the association between network centrality and creativity is stronger only when the level of knowledge sharing is high. This may be attributable to the environment of the heterogeneous team. When the team was first formed, members from the different companies still needed time to acquaint with each other. Hence, the team coordination would heavily rely on a few leaders from the hosting company. Those were most likely to occupy the central positions of the network. Occupying these positions did not necessarily imply that they were creative. As our findings show that the person at the central position must share their work related knowledge to be associated with the high level of creativity. Another possible explanation could be that other variables should be considered. Researchers have argued that central network position is insufficient if motivation and ability to share knowledge are not taken into account. Finally, we found that gender diversity was significant while other research did not confirm the significance. This may be sample specific and should be studied further in future research.

Our findings show the importance of knowledge sharing in fostering individual creativity. Managers in the parent company should provide strong support and incentives for team members to share knowledge, especially those who are in the central network positions. This can capitalize on the members' creative abilities and their location "linking" the diverse team members and allow them to generate innovative ideas. At the same time, managers should encourage team members from the parent company to share ideas and opinions with employees from other companies. This way, a team can make best use of members' knowledge to boost creativity.

The heterogeneous team model is based on reduced R&D costs and complementary knowledge that employees contribute to a team project. The core problem of integrating team members from various firms and completing software development projects is to develop ways to select best-fit employees and form an efficient R&D team. Under a heterogeneous team model, the company responsible for R&D team-building must develop ways to stimulate the employee knowledge sharing and creativity. Specific tasks include encouraging various types of communication among the employees from different companies, planning team activities that are aimed toward improving mutual trust, and awarding employees who are willing to share knowledge and help others.

6. CONCLUSION

This study provides insights about individual creativity in a heterogeneous team where the team members come from different independent companies. Based on existing research, we theoretically develop the cross-level interaction effect of knowledge sharing on the relationship between network central position and individual creativity. Using a data set collected on an IT company's project team that was formed under a heterogeneous team model with the members coming from four different companies, we empirically demonstrated that the relationship between network centrality and creativity is moderated by knowledge sharing.

Specifically, the relationship between network centrality and creativity is stronger when the level of knowledge sharing is high. When the knowledge sharing level is low, the relationship is almost flat. In addition, while we find that the correlations of creativity and network centrality and knowledge sharing are both positive, our results also show that the correlation with knowledge sharing is stronger. We draw theoretical implications for future research and practical implications for strategies to improve employee knowledge sharing and creativity in a heterogeneous team.

ACKNOWLEDGEMENT

This research is supported, in part, by a grant [Grant 71572196] from the National Natural Science Foundation of China.

REFERENCES

- [1] Ancona, D. G. and Caldwell, D.F. (1992). Demography and design: Predictors of new product team productivity. *Organization Science*, 3, 321-341.
- [2] Bantel, K. A. and Jackson, S.E. (1989). Top management and innovations in banking: Does the composition of the top team make a difference! *Strategic Management Journal*, 10, 107-124.
- [3] Burt, R. S. (1992). Structural Holes: The Social Structure of Competition. *Cambridge: Harvard University Press*, 207-301.
- [4] Granovetter, M.S. (1992). Networks and organizations, Structure, form and action, Problem of Explanation in Economic Sociology, *Boston: Harvard Business School Press*., 34-78.
- [5] Hansen, T. M. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization sub-units. *Administrative Science Quarterly*, 44(1): 82-111.
- [6] Jie, X.M. and Zuo, L.L. (2013). Characteristics of collaborative innovation networks and innovation performance of firms: The mediating effect of knowledge absorptive capacity. *Nankai Business Review*, 16(3): 47-56.
- [7] Mittal, S., and Dhar, R. L. (2015). Transformational leadership and employee creativity: mediating role of creative self-efficacy and moderating role of knowledge sharing. *Management Decision*, 53(5), 894-910.
- [8] Mumford, M. D. and Gustafson, S. B. (1988). Creativity syndrome: Integration, application, and innovation. *Psychological Bulletin*, 103: 27-43.
- [9] O'Reilly, C. A., Caldwell, D.F. and Barnett, W.P. (1989). Work group demography, social integration, and turnover. *Administration Science Quarterly*, 34, 21-37.
- [10] Pfeffer, J. (1983). Organizational demography: Implications for management. *California Management Review*, 28, 67-81.
- [11] Reagans, R., & Zuckerman, E. W. (2001). Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organization science*, 12(4), 502-517.
- [12] Uzzi B. (1997). Social structure and competition in interfirm networks: the paradox of embeddedness. *Administrative Science Quarterly*, 42(1): 35-67.
- [13] Wang, M.S. (2010), "Influence of knowledge sharing and project complexity on group creativity: taking the development of information system for example", *Journal of e-Business*, Vol. 12. No. 1, pp. 73-102.
- [14] Wu, X.B. and Liu, X.F. (2007). the research of technology transfer processes in global manufacturing networks. *Technology Economics*, 26(2): 1-6.
- [15] Zenger, T. R. and Lawrence, B.S. (1989). Organizational demography: The differential effects of age and tenure distributions on technical communication. *Academy of Management Journal*, 32, 353-376.
- [16] Zheng, R.W. and Li, S.Q. (2001). The relationships of organization justice, trust and knowledge sharing behaviors. *Journal of Human Resources Management*, 1(2): 69-93.

A Net Loan Monitoring Platform for University Students Based on Visual Micro-blog

Ruijun Zhang^{1,2}, Caiyan Lin², Xinxin Zheng², Lixin Gao², Jiaxin Tong², Shangyang Yu³

¹Center of Service Science and Engineering, Wuhan University of Science and Technology,
Wuhan, 430065, China

²School of Management, Wuhan University of Science and Technology, Wuhan, 430065, China

³College of Science, Wuhan University of Science and Technology, Wuhan, 430065, China

Abstract: In order to prevent college students from tragic events, which mostly result from falling into illegal net loan, a net loan risk monitoring platform of college students was established based on the microblog visualization. Based on the calculation model of risk degree (CMRD) and the calculation model of relational closeness (CMRC), a visual system, a user relationship analysis system and an early warning system were constructed on the platform. Through the CMRD, the risk degree of micro-blog net loan contents can be worked out to decide whether warning behaviors are necessary. Through the CMRC, the closeness of net loan bloggers with users can be obtained a propagation map. The key nodes can be cut off by internet security officers in time, which can prevent further propagation of the contents of the microblog net loan. The platform can effectively alleviate the tense situation of college students and illegal online loans.

Keywords: college students' net loan, microblog visualization, risk monitoring, calculation model of relationship closeness, calculation mode of risk degree, propagation map.

1. INTRODUCTION

Net loan is a new industry of Internet finance in recent years. It has the advantages of low interest rate, simple procedure, installments repayment, etc, which is exactly popular among young people. However, it is also easy to be used by illegal merchants ^[1]. Earlier, some scholars conducted research on bad net loans: Istrate E and others analyzed the extent of bad net loan issues, the causes and the solutions adopted to resolve the bad net loans ^[2]. In recent years, the high frequency new net loan events as “nude loan” and “usury” impact college students badly, which sometimes even threaten their lives ^[3]. The spread of net loans relies heavily on social media and web advertising. As a new form of social media, microblog Microblog provides a sharing platform for users to share new things and express new opinions ^[4]. Therefore, more and more net loan platform and personal loan borrowers in Microblog posted net loan information to attract the attention of college students, but some of the information is with safe risk, and serious consequences may be made without preventing behaviors.

Domestic and foreign scholars have done some researches on visualization and online public opinions. In visualization, Zhao Hua and others made visualization analysis in Microblog users interest identification results ^[5], Zhao J, et put FluxFlow combined with machine learning in identifying abnormal behavior about popular topics of social media ^[6], Michael Jendryke and others used the method of Microblog visualization on Sina Weibo to analyze the relationship between news and census ^[7], Siming Chen and others made visualization analysis of large amounts of data on social media research, summarizes the visual analysis process ^[8]. In terms of internet public opinion, Xue Gang Chen and others made the network public opinion trend prediction and evaluation research ^[9], Gu Q and others analyzed the evolution process of network public opinion over time and information source quality evaluation with web crawler tools, which can automatically detect and track hot

public opinions ^[10].

Based on the above theoretical basis, this paper studies a set of early-warning mechanism in the aspect of college students' net loan risk monitoring. In addition, the improvement and innovation is made on Kuo Y H and others' comprehensive analysis method of micro-blog emotion from user's social interaction mode and text view ^[11], the risk calculation model and relationship intimacy calculation model is made. On the basis of these two models, the net loan risk monitoring system of college students is constructed, which can effectively analyze the risk net loan information and adopt corresponding early-warning behaviors.

2. CORRELATION MODELS

2.1 Calculation model of risk degree

The overall risk value of microblog content can be defined as *DaValue*.

A complete microblog contains emoticons, pictures, links and words, so the calculation model of risk degree needs to be considered from these four aspects.

2.1.1 Building a list of net loan dangerous vocabulary

The list of risk words for net loan contains the following three subsets:

- The high frequency words selected from the examples of microblog, such as *The Next Day's Money*, *Small Private Loan*, *Emergency*, etc. The risk value of the *i*-th high-frequency vocabulary is set as a_i .
- The names of several major campus loan platforms include aliases, such as *Credit Ease*, *Elite Loans*, *99 Staging*, etc. The risk value of *i*-th campus loan platform is set as b_i .
- There are common terms of loan, such as Issuing a bid, *Bidding and Freezing a deposit*, etc. The risk value of *i*-th loan common word is set as c_i .

If there is a risk-word list of net loan (denoted as T_{dan}) in the microblog content as a stored word, then it is set according to the frequency of its occurrence and the risk value of its similar words stored in the list. If its risk value exceeds the lowest value in T_{dan} , then classify it into T_{dan} .

2.1.2 R language processing participle and definition of dangerous values

The Rwordseg program package in R language is used to handle participle words by removing punctuation and dividing the sentences into several words or phrases, and the table function in R language is used for statistic f_r , which denotes the frequency that each phrase appears. High-frequency words are listed into high-frequency vocabulary in the T_{dan} . In daily language, terms, which are meaningless to the whole sentences, such as *do*, *is* and *have*, are extremely frequent and can be removed from the T_{dan} by computer or hand. The risk value of the name of some campus loan platforms are set in accordance to the development of the platform and its credit. But those frequent words about loan are set value according to their importance in the process of net loan.

As to the expressions and pictures obtained, their risk values can be set correspondingly to their definition combined with its multiplexing values in the field of net loan. Suppose the risk value of the *i*-th expression is δ_i and the *i*-th picture is λ_i .

As to the capture link, the final risk value can be figured out by analyzing its content, handling with participles and the risk judging of each word. Suppose the risk of the *i*-th dark link is π_i .

$$\pi_i = \sum \frac{n_x}{n} \sum \pi_{x,y} \quad (x = a, b, c, \delta, \lambda, \pi) \quad (1)$$

In the formula, n represents the total number of the participle words after segmentations, n_x represents the

number of the words of x , and $\pi_{x,y}$ represents the risk of the y -th word of x type.

2.1.3 Calculation model of risk degree

Combined with the above description, the risk value of microblog content is recorded as:

$$CoValue = \sum X \quad (X = a_i, b_i, c_i, \delta_i, \lambda_i, \pi_i) \quad (2)$$

Besides the impact of content, the risk of microblog net loan is also related with the force of the publishers, which shows specifically on publisher's fans F_c , amount of its likes D_c , amount of its comments L_c and amount of its retweets Z_c [12]. The publisher's influence is recorded as *InValue*.

$$InValue = \sum \lg(x+1) \quad (x = F_c, D_c, L_c, Z_c) \quad (3)$$

Combined with the above description, the calculation model of comprehensive risk degree is obtained as follow:

$$DaValue = InValue \cdot CoValue \quad (4)$$

2.2 Calculation model of relationship closeness

In order to predict the propagation direction of net loan information and understand the relationship between net loan creditors and borrowers, a calculation model of relationship closeness was established on the basis of LDA's microblog user fan intimacy evaluation model [13]. Through analyzing microblog user's relevance, interaction degree and interest similarity, the final calculation model of relationship closeness is ascertained. The model was used to predict the possibility of fans reposting information from bloggers. It should be noticed that the propagation pattern of net loan is called multi-level. The following algorithm is based on the first level of propagation, and closeness values of other levels of propagation can be calculated by this method.

2.2.1 Microblog user's relevant relationship

The relationship between users can be divided into four kinds as follows: Attention (the blogger single concerns), the fans (the uses single concern blogger), mutual fans (the blogger and fans concern mutually) and no relation. If Microblog nodes only establishes the relationship of one-way attention or the fans, it cannot show that there is a mutual trust relationship between both sides, when the two sides reach a certain familiarity forming the friend relationship, interactive feedback has a certain similarity between nodes [14]. Based on this relationship, we believe that the probability of forwarding in these four relationships is from high to low: mutual fans, fans, attention, and no relation. Therefore, the correlation can be simply expressed as equation (5).

The relevance is defined as *Att_degree*:

$$Att_degree = \begin{cases} 0 & No\ attention \\ 1 & Attention \\ 3 & Fans \\ 4 & Mutual\ fans \end{cases} \quad (5)$$

2.2.2 Interaction strength of Microblog users

According to its characteristics, users can use microblog to interact with others by liking, retweeting, @ and commenting. These four methods are also most commonly used. The degree of interaction can be set with these four interactive ways. From the perspective of social capital theory, closeness relationship is based on social interaction [15], the frequency of interaction determines the degree of closeness and the lasting maintenance of the relationship.

In four kinds of interaction, the theoretical research shows that the like is the most simple interaction, forwarding behavior is also relatively simple, and the comment is relatively deep interaction [14], @ was conducted on the basis of the forwarding or comments, it can be simply represented as equation (6).

$$I = \begin{cases} 1 & \text{Like} \\ 2 & \text{Comment} \\ 3 & \text{Retweet} \\ 4 & \text{@} \end{cases} \quad (6)$$

In order to reflect the interaction between microblog users and fans better, this paper uses bidirectional calculation and weighting method to represent. Define R_{ui} as the interaction degree between the users and his i -th fan^[13]. Combined with equation (6), it can be denoted as:

$$R_{ui} = \frac{I}{2} \left(\frac{\sum D_{fi} \cup L_{fi} \cup Z_{fi} \cup A_{fi} + \sum D_{bi} \cup L_{bi} \cup Z_{bi} \cup A_{bi}}{t} \right) \quad (7)$$

In the formula, D_{fi} , L_{fi} , Z_{fi} and A_{fi} denotes the number of the liking, commenting, retweeting and @ of the i -th fan f_i for the blogger respectively in a certain period of time t . D_{bi} , L_{bi} , Z_{bi} and A_{bi} denotes the number of liking, commenting, retweeting and @ of the blogger for the i -th fan respectively in a certain period of time t .

2.2.3 Interest similarity of microblog users

Interest similarity of microblog users can be analyzed by analyzing the topics they participated and the microblogs they retweeted, the final total interest similarity of users will be recorded as S_{bi} .

When a microblog topic is discussed by user, it will be added # *topic name* # and its hyperlinks display in blue words. Each topic that the user participated in will be compared with the topic that the i -th fan participates in. If it is the same, then T_{bi} increases by one, where T_{bi} denotes the degree of topic similarity between the blogger and his i -th fan in a period of time t , and the initial value of T_{bi} is 0.

Users can also forward something according to their own interests. They can attach their own sharing experience, and the bottom of their retweet includes blue words as @ *the original blogger*. By comparing the words, the similarity of interest between users and fans will be reflected further. Through the topic similarity method and contrasting one by one, if it is same then the similarity F_{bi} plus one, where F_{bi} denotes the degree of topic similarity between the blogger and his i -th fan in a period of time t , and the initial value of F_{bi} is 0.

The final interest similarity of users:

$$S_{bi} = \frac{2 \left[(B_i - 1) \frac{T_{bi}}{B_i} + (F_i - 1) \frac{T_{bi}}{F_i} \right] \cdot \sqrt{\frac{1}{B_i} + \frac{1}{F_i}} + \left[(B_{ir} - 1) \frac{F_{bi}}{B_{ir}} + (F_{ir} - 1) \frac{F_{bi}}{F_{ir}} \right] \cdot \sqrt{\frac{1}{B_{ir}} + \frac{1}{F_{ir}}}}{(B_i + F_i - 2)t} + \frac{\left[(B_{ir} - 1) \frac{F_{bi}}{B_{ir}} + (F_{ir} - 1) \frac{F_{bi}}{F_{ir}} \right] \cdot \sqrt{\frac{1}{B_{ir}} + \frac{1}{F_{ir}}}}{(B_{ir} + F_{ir} - 2)t} \quad (8)$$

In the formula, B_i represents the total number of topics that bloggers participated in the discussion in a period of time t , F_i represents the total number of fans participated in the discussion of microblog topics in a period of time t , B_{ir} represents the total number that bloggers retweeted microblogs in a period of time t , F_{ir} represents the total number of microblogs retweeted by fans in a period of time t .

2.2.4 Calculation model of relationship closeness

Combining the above three aspects, the calculation model of relationship closeness can be obtained. C_{bi} is set to represent the comprehensive closeness between bloggers and the i -th fan:

$$C_{bi} = \prod X \quad (X = R_e, R_{ui}, S_{bi}) \quad (9)$$

3. RISK MONITORING PLATFORM OF UNIVERSITY STUDENTS' net loan BASED ON THE MICROBLOG VISUALIZATION

Based on the above calculation model of relational closeness and calculation model of risk degree, a net loan risk monitoring platform of college students is designed, which can be generally divided in visualization system, analysis system of users' relationship and early-warning system. The architecture is shown as Figure 1.

Visualization System: Web crawler crawls cases related with microblog net loan, through the R language and

microblog visualization tools of Peking University (or PKUVIS), and then several propagation relationship diagrams will be obtained. After words segmenting, data filtering and other operations, the system stores the organized data into the database platform. Meanwhile, experts in the field of net loan can add some information that it is hard for the system to identify, such as some sensitive words that has not been recorded into T_{dan} for some time.

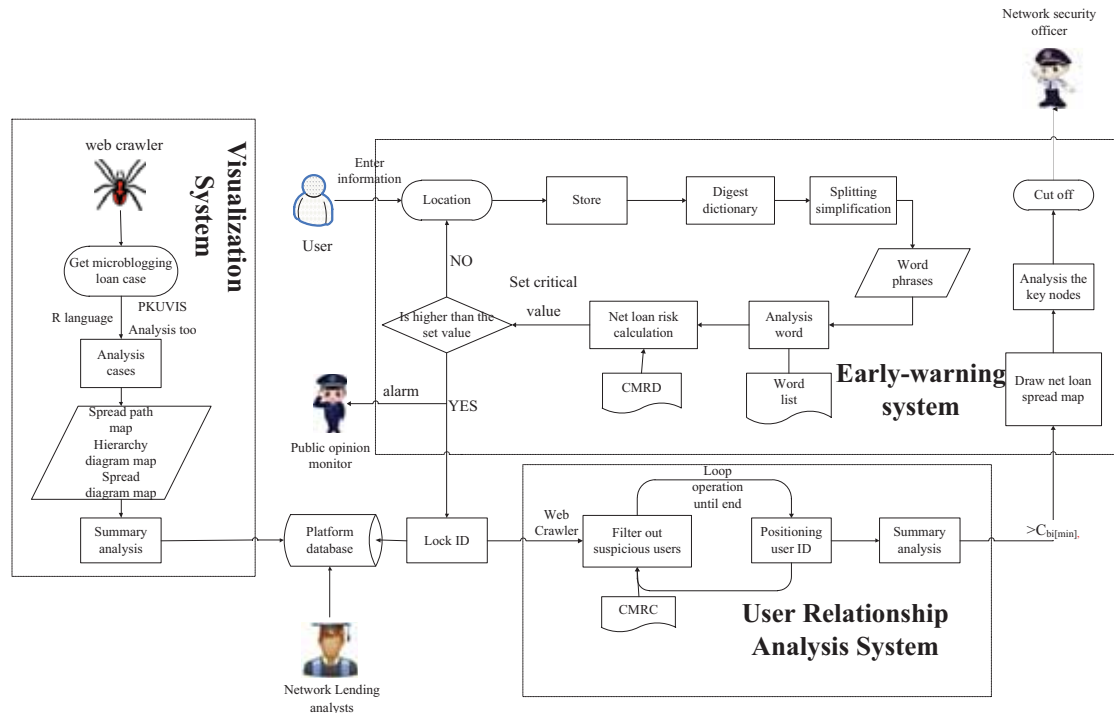





Figure 1. Platform architecture

Early-warning System: the platform define the position of information input by users, split the content sentences ^[16] and simplify the data, and calculate the overall danger value $DaValue$ according to the risk calculation model of T_{dan} . If it belongs to first kind of risk, sign it as , platform will ask public opinion supervisors to make warnings and lock the propagation resource, combined with analysis system of microblog users relationship select which is greater than the minimum value $C_{bi[min]}$, the net loan transmission graph is obtained and the key nodes are analyzed, and then the net security officer is involved to cut it off. If it is the second kind of risk, the platform will ask the user whether to cut off the key node of the transmission graph, and sign it as . If it is the third kind of risk, the platform will continue to pay attention to this information and sign it as .

User Relationship Analysis System: The platform focuses on the high-risk owners' ID and locates the suspicious users of various stages of propagation to summarize and dealt with, through the network crawler and CMRC.

4. DEVELOPMENT OF PLATFORM AND EXPERIMENTAL DATA

4.1 Development technology

The platform uses B/S architecture to develop, SQL Server 2008 is set as the platform of the background database, R language is used to handle the business logic of the middle business layer. In addition, C# language and the .NET Framework are applied as front-end Web development tools, the PKUVIS Analysis tool of Peking University is used as the information display tool of microblog visualization.

4.2 Experimental Data

Firstly select a microblog about net loan. Then the result, which contains the experimental data, propagation route map and the distributed map of propagation information, will be obtained by functional testing through the platform. The core of processing data in the platform is parallel data mining because the Internet Information is complicated and gigantic^[17].

For example, there is the microblog news about net loan that has been forwarded for more than 2000 times. The original content is as [Xiamen net loan suicide girl's last week: under crazy struggle, she considered suicide for more than once]. No one knows Xueqi (a pseudonym) experienced what kind of psychological struggle in a small hotel near a university in east of Quanzhou. On the early morning of April 11, she chose to end her life and the police initially regard it as suicide by burning charcoal. On the 14th, the paper contacted with two of her closest friends who told reporters: <https://wx4.sinaimg.cn/large/005vnhZYgy1feontjyglrj30fq0f20x5.jpg>. To reduce redundancy, a propagation chain is taken as an example.

Set $C_{bi[min]} = 0.00020$, the critical value of the first kind of risk is 230.0, the critical value of the second kind of risk is 100.0

Table1. The information chart of original microblog data

No.	User name	Retweetings	Likes	Comments	Fans	Number of microblogs	Concerns
1	Pengpai News	6137.00	22440.00	23848.00	5604398.00	28434.00	310.00

Table2. The information chart after the split of original microblog

No.	High frequency word	Risk value	High frequency word	Risk value	High frequency word	Risk value
1	Net loan	4.68	Commit suicide	5.78	Myself	0.78
2	Suicide	5.88	No	1.23	Life	2.65
3	Girl	2.12	College	3.24	Police	3.33
4	Last	3.14	Hotel	1.87	First step	1.24
5	A week	4.33	Experience	1.67	Identification	2.66
6	Crazy	3.12	Psychology	1.21	Choice	0.54
7	Dun	4.55	Struggle	2.45	End	4.21
8	More	1.23	Early morning	1.22	Thought	0.54
9	Once	0.78				

According to the content of this microblog and combining with CMRD:

Risk value of high frequency word: $a = 4.68 + 5.88 * 3 + \dots + 0.54 = 76.21$;

Risk value of platform word: $b = 0$;

Risk value of loan words $c = 0$;

Risk value of expressions $\delta = 0$;

Risk value of pictures $\gamma = 2.33$;

Risk value of dark links $\pi = 0$;

Risk value of content information is obtained with equation(2): $CoValue = 78.54$

Risk value of publishers' force is obtained with equation (3): $InValue = 19.27$

Risk value of final information is obtained with equation(5): $DaValue = 1513.4658 \in [230, +\infty)$ It is the first kind of risk and it is necessary to obtain the propagation graph according to CMRC

Table3. The required information chart of CMRC after handling

Number	Dfi/t	Dbi/t	Lfi/t	Lbi/t	Zfi/t	Zbi/t	Afi/t	Abi/t	Re
1	7.00	-	3.00	-	0.00	-	0.00	-	-
2	0.00	2.00	1.00	4.00	0.00	2.00	0.00	0.00	1.00
3	5.00	1.00	1.00	2.00	0.00	2.00	0.00	0.00	1.00
4	1.00	3.00	1.00	4.00	1.00	1.00	0.00	0.00	1.00
5	0.00	0.00	2.00	2.00	0.00	1.00	0.00	0.00	1.00
6	-	1.00	-	0.00	-	1.00	-	0.00	3.00

In the table, $t = 48h$.

The degree of interaction and interest similarity between users at all levels is obtained with equation (7), (8) and Table 3.

$$R_{1,2} \approx 0.35, R_{2,3} \approx 0.15, R_{3,4} \approx 0.26, R_{4,5} \approx 0.15, R_{5,6} \approx 0.15$$

$$S_{1,2} \approx 0.016, S_{2,3} \approx 0.008, S_{3,4} \approx 0.003, S_{4,5} \approx 0.026, S_{5,6} \approx 0.010$$

The ultimate degree of closeness at all levels is obtained with equation (9)

$$C_{1,2} = 0.00560, C_{2,3} = 0.00120, C_{3,4} = 0.00078, C_{4,5} = 0.00390, C_{5,6} = 0.00270$$

They are all more than or as same as $C_{b[\min]} = 0.00020$, and a propagation graph in the red box of figure 2 is obtained.

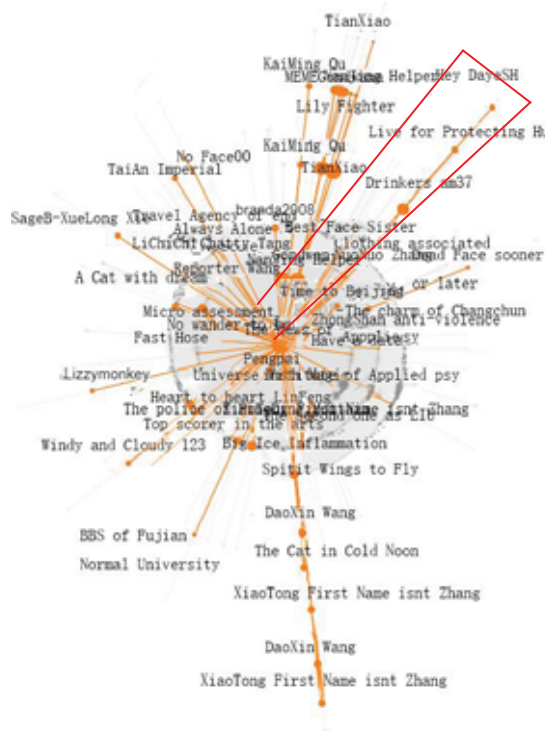


Figure 2. Propagation Route Map

In Figure 2, each circular node represents a user, and the others retweet and comment around the central microblog. The bigger the node is, the greater the impact of this user will be. By combining the propagation route map with those second-retweet users IDs, it can be observed more distinctly and directly.



Figure 3. Distributed Map

According to the information of selected microblog, the distributed map can be drawn with population distribution. Each point represents each blogger who retweeted. In the same region, the intensity of the points is proportional to the amount of net loan information forwarding. From Figure 3, it is clear that the retweet frequency of this microblog net loan reaches highest in Guangdong Province. Therefore, its network supervision department should focus on it and take targeted programs.

5. CONCLUSION

The article designed a net loan risk monitoring platform of college students based on the microblog visualization. The platform combines calculation model of intimacy and risk degree, it can analyze and predict the transmission direction of net loan information in time, and draw the net loan transmission map. By cutting off the key nodes, the internet security officers will stop the illegal net loan information from the source and avoid the occurrence of malignant events. The platform takes advance warning mechanism of microblog contents, which can effectively prevent illegal net loan information through micorblog into the field of college students, and provide supervision departments convenient tools to control the spread of illegal net loans. The methods studied in this paper can purify the environment of microblog to some extends, reduce the probability of college students falling into illegal net loans and protect the safety of life and property of college students.

ACKNOWLEDGEMENTS

This research is supported by the 2017 National Undergraduate Innovative Training Project - An Net Loan Monitoring Platform for University Students Based on Visual Micro-blog, under the Grant NO. 201710488005 , by the Humanities and Social Sciences Research Project from the Hubei Provincial Department of Education, under the Grant NO.18D014, by the Service Science and Engineering Research Project from Wuhan University of Science and Technology ,under the Grant NO.CSSE2017GA07.

REFERENCES

- [1] Cai Yu.(2016). Current situation and strategy of college students' online loan [J]. Modern economic information, (15):309+311(in Chinese).
- [2] Istrate E, Das Gupta D, Weissburg P. (2007). Toward Developing a Structured Approach to the Diagnosis and Resolution of Nonperforming Loans: the Case of China and India[J]. Review of Policy Research, 24(4):345-365.

- [3] Zhang Jie, Mu Shu. (2017). Analysis of college students' online loan cognition and its risk prevention -- based on the survey data of tianjin university [J]. Credit, 35(4):22-26(in Chinese).
- [4] Zhao Y, Qin B, Liu T, et al.(2016). Social sentiment sensor: a visualization system for topic detection and topic sentiment analysis on microblog[J]. Multimedia Tools & Applications, 75(15):8843-8860.
- [5] Zhao Hua, Ji Xiaowen, Zeng Qing-tian, et al. (2015). Interest identification and visualization of microblog users based on topic related space [J]. Computer science, 42(S1):500-502+509 (in Chinese).
- [6] Zhao J, Cao N, Wen Z, et al. (2014). FluxFlow: Visual Analysis of Anomalous Information Spreading on Social Media[J]. IEEE Transactions on Visualization & Computer Graphics, 20(12):1773-1782.
- [7] Michael Jendryke,Timo Balz,Mingsheng Liao.(2017). Big location - based social media messages from China's Sina Microblog network: Collection, storage, visualization, and potential ways of analysis[J]. Transactions in GIS, 21(4):825-834.
- [8] Chen S, Lin L, Yuan X.(2017). Social Media Visual Analytics[J]. Computer Graphics Forum, 36(3):563-587.
- [9] Chen X G, Duan S, Wang L. (2017). Research on trend prediction and evaluation of network public opinion[J]. Concurrency and Computation: Practice and Experience, 29(24): 4212-4218.
- [10]Gu Q, He X, Wang x. (2014). Study on evolution trends of network public opinion based on hyperlink analysis[J]. Journal of Digital Information Management, 12(6):421-428.
- [11] Kuo Y H, Fu M H, Tsai W H, et al.(2016). Integrated microblog sentiment analysis from users' social interaction patterns and textual criticism [J]. Applied Intelligence,44(2):399-413.
- [12] Lliu Gongshen, Meng Kui, Xie Jing.(2014). An early warning algorithm for microblog [J]. Computer science, 41(12):33-37(in Chinese).
- [13] Wang Qiuxun, Yu Haoliang, Xu Haocheng, etc. (2016). Based on LDA's microblog user fan affinity evaluation model [J]. Computer application and software, 33(10):67-71(in Chinese).
- [14] Xiong Jianying. (2017). A study on the model of emotional anomaly of microblog users based on trusted feedback [J]. Intelligence science, (4):48-53(in Chinese).
- [15] Yao Zhengyu.(2013). Research on Intimacy and Social Network Interaction Behavior[C].//Communication and China-Fudan Forum. Networked Relationship: New Communication and Current Chinese Proceedings, 1(in Chinese).
- [16] Wubben S, Antal V D, B Krahmer e. (2012). Sentence Simplification by Monolingual Machine Translation [C] / / Proceedings of the Annual Meeting of the Association for Computational Linguistics, 1015-1024.
- [17] Ni Yingjie, Wang Yuke, Zhang Jun. (2009). Network mass information processing platform based on high-performance data mining [J]. Computer engineering and science, 31(S1):129-132(in Chinese).

Inhibitors of Continuance Intention to Use Mobile Social

Networking Sites: The Effects of Stress and Regret

Chunmei Gan, Tingting Li

School of Information Management, Sun Yat-sen University, China

Abstract: [Purpose /significance] This study aims at exploring factors influencing continuance intention to use mobile social networking sites (SNS). Specifically, this study attempts to investigate the effects of stress and regret. [Method /process] 226 valid data from working stuff were collected via questionnaires and further analyzed using SEM. [Result /conclusion] Results show that, stress and regret have significant impacts on continuance intention to use mobile SNS. Also, technology-work conflict and technology-personal conflict have significant influences on stress, and excessive use has a significant effect on regret.

Keywords: mobile social networking sites, continuance intention, stress, regret, user behavior

1. INTRODUCTION

With the development of mobile technologies and smart phones, mobile social networking sites (SNS) (such as WeChat) become more popular in individuals' life. Tencent official report reveals that WeChat monthly active account has reached 980 million, a year-on-year growth of 15.8% in the third quarter of 2017 ^[1]. However, as the competition is fierce, user continuance is still a big issue for service providers. Considering that SNS users have a high motivation and the freedom to discontinue using an SNS platform, especially when they feel the need to avoid stressful or self-evaluated feelings of painful emotion situations, it is necessary to investigate the inhibitors of continuance intention to use mobile SNS. Previous studies have mainly focused on the motives that foster continuance to use mobile SNS. To better understand factors that affecting user behavior of mobile SNS, more attention should be paid to the inhibitors.

Prior studies have investigated the factors affecting SNS user behavior. For example, Chaouali revealed that information overload and social overload affect emotional exhaustion, further influences mobile SNS continuance intention ^[2]. Also, stress is found to be negatively associated with SNS user behavior. Luqman et al. found that stress can result in the decision to quit Facebook ^[3]. In addition, several studies have revealed the effects of regret on IS user behavior. Woollaston confirmed that use context, which is the excessive or improper use of intrinsically rewarding information system, can violate users' moral norms and values ^[4]. Turel examined the antecedent of regret feeling, satisfaction and consequences of discontinuance intention to use Facebook ^[5]. As can be seen from prior studies, both stress and regret make users feel uncomfortable, so they would experience a cognitive dissonance, and consequent motivation for corrective action will arise. However, to reveal the effects of stress and regret in the context of mobile SNS, more attention should be paid to such questions as how stress and regret develop as well as what effects they have on user continuance. Thus, this study aims to explore the inhibitors that influence user continuance on using mobile SNS.

In addition, Zheng and Lee found that technology-personal conflict, technology-work and technology-family conflict have significant impacts on stress, but they did not study the negative effects of stress on continued use ^[6]. Also, prior research argued that users who tired or drained from doing something develop the intention to shape their current situation ^[7]. Therefore, this study attempts to integrate the effects of stress and regret, and develops an integrated model to explore the inhibitors affecting user continuance of mobile SNS. Data were collected from 226 working stuff via questionnaires, and further analyzed using

structural equation model. The findings of this study have both theoretical and practical contributions.

2. RESEARCH MODEL AND HYPOTHESES

Stress refers to an individual's psychological response to a stressor such as an environmental condition^[8]. Users who are exhausted from doing something may develop the intention to change his or her behaviour^[7], such as temporarily or permanently discontinue using IS. Prior studies have shown that stress is an inhibitor that influences continue intention to use SNSs. Beauty and Pinsonneault confirmed that stress is an inhibitor that influences their intention to continue using SNSs^[9]. Luqman et al. found that psychological and behavioral consequences lead users to discontinue or reduce the use of Facebook due to stress^[3]. In addition, Maier et al. found that discontinuous use is users' coping strategy toward stress resulting from SNS exhaustion and social overload^[10]. For example, when users consider using a stressful mobile SNS is a problematic behavior, they will feel more exhausted, and intentions to discontinue SNS usage arise. The higher stress the users feel, the less they will intend to continue using the services. Hence, the following hypothesis is developed.

H1: Stress negatively affects continuance intention.

Regret is described as a painful emotion, accompanies actual or contemplated violation of internal values and rules in any situation^[11]. Users may feel regret due to their unpleasant feelings about regret-producing system or their usage behaviors that make them dissatisfied with themselves^[5]. To reduce their regret, users may modify their current situation when they realize that their behavior can't satisfy their expectations^[12]. It is argued that regret has a significant impact on user continuance. Woollaston confirmed that some hedonic IS use context, where the excessive or improper use of intrinsically rewarding IS, can violate users' moral norms and values^[4]. Based on social cognitive theory and theory of planned behavior, respectively, Turel confirmed that regret feeling is a unique driver of IS discontinuance decisions^[5]^[13]. That is to say, regret, which refer to subjective and self-evaluated feelings of painful emotion from SNS usage would be involved in the judgmental process, and influence discontinuance decisions. For example, when users sense regret, they will feel something is wrong, make them more aware of their problems, and therefore increase their motivation and ability to reflect on the discontinuance decision. The more regret users develop, the less intention they will have of using the services continuously. Therefore, the following hypothesis is put forward.

H2: Regret negatively affects continuance intention.

Technostress reflects users' psychological reaction of stress caused by their experience of using IS in the organizational context^[14]. Previous studies investigated the techno-stressors that create technostress^[15]. For example, Fox et al. indicated the interrelationship between technostress and strain among intensive ICT users^[16]. Drawing from P-E model, stress stems from the absence of equilibrium in the relationships. When individuals are limited in time, energy, and even body health conditions, the increased inappropriate use of MSNSs in different environments could lead to conflicts with family, work or school. Zheng et al found that because of the widespread use of social networking sites, stressors which create from family relationships, external environment, and personal conflicts are technology-family conflict, technology-work conflict, and technology-personal conflict^[6]. For example, when individuals produce conflict with personal, family, or work due to using mobile SNS, it will affect their body health, work performance, and reduce their quality of life. Such phenomenon may make them feel nervous and stress. The more technology-family, technology-work, and technology-personal conflict the users feel, the higher stress the users feel. Thus, the following hypotheses are assumed.

H3a: Technology-family conflict positively affects stress.

H3b: Technology-work conflict positively affects stress.

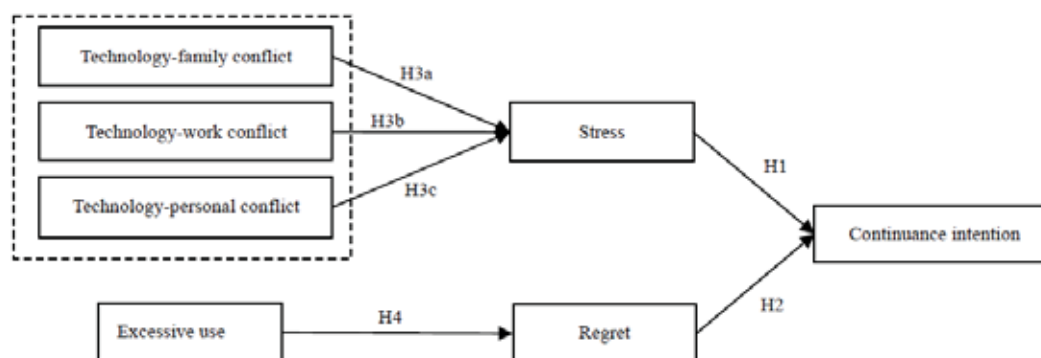
H3c: Technology-personal conflict positively affects stress.

Excessive use reflects the extent to which IS use is longer than the time planned ^[17]. Luqman indicated that excessive social, hedonic, cognitive use may be perceived as a negative stimulus and an intrusion in an individual's life ^[3], thereby causing regret. Turel found that in some cases the use of mobile SNS such as Facebook and especially excessive use and conflict stemming from high levels of addiction to such systems can present such violations that result in regret feelings ^[5]. For example, when users sense they lose control over the time they spend on website, they experience a discrepancy between what they believe is right to do, and what they are actually doing, and regret feelings arise. The greater the degree of unreasonable use, the more regret users develop. Thus, the following hypothesis is proposed.

H4: Excessive use positively affects regret.

Figure 1 presents the research model.

Fig 1. Research model



3. RESEARCH METHODOLOGY

3.1 Instrument development

To ensure validity and reliability, items were adapted from prior studies. Items for technology-family conflict were adapted from Turel ^[18]. Items for technology-work/study conflict were adapted from Hong et al. ^[19]. Items of technology-personal conflict were adapted from Zheng and Lee ^[6]. Items for stress were adapted from Ayyagari et al. ^[15]. Items for excessive use were adapted from Scott and Andrew ^[17]. Items for regret and continuance intention were adapted from Turel ^[5] and Bhattacharjee ^[20]. All items were measured with a seven-point Likert scale, ranging from “strongly disagree (1)” to “strongly agree (7)”. Appendix shows the items in this study.

Since the investigation was carried out in the context of Chinese mobile social networking sites, the method of back-translation was used to ensure equivalence of translation ^[21]. A native Chinese speaker translated the original English questionnaire into Chinese. Then another researcher translated the questionnaire into English to ensure content is consistent. And they reached agreement on the final version of the questionnaire. In addition, 25 users with rich usage experience were invited to participate in the pilot study and their feedback to the questionnaire was used to modify and improve the comprehensibility and clarity of the measurement items.

3.2 Data collection

The data were collected from working stuff, and a professional survey platform sojump was used to distribute our questionnaires online. The data used in this study were collected in China, where there are noticeable characteristics in the usage of Internet and SNSs. CNNIC official report reveals that the number of mobile Internet users reached 753 million in China, an increase of 57.34 million from the end of 2016, and 97.5% of Internet users use mobile phones to access the Internet ^[22]. Furthermore, this study did not focus on

any specific mobile SNS, because most of the popular SNS applications have similar functions. This will ensure good generalizability of this research. Data were collected for two weeks and 244 completed questionnaires were obtained. After eliminating outliers and removing invalid responses, 226 valid samples were used for further analysis.

Table 1 depicts the demographic information of the respondents. With respect to the demographics of the respondents, the current research finds that 54% of the respondents are females and 46% are males. The age of most respondents ranged from 23 to 27 years. 62.18% use mobile Facebook every day, 17.13% from 3 to 6 days a week, and 20.67% from 1 to 2 days a week. 54.9% have joined Facebook for more than 3 years, 10.2% from 2 to 3 years, 16.8% from 1 to 2 years, and 18.2% for less than 1 year. 48.7% used mobile SNS frequently each day, 25.7% several times a day, 10.6% more than ten times a day, 6.6% once a day, 4.4% several times a week, and 4.0% rarely use.

Table 1. The demographic information of the respondents

Variables	Levels	Count	(%)	Variables	Levels	Count	(%)
Gender	Male	104	46	Age	<18	0	0
	Female	122	54		18-22	42	18.6
Education	High school or below	53	23.5		23-27	119	52.7
	Bachelor degree	139	61.5		28-31	34	15
	Master's degree	31	13.7		32-36	25	11.1
	PhD or higher	3	1.3		>=37	6	2.7
Usage	Frequently each day	110	48.7	Usage time	Within half year	16	7.1
Frequency	more than ten times a day	24	10.6		0.5-1 year	25	11.1
	Several times a day	58	25.7		1-2 years	38	16.8
	Once a day	15	6.6		2-3 years	23	10.2
	Several times a week	10	4.4		More than 3 years	124	54.9
	Rarely use	9	4.0				

4. DATA ANALYSIS

4.1 Measurement model

The evaluation of the measurement model is mainly conducted by examining the validity of the convergent validity and the discriminant validity. Convergent validity reflects whether there is a high correlation between the measure items of an instrument, and is examined by composite reliability (CR), average variance extracted (AVE) and Cronbach's alpha. Values of AVE, CR and Cronbach's alpha for each construct should be at least 0.5, 0.7 and 0.7, indicating that the scales have good convergent validity and reliability^{[23][24]}. As shown in Table 2, CR values range between 0.736 and 0.916, and those for AVE are between 0.502 and 0.785, and all Cronbach's alpha values were greater than 0.70, thereby indicating valid measures.

Table 2. Construct reliability and convergent validity

	Cronbach's α	CR	AVE
Technology-family conflict (TFC)	0.712	0.736	0.502
Technology-work conflict (TWC)	0.887	0.888	0.725
Technology-personal conflict (TPC)	0.836	0.837	0.633
Excessive use (EU)	0.812	0.812	0.593
Stress (ST)	0.916	0.916	0.785
Regret (RE)	0.929	0.929	0.766
Continuance intention (CI)	0.740	0.752	0.512

Discriminant validity reflects whether two factors are statistically different. The criterion for discriminant validity is that the square root of the AVE value of each variable should be greater than the correlation coefficient of that variable with other variables. As shown in Table 3, the square root of the AVE value for each variable was significantly greater than its correlation with other variables. Thus, the discriminant validity is confirmed.

Table 3. Correlation Coefficient Matrix.

	TFC	TWC	TPC	EU	ST	RE	CI
TFC	0.708						
TWC	0.657	0.852					
TPC	0.602	0.702	0.796				
EU	0.649	0.725	0.726	0.770			
ST	0.549	0.691	0.750	0.619	0.886		
RE	0.478	0.534	0.535	0.737	0.456	0.875	
CI	-0.374	-0.446	-0.467	-0.498	-0.524	-0.541	0.715

4.2 Structural model

The results are presented in Fig 2, explaining 54.3%, 61.7% and 39.0% of the variance in stress, regret, and continuance intention, respectively. All hypotheses were supported except for the relationship between Technology-family conflict and stress ($\beta=0.039$, $t=0.475$), thus H1, H2, H3b, H3c and H4 are supported, and H3a is not supported. Also, as shown in Table 4, the overall model fit indices demonstrate that the model is acceptable.

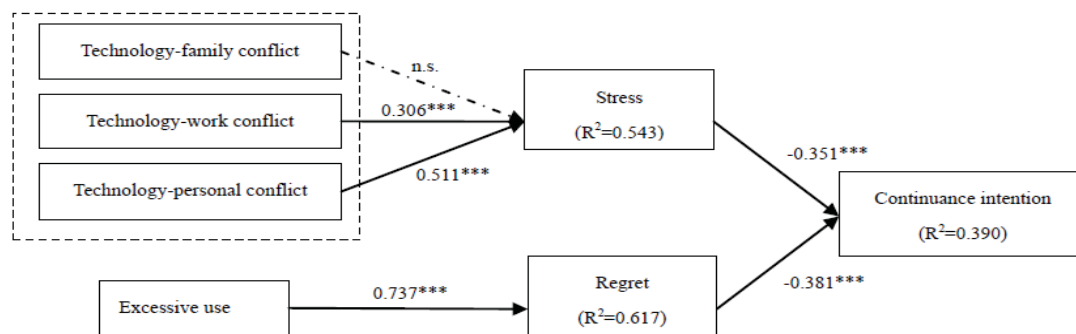


Fig 2. Structural model

Table 4. Fit indices for the estimated model.

	Chi-square /df	RMSEA	CFI	IFI
Critical value	< 3.0	< 0.08	> 0.90	> 0.90
Actual value	1.985	0.066	0.941	0.942

5. DISCUSSION, IMPLICATIONS AND LIMITATIONS

5.1 Discussion

This study aims to explore inhibitors of continuance intention to use mobile SNS. Results show that technology-personal conflict and technology-work conflict have significant impacts on stress. This finding is consistent with the previous study of Zheng and Lee [6]. Due to the use of mobile SNS, users may experience

physical problems, such as backaches, eye strain and headache. Such phenomenon may makes uses feel nervous and stress. Furthermore, the working stuff may use SNSs for their work needs. Untimely attention to the message may lead to loss of work interests; on the other side, excessive attention to information can also affect their work performance and concentration. However, the results show that technology-family conflict has no effect on stress. The possible reason may be that this study focuses on different samples, i.e. working stuff. Generally, working stuff has less time to spend with their families because of the place and time constraints. The existence of mobile SNS enables them to greet and care for their families whenever and wherever they want, and therefore has no significant relationship to stress.

As expected, excessive use of mobile SNS is found to be positively related to regret. Prior studies found that addiction to using the sites directly influences the feeling of regret ^[18]. When users find that their use of mobile SNS is longer than the time planned, they will produce regret feelings. They consider that time should be spent on more meaningful things, such as learning a new knowledge or new skill. Therefore, reasonable use can effectively reduce the sense of regret.

Among the inhibitors of mobile SNS continuance intention, stress is found negatively related to mobile SNS continuance intention. This finding is consistent with Beaudry ^[8]. Beaudry found that individuals avoid using SNS when they perceive SNS usage as a stressful experience ^[9]. That is to say, when user is stress from mobile SNS, they would shape their current situation to stop using the sites. As users adjust their stress in their lives, users rely more on mobile SNS. But the social stress from the convenience of mobile SNS itself adds further stress on users, which further increase the likelihood of users reducing their use of mobile SNS. What's more, the current research finds that regret negatively affects continuance intention. This is consistent with the conclusion of a previous research on SNSs activities, which indicated that if people feel regret, it is reasonable to expect that this negative emotion can be fixed by discontinuance actions ^[25]. Hence, when mobile SNS users realize that their behavior is wrong and feel regret from mobile SNS, they will temporarily or permanently discontinue using mobile SNS to reduce their regret.

5.2 Implications and limitations

This study has theoretical implications. First, prior studies were more concerned about the enablers of user continuance intention and considered less on the influences of personal feelings, such as regret. However, this study mainly focused on inhibitors of mobile user behaviors and explored effects of regret and stress. It will be helpful on better understanding user behavior in mobile contexts. Second, data were collected from the working staff, while prior studies mainly focused on samples such as university students. Findings of this study thus will enrich the related study of mobile user behaviors.

Furthermore, this study has practical implications. For users, mobile SNS is a double-edged sword and need to be used rationally. Users should handle their conflicts of technology with personal and work in an appropriate way. In order to relieve the stress of users, service providers are advised to set relaxing games in social networking sites. They can recommend new knowledge or skills based on the users' habits. In addition, when users overuse social networking sites, service providers should friendly remind users of paying attention to their health. For example, when users use mobile SNS continuously for more than two hours, he or she will be reminded to rest to protect his or her eyesight. To reduce users' regret, service providers are expected to encourage users to use SNSs reasonably.

This study is subject to several limitations. First, different user groups may behave differently. Further study may consider conduct difference analysis among e.g., working staff and student samples. Second, data were collected from Chinese users. Future research may consider the effects of cultural differences. Third, this study only focused on the impacts of stress and regret. Further study may consider other inhibitors such as

perceived overload.

APPENDIX: ITEMS AND SOURCES

Construct	Items	Sources
Technology-family conflict	TFC1 The use of mobile SNS keeps me from my family and friends more than I would like.	Turel ^[18]
	TFC2 The use of mobile SNS takes up time that I feel I should spend with my family and friends.	
	TFC3 (R) I generally seem to have enough time to use my mobile SNS and to spend time with family and friends.	
Technology-work conflict	TWC1 Mobile SNS usage influences my work.	Hong ^[19]
	TWC2 I neglect work to spend more time on mobile SNS usage.	
	TWC3 My work performance and concentration are influenced by mobile SNS usage.	
Technology-personal conflict	TPC1 I experience physical problems because of mobile SNS use (e.g., backaches, eye strain, and headache).	Zheng ^[6]
	TPC2 Using mobile SNS at night influences my sleep.	
	TPC3 I lose sleep due to late-night mobile SNS using.	
Stress	ST1 I feel drained from activities that require me to use mobile SNS.	Ayyagari ^[15]
	ST2 I feel tired from my mobile SNS activities.	
	ST3 I feel burned out from my mobile SNS activities.	
Excessive use	EU1 I think the amount of time I spend using mobile SNS is excessive.	Scott ^[17]
	EU2 I spend a usually large amount of time using mobile SNS.	
	EU3 I spend more time using mobile SNS than most other people.	
Regret	RE1 Excessive use of mobile SNS makes me feel guilty.	Turel ^[5]
	RE2 Excessive use of mobile SNS makes me feel ashamed.	
	RE3 Excessive use of mobile SNS makes me be angry at self.	
	RE4 Excessive use of mobile SNS makes me be dissatisfied with self.	
Continuance intention	CI1 My intentions are to continue using this mobile SNS than use any alternative means.	Bhattacharjee ^[20]
	CI2 (R) If I could, I would like to discontinue my use of this mobile SNS.	
	CI3 (R) If I could, I would like to use other alternatives rather than this mobile SNS.	

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71403301, the National Undergraduate Training Program for Innovation and Entrepreneurship of Sun Yat-sen University under Grant 201701003 and the Fundamental Research Funds for the Central Universities under Grant 16WKPY35.

REFERENCES

- [1] Tencent. (2017). <https://www.tencent.com/zh-cn/articles/8003451510737482.pdf>, 2017-12-30. (in Chinese)
- [2] Chaouali, W. (2016). Once a user, always a user: enablers and inhibitors of continuance intention of mobile social networking sites. *Telematics & Informatics*, 33(4), 1022-1033.
- [3] Luqman, A., Cao, X., Ali, A., Masood, A., & Yu, L. (2017). Do you get exhausted from too much socializing? empirical investigation of facebook discontinues usage intentions based on sor paradigm. *Computers in Human Behavior*.
- [4] Woollaston V (2013) Facebook users are committing ‘virtual identity suicide’ in droves and quitting the site over privacy and addiction fears. *Daily Mail*, London.

- [5] Turel, O. (2015). Quitting the use of a habituated hedonic information system: a theoretical model and empirical examination of Facebook users. *European Journal of Information Systems*, 24(4), 431-446.
- [6] Zheng, X. B., & Lee, M. K. O. (2016). Excessive use of mobile social networking sites: Negative consequences on individuals. *Computers in Human Behavior*, 65, 65-76.
- [7] Lavee, Y., McCubbin, H. I., & Olson, D. H. (1987). The effect of stressful life events and transitions on family functioning and well-being. *Journal of Marriage & Family*, 49(4), 857-873.
- [8] Lazarus, R. S. (1966). Psychological stress and the coping process. *Science*, 156.
- [9] Beaudry, A., & Pinsonneault, A. (2005). Understanding user responses to information technology: A coping model of user adaptation. *MIS Quarterly*, 29(3), 493-524.
- [10] Maier, C., Laumer, S., Weinert, C., & Weitzel, T. (2015). The effects of technostress and switching stress on discontinued use of social networking services: a study of facebook use. *Information Systems Journal*, 25(3), 275-308.
- [11] Ireland, J. L. (2003). Shame, guilt, and alcoholism: Treatment issues in clinical practice, 2nd edition. *Psychologist*, 16(8), 430-430.
- [12] Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: an interpersonal approach. *Psychological Bulletin*, 115(2), 243-267.
- [13] Turel, O. (2016). Untangling the complex role of guilt in rational decisions to discontinue the use of a hedonic information system. *European Journal of Information Systems*, 25(5), 1-16.
- [14] Tarafdar, M., Tu, Q., Ragu-Nathan, B. S., & Ragu-Nathan, T. S. (2007). The impact of technostress on role stress and productivity. *Journal of Management Information Systems*, 24(1), 301-328.
- [15] Ayyagari, R., Grover, V., & Purvis, R. (2011). Technostress: Technological Antecedents and Implications. *MIS Quarterly*, 35(4), 831-858.
- [16] Fox, J., & Moreland, J. J. (2015). The dark side of social networking sites: an exploration of the relational and psychological stressors associated with facebook use and affordances. *Computers in Human Behavior*, 45(45), 168-176.
- [17] Scott E. Caplan, & Andrew C. High. (2006). Beyond excessive use: the interaction between cognitive and behavioral symptoms of problematic internet use. *Communication Research Reports*, 23(4), 265-271.
- [18] Turel, O., Serenko, A., & Bontis, N. (2011). Family and work-related consequences of addiction to organizational pervasive technologies. *Information & Management*, 48(2-3), 88-95.
- [19] Hong, F. Y., Chiu, S. I., & Huang, D. H. (2012). A model of the relationship between psychological characteristics, mobile phone addiction and use of mobile phones by Taiwanese university female students. *Computers in Human Behavior*, 28(6), 2152-2159.
- [20] Bhattacharjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, 25(3), 351-370.
- [21] Van, d. V. F. J. R., & Leung, K. (1997). Methods and data analysis for cross-cultural research.
- [22] CNNIC. (2018). <http://www.cnnic.net.cn/hlwfzyj/hlwxbzg/hlwjtjbg/201803/P020180305409870339136.pdf>, 2018-03-24 (in Chinese)
- [23] Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.
- [24] Chin, W. W. (1998). The partial least squares approach for structural equation modeling. In Marcoulides G A, *Modern methods for business research*, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 295-336.
- [25] Klass, E. T. (1978). Psychological effects of immoral actions: the experimental evidence. *Psychological Bulletin*, 85(4), 756-771.

Product Information Diffusion in a Social Network and Marketing

Implications: A Case Study of Huawei Mobile Phone

Zhang Ling^{12*}, Luo Manman¹, Zhu Lijun²³

¹School of Management, Wuhan University of Science and Technology, China

²Center for Service Science and Engineering, Wuhan University of Science and Technology, China

³Institute of Scientific and Technical Information of China, Beijing, China

Abstract: There is a need to consider how to spread marketing information into the largest area in a social network. In this paper, Tweets of Huawei Mate 9 were collected to analyze users' information behavior such as tweeting, forwarding, and commenting on tweets. First, the network topology is described as topology structure diagram; second, the Independent Cascade Model (ICM) is used for simulating information propagation; and finally, the article discusses how to identify the influential nodes to maximize the spread of business marketing information. The findings show us how to choose the influential nodes in an enterprise's marketing campaign conducted in a social network. The result shows that enterprises must pay attention to official nodes but also accident nodes. We suggest that the enterprise should pay more attention to the individuals who are characterized by their occupation, interests and are influential in a friend circle or interests-oriented circles.

Keywords: Social Network; Information Propagation; Independent Cascade Model

1. INTRODUCTION

Human social relationships are bound by time and space. However, the evolution of information and communication technologies tools have allowed people to inexpensively and reliably share information anytime and anywhere through social media. Companies such as Philips, HP, and Microsoft have adopted seeding strategies that target influential nodes in social networks to launch new products^[1]. For example, Twitter, one of the most popular social media techniques, has evolved into a practical means for sharing opinions on almost all aspects of everyday life^[2, 3].

Twitter is a popular microblogging service through which users send and receive text-based posts, known as "tweets", consisting of up to 140 characters. In the process of using Twitter, the users' behavior such as forwarding, or commenting can promote the spread of information in a social network. Many researchers conducted their studies on Twitter because the retweeted times is an clear indicator to the diffusion process^[4-6]. Besides, today enterprises regard social networks as an important platform for launching new products and receiving the market feedback on a product. Therefore, it is significant to study the information diffusion in social networks to make the viral-marketing strategy successful. Specifically, how to identify the influential in an online marketing campaign, and how these influential are connected in a social network.

In this paper, the characteristics of the diffusion of an enterprise's product information in a social network were analyzed by using the Twitter data. First, the diffusion network topology was visualized and analyzed. Second, the independent cascade model was used to simulate the diffusion of enterprise mobile phone product information. Third, the influential nodes were identified, and the characteristics of influential nodes were investigated, marketing implications were discussed and concluded.

¹ Corresponding author. Zhangling, Email: beauty_ling@hotmail.com

2. LITERATURE REVIEW

2.1 Diffusion models

In recent years, the social network information propagation has become a research hotspot. The research involves network topology analysis^[7], text content analysis, large-scale data processing and so forth. At the same time, in order to express and predict the process of information diffusion in social networks, researchers have proposed information diffusion models^[8]. Some frequently used models are independent cascade model (ICM)^[9], linear threshold model (LTM)^[10] and epidemics model^[11]. Guille^[12] noted that the diffusion models could be divided into graph based models and non-graph based models according to the propagation rule that considers the interaction between nodes or not. This study aims to find out the influential nodes and explain why they are influential. Hence, we choose the graph-based models to simulate the diffusion process.

Meanwhile, some scholars have carried on the algorithm optimization research on those models. Because of the complexity of social networks, a more scientific approach is to apply the models in some typical network cases. Algerian author Samir Akrouf^[13] et al. analyzed the information propagation process and the influence of a set of nodes in two different networks: an egocentric contact network created by explicit relationships from Flickr social service, and an implicit video-commenting network created by commenting relationship from YouTube service. As a result, the research noticed that ICM performed better on implicit networks with stronger ties since it is based on the interactions between nodes. The article gives us new insights that it is useful to estimate the network attribute before choosing the information diffusion model.

2.2 Influential nodes identification

In the relative study about identifying influencers in a large-scale spreading, it has been accepted that the ability of influencers to initiate a large-scale spreading is attributed to their privileged locations in the underlying social networks^{[14] [15-18]}. The most straightforward measurement of influence is using centrality-based heuristics. In recent years, an increasing number of predictors have been adopted to ranking node's influence in a social network, among which the most universally used ones include degree centrality^[19], betweenness centrality^[20], k-core^[21] and PageRank^[22]. Specifically, in the study of information diffusion, some scholars choose influential nodes as initial active nodes and simulate the information propagation process. Li^[23] proposed a descriptive diffusion model to take dependencies among the topics into account, to identify the most influential nodes for specific contagion, and they applied the proposed model on an ISIS Twitter dataset, aiming to predicting the diffusion volume. Lu, et al^[24] proposed a Score Cumulate model to evaluate the initial influence by using PageRank, and they applied the model in two real-world networks, Facebook and e-print arXiv (a scientific co-author network). Kwak^[18] identified influentials on Twitter, and ranked users by the numbers of followers, PageRank and betweenness centrality. Thus, it is proper to identify influential nodes in a social network by the centrality measurements, through a simulation experiment.

3. METHODOLOGY

The diffusion network was constructed through data import and processing procedure, then influentials were identified from performing the simulation experiment. Finally, influential node's background were investigated, as well as tweet content, etc. The methodology could be shown as below:

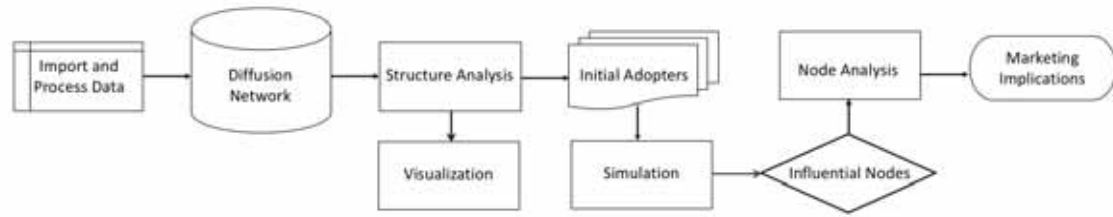


Figure 1. The overview of the methodology

3.1 Data import and processing

There are many social network analysis tools for extracting, analyzing and visualizing social network data. We use NodeXL (a free Excel plug-in) ^[25] developed by the Social Network Foundation. NodeXL is user-friendly for people who are not able to program. Our tweet data were sourced from Twitter. “Huawei Mate 9” was used as the keyword, searched daily through the node XL from January 3, 2017 to January 13, 2017. Then tweets data that is inaccessible to tweet relationship type were removed. Finally, the extracted diffusion network contains 5791 unique node (users) and 8386 links (relationship) between them. The relationship between vertices (users) include the original tweets (Tweet), comments (Replies to), and mentions (Mention).

3.2 Network structure analysis

After the Twitter diffusion network was constructed, the structure was further analyzed both visually and quantitatively. On one hand, network structure was visualized using a cluster-layout algorithm provided by NodeXL. On other hand, each node’s network metric was calculated, including the betweenness centrality and the PageRank. The nodes were ranked, and the top-ranked nodes in each measurement were selected as the corresponding heuristic initial adopters. The selected initial adopters are applied to the regarding heuristic to simulate information propagation according to independent cascade model.

3.3 Simulation

Independent Cascade Model was proposed in the context of marketing by Goldenberg, Libai, and Muller ^[9]. Given a network $G = (V, E)$ where V is the set of vertices, and E is the set of existing edges in the network. A vertex $v \in V$ is said to be **active** if the information has reached the vertex and was accepted by it. If the information didn’t reach the vertex or the vertex rejected it, then the vertex is said to be **inactive**. Each inactive vertex tends to become active, and it can switch from inactive to active, but it cannot switch from active to inactive. Given a set of initial active vertices A_0 , vertex v first becomes active in step s and is given a chance to activate each of vertex v ’s inactive neighbors with a probability P_{vw} for success. If v succeeds to activate one of its inactive neighbors, say w , in step $s+1$, then the new active vertex w will be added to A_s to form the new active vertices set A_{s+1} , and w will adopt the same activation action to activate its inactive neighbors. Each vertex is only given one chance to activate its neighboring vertices. If v fails to activate w in step $s+1$, it cannot make further attempts to activate w in subsequent steps. The propagation process ends when there are no more vertices can become active in step s . In our study, the probability of succeeding to activate vertices was set with 0.5.

Selecting an appropriate set of initial active nodes is a key step to simulating information propagation. On one hand, the initial set of active nodes were selected based on high betweenness centrality. On the other hand, high PageRank was considered as supplemental measurement since it is a common measure to gauge the importance of a node. Hence, the metrics of *betweenness centrality* and *PageRank* were adopted to select top-rank node sequence as initial adopters to initiate the diffusion process. If the total number of activated nodes larger, the initial adopters under this measurement is more influential.

4. FINDINGS

4.1 Network structure and visualization

The extracted “Huawei Mate 9” diffusion network for vertices (nodes) contained 5,791 individuals. While the edges which are ties or connections between nodes contained 8,686 connections(edges) Figure 2 is the extracted diffusion network graph which showing the interconnection of users, including retweet, comment relationship. A vertex is created when a user posted an original tweet. An edge is created when user(s) respond to the original tweet: retweet or comment.

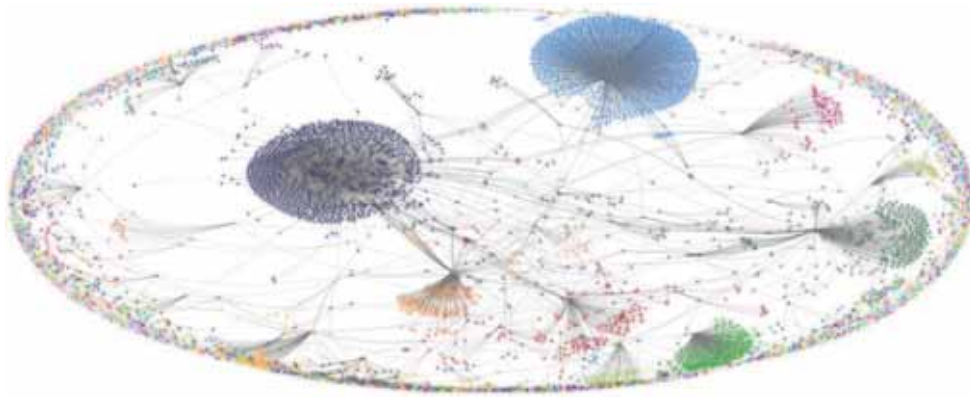


Figure 2. A Network Graph of #Huawei Mate 9

Next, network basic metrics, node's centrality and PageRank have been calculated using NodeXL, as shown in Table 1. These metrics help us characterize the research network as shown in Table 2. At the same time, we ranked the nodes according to the *betweenness centrality* and *PageRank*.

Table 1. Structure Characteristics of the User Networks

Graph Type	Directed Network
Type of relationship	Implicit
Number of vertices	5 791
Number of edges	8 386
Graph density	0.00019
Connected component	1 887
Maximum number of vertices in connected component	3 270
Maximum number of edges connected component	7 496
Diameter	13
Average distance	4. 351 803

Table 2. Metrics statistics

Metric Statistics	Value
Minimum overall degree	1
Maximum overall degree	1 032
Average overall degree	2.889
Minimum out-degree	0
Maximum out-degree	25
Average out-degree	1.448
Minimum in-degree	0
Maximum in-degree	1 031
Average in-degree	1.448
Minimum betweenness	0
Maximum betweenness	5027993.179
Average betweenness	6 197.929

4.2 Simulation results

The diffusion process algorithm for ICM was carried out on this directed network for each heuristic $h \in H$, where $H = \{PageRank, betweenness\}$. For each heuristic the diffusion process was run 10 times for a differing number of initial active nodes. There were four sets of initial active nodes consisting of 5, 10, 15, or 20 nodes, respectively $A_0 \in \{5, 10, 15, 20\}$. An average was computed after each run of ten using one of the four heuristics and in of the four sets of initial active nodes. This is the average of active nodes after each run was computed. This average is the influence of the initial set. These results are displayed in Figure 3 which shows the performance of the algorithms used in the ICM on the network.

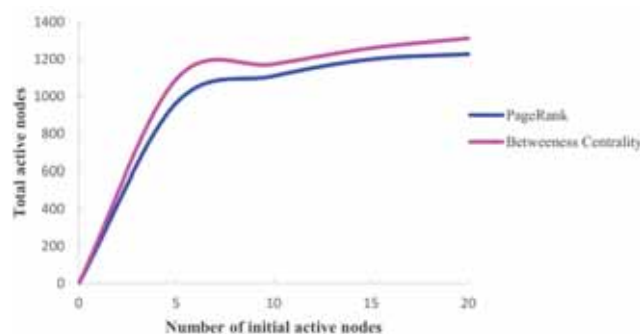
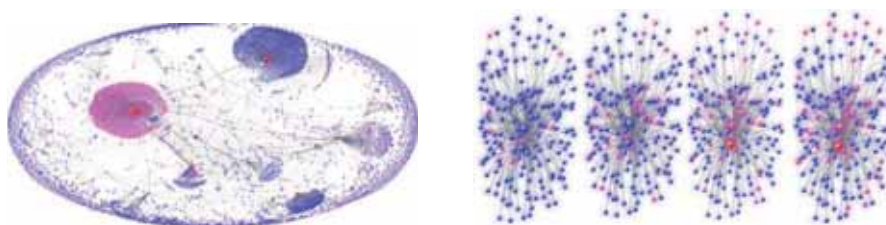
**Figure 3. Simulation results for Independent cascade model**

Figure 4. The left is an instance after a diffusion run using the ICM. The right is the activated status of a branch group node “androidheadline” change with Different number of initial active nodes. Noted that red, fuchsia, blue respectively represents nodes of *initial active*, *activated*, *inactive*.

1) It can be seen from Figure 3 the results converge to be proximate and close when the initial active set is chosen based on PageRank and betweenness centrality. The initial active nodes that have high overall degree

activate the largest number nodes: 22.68% of the nodes. The activation effect of high PageRank is not as good as the high betweenness centrality nodes.

2) From Figure 3, the curves become steady when the initial active node sets contain more than seven (7) nodes. The first seven (7) targeted nodes (with high PageRank, betweenness centrality) influence a large fraction of the network. When the initial active nodes are less than seven (7), the activation effect of choosing nodes with high betweenness centrality values seems the same as choosing high PageRank values. When the initial active nodes are larger than seven (7), the activation effect of the node with high betweenness centrality is relatively weakened.

3) Figure 4 visualizes the diffusion process using the ICM (independent cascade model). NodeXL enables us to visualize the nodes based on their metrics. The left part illustrates an instance of diffusion process choosing 10 initial active nodes (in red) based on the high betweenness centrality. By the end of the process, the initial adopters have influenced about 45% of the nodes (infected nodes are in fuchsia). The right part shows the node “*androidheadline*” has not been activated when there are 5 or 10 initials activation nodes. While the number of initial nodes achieved 15 or 20, although node “*androidheadline*” has become the activated node, there is a little diffusion change in the branch group.

4.3 Influential nodes identification

In our study, total number of activated nodes was defined as the influential. Larger the total number of activated nodes, higher the influential based on the regarding initial adopter strategy. Based on the experimental results represented in Figure 3. the heuristic for the diffusion process based on: **PageRank**, **betweenness centrality** converge to be proximate. Top 20 nodes in PageRank, betweenness centrality, and obtained 27 non-repeat nodes were merged. The influential diffusion nodes set $S = \{huaweimobile, huaweimobileuk, androidauth, threeluk, youtube, nobunaga_s, huawei, androidheadline, jet, huawei_japan_pr, droid_life, huaweimobileksa, hamadsalleeh, androidcentral, wsj, xataka, techzilla, princepipa, majuzub, khajochi, huaweimobileesp, huaweimobilemy, this_is_e, freeconteston, metrini, mobilenewsmag, rkii2306\}$. The nodes into were divided into four categories "Huawei official", "media", "mobile review", and "ordinary user".

5. DISCUSSIONS

The simulation result is consistent with the Akrouf's^[13] experiment results, which also applied the independent cascade model to an implicit network. It indicates that the propagation rule of independent cascade model is proper to fit the diffusion network in marketing domain.

In this study, 27 influential nodes were identified, and classified the nodes into four categories according to their relationships with Huawei's new product “Huawei Mate 9”, namely, “official”, “media”, “mobile review”, and “ordinary user”. Not all influential diffusion nodes are Huawei's official platform, information media, digital products reviewers, some of them are ordinary users and their influence we didn't expect. They also promote the information diffusion in our case. Thus, influential nodes unexpected arise in the event were called as “*accident node*”. In this paper, “*accident node*” are $\{nobunaga_s, princepipa, majuzub, metrini\}$.

The average rank of Node “*nobunaga_s*” is 6. It is account of a Japanese actor, who published one tweets on January 5. The tweet's content is “I heard that Huawei Mate 9's performance is very well, playing games without crashing, I would like to chose Huawei Mate 9 to be my next mobile phone.” This tweet immediately caused more than 200 forwarding and nearly 1900 “like” (thumbs up). The node “*majuzub*” has high betweenness centrality rank, it is a Japanese religion scholar, who had released a tweet on January 7. The tweet says, “I am very interested in the new Huawei Mate9 which equipped with a Leica camera”, and this tweet was responded by Huawei's Japanese official “*huawei_japan_pr*”. The node is interested in the Huawei's mobile

and pay attention to the mobile purchase information constantly. The tweet content of “*nobunaga_s*” and “*majuzub*” is shown in Figure 5. From this case, it is worth for Huawei market staff to further consider the marketing opportunities of the user’s identity of scholar. Moreover, the official’s response behavior may draw user’s attention on mobile product and lead to users’ purchase behavior eventually.



Figure 5. “accident node” tweets content

From the view of average ranking of node “*rkii2306*”, it just an ordinary user, but its betweenness centrality ranking is very high. Betweenness centrality essentially reveals how important each node is in providing a “bridge” between different parts of the network^[14]. The higher the betweenness centrality, the stronger the intermediary effect of the node, and the stronger dependence of other nodes. The node’s identity is Japanese android lovers, who have strong interest in the new mobile phones on the market. The reason of its high betweenness centrality is because that it locates in the two core node branches group. One branch is based on Huawei official node “*huawei_japan_pr*” as core node, another branch is based on accident node “*nobunaga_s*” as core node. The local network structure where node “*Rkii2306*” is shown in Figure 6.

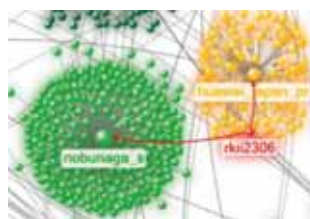


Figure 6. Node “*rkii2306*” in the Diffusion Network

6. CONCLUSIONS

This article used the NodeXL to visualize a product information diffusion network. The network topology represents the users’ tweeting, forwarding, commenting relationships. Meanwhile, the article used independent cascade model (ICM) to simulate the information diffusion process. To identify the influential nodes, betweenness centrality and PageRank were considered as the measurements. The simulation results show that in the initial stage of the information, Huawei’s official and media have an ability to trigger a large cascade. Then in the middle stage, some hub nodes such as mobile review, ordinary users emerge, leading to small-scale subsequent information dissemination. It is worth to explore the influence of those “accident node”. In our case, some accident nodes are scholars, game lovers and so on. They can guide target customer groups, promote enterprise marketing effectively; accordingly promote a large scale of information dissemination. Based on the discussions, we suggest that the enterprise should pay more attention to the individuals who are characterized by their occupation, interests and are influential in a friend circle or hobby-oriented circles.

Our contribution is that independent cascade model was applied to a marketing empirical diffusion network, and found it is proper to fit the product information diffusion process. We also found it significant to consider “accident node”. And the tweet content related to novel technology could attract more participation in ordinary users. However, the limitation exists. First, the diffusion model is basic and simple, there is much room for improvement. For example, the probability of succeeding could be adjusted according to the node’s attributes

and local network properties. Second, we didn't consider the culture influence on the diffusion process. In this case, the accident nodes are from Japan. It is worthwhile to further comprehend the nationality to improve the diffusion model, or to propose a more comprehensive measurement of the influentials.

ACKNOWLEDGEMENT

Research was sponsored in part by the National Social Science Fund Project "Study on dynamic optimization mechanism of information diffusion in social networks", Agreement Number 15CTQ029.

REFERENCES

- [1] Libai, B., Muller, E., & research, R. P. J. O. M. (2013). Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. *journals.ama.org*, 50(2), 161–176.
- [2] Richardson, M., & Domingos, P. (2002). Mining knowledge-sharing sites for viral marketing (pp. 61–70). Presented at the the eighth ACM SIGKDD international conference, New York, New York, USA: ACM Press.
- [3] Kaplan, A. M., & Haenlein, M. (2012). Social media: back to the roots and back to the future. *Journal of Systems and Information Technology*, 14(2), 101–104.
- [4] Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems*, 29(4), 217–248.
- [5] Ferrara, E., & Yang, Z. (2015). Measuring Emotional Contagion in Social Media. *PLOS ONE*, 10(11), e0142390–14.
- [6] Tsugawa, S., & Ohsaki, H. (2017). On the relation between message sentiment and its virality on social media, 1–14.
- [7] Udanor, C., Aneke, S., & Ogbuokiri, B. O. (2016). Determining social media impact on the politics of developing countries using social network analytics. *Program*, 50(4), 481–507.
- [8] Even-Dar, E., & Shapira, A. (2007). A Note on Maximizing the Spread of Influence in Social Networks. In *Internet and Network Economics* (Vol. 4858, pp. 281–286). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [9] Goldenberg, J., Libai, B., & Muller, E. (2001). Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters*, 12(3), 211–223.
- [10] Granovetter, M. (1978). Threshold Models of Collective Behavior. *American Journal of Sociology*, 83(6), 1420–1443.
- [11] Hethcote, H. W. (2000). The Mathematics of Infectious Diseases. *SIAM Review*, 42(4), 599–653.
- [12] Guille, A., Hacid, H., Favre, C., & Zighed, D. A. (2013). Information diffusion in online social networks: a survey. *ACM SIGMOD Record*, 42(2), 17–28.
- [13] Akrouf, S., Meriem, L., Yahia, B., & Eddine, M. N. (2013). Social Network Analysis and Information Propagation: A Case Study Using Flickr and YouTube Networks. *International Journal of Future Computer and Communication*, 246–252.
- [14] Watts, D. J., & Dodds, P. S. (2007). Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, 34(4), 441–458.
- [15] Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics*, 6(11), 888–893.
- [16] Pei, S., & Makse, H. A. (2013). Spreading dynamics in complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(12), P12002.
- [17] Yoganarasimhan, H. (2011). Impact of social network structure on content propagation: A study using YouTube data. *Quantitative Marketing and Economics*, 10(1), 111–150.
- [18] Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? (p. 591). Presented at the the 19th international conference, New York, New York, USA: ACM Press.
- [19] Albert, R., Jeong, H., & Barabási, A.-L. (1999). Internet: Diameter of the World-Wide Web. *Nature*, 401(6), 130–131.

- [20] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215–239.
- [21] Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks*, 5(3), 269–287.
- [22] Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7), 107–117.
- [23] Li, Q., Kailkhura, B., Thiagarajan, J. J., Zhang, Z., & Varshney, P. K. (2016, November). Influential Node Detection in Implicit Social Networks using Multi-task Gaussian Copula Models. *arXiv.org*.
- [24] Lu, F., Zhang, W., Shao, L., Jiang, X., Xu, P., & Jin, H. (2016). Scalable Influence Maximization under Independent Cascade Model. *Journal of Network and Computer Applications*, 1–26.
- [25] M, S., A, C., N, M.-F., B, S., E, M. R., J, L., & C, D. (2010). NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010/2013/2016. <https://www.smrfoundation.org>. Retrieved March 19, 2018.

The Analyzing on the Connotation, Properties and Research Trend of Enterprise Social Media

Yu Wang¹, Daqing Zheng^{2}, Lin Huang³*

^{1,2,3}School of Information Management & Engineering, Shanghai University of Finance and Economics, Shanghai, 200433, China

Abstract: The expansiveness and importance of Enterprise Social Media (ESM) have attracted the attention of researchers. ESM's connotation, properties and functions are being actively explored. And the discovery of some new social media properties such as diversity, polycentricity and network externalities, as well as their further application in business management scenarios, will surely drive further ESM research. Taking this reality into account, we will redefine the connotation of ESM, sort out the properties of ESM, analyze the overall characteristics of ESM related literatures, summarize the current status of ESM research, and then look forward to the ESM future research blueprint. Our paper will help scholars and business managers understand that, the ESM has a profound impact and the application of ESM is inevitable, which will bring opportunities and challenges in the IS field.

Keywords: Enterprise Social Media, Enterprise Social Network, Social Media

1. INTRODUCTION

Social media is an Internet-based program that promotes users to create and share content. It's one of the most important applications in the past decades. ESM is a further extension of social media in the business environment. Today, more and more enterprises begin to use ESM for external communication and internal collaboration. ESM thus become one of the most important enterprise-level applications, and there are many mature ESM products, including foreign Yammer, Jive and domestic MingDao and so on. According to the research firm, the market size of ESM based enterprise collaboration software in 2019 will reach to 8.14 billion dollars.

ESM is fundamentally changing the way companies communicate, collaborate, operate and create values ^[1]. ESM, as an important part of enterprise application system for web 2.0, can generate the collective wisdom, promote collaboration and innovation. It can also help an enterprise to better establish connection based on knowledge and information among employees and further establish enterprise social knowledge network to achieve knowledge management. In addition, ESM can promote the connection among employees, then affect the social capital of individuals in the organization and further influence the performance of employees. For managers, ESM is provided to respond promptly to user feedback and maintain corporate image, ultimately impact enterprise performance.

According to the existing ESM research and considering the new properties and focuses to be urgently excavated and clarified, the logical structure of this paper is as follows. First, we summarize the classification, connotation and main properties of ESM; secondly, we collect the key literatures of ESM; thirdly, we measure and analyze the literatures, summarize the research status and explore the breakthrough of forthcoming research; finally, we give the research trend of ESM.

2. ESM CLASSIFICATIONS, CONNOTATION AND PROPERTIES

* Corresponding author. Email: zhengdaqing@sina.com

2.1 ESM classifications

Social media into the business environment, has undergone a gradual development process. With Facebook and other social media sites gathering more and more users, managers realize that mainstream social media sites are a good place to engage with customers and can communicate with customers in a convenient and efficient manner, which prompted many enterprises to start using social media. In view of the possible negative impact of using public social media, including the intrusion of employees' leisure and entertainment into the workplace and the leakage of internal information, the need to build an enterprise-owned social media is highlighted, such ESM can support internal staff contact and collaboration, but also to avoid the disadvantages of public social media^[2]. Therefore, the current ESM is mainly divided into two types: Enterprise external public social media (EEPSM) and enterprise internal private social media (EIPSM). EEPSM is an enterprise level application system based on public social media such as Facebook, Twitter and Blog. It's used to support the communication between enterprises and external organizations and individuals such as customers, suppliers and the public. EIPSM mimics mainstream social media platforms in terms of appearance, functionality and user experience, primarily for internal communication and collaboration. In addition, MITRE has developed an internal social media platform and allows employees to invite external business partners to join, which may mean that ESM will be an enterprise level application with seamless integration both internally and externally.

2.2 ESM connotation

As for the connotation of ESM, Kuegler et al. argue that ESM is to apply mature and public social media tools into enterprises so that employees can use it to work and self-management^[3], highlighting the characteristics of the social network. Leonardi et al. defined the connotation of ESM by summarizing the properties of ESM. They considered visibility, persistence, editability and association^[4], emphasizing ESM as a social media whose information and data can be edited and persisted for a long time, this view is similar to Wu^[5]. Richter et al. propose that the phenomenon of enterprises using public social media is ESM, thus implicitly indicating a typical scenario with two kinds of ESM applications^[6]; In addition, there were some scholars also define ESM from the perspective of resource management and knowledge management^[7-9]. Taking a variety of viewpoints, we have Table 1.

Table 1. ESM connotation analysis

	Connotation of ESM	Construction perspective	References
1	Achieve efficient, transparent and convenient communication and collaboration among internal employees	Enterprise internal social network	[3]
2	Web-based software platform with visibility, persistence, editability and association	Enterprise internal social network, social media	[4, 5]
3	Social network is applied to enterprise environment	Public social network, enterprise internal social network	[6]
4	Through the 'following', users form an enterprise internal network platform to provide continuous data resources to the shared pool	Data resource management, knowledge management	[7, 8]
5	Symbolic social interaction system to promote members' social identity and collaboration, especially knowledge sharing	Knowledge management, social network	[9]

Based on the analysis of existing research on the connotation of ESM, we think ESM is an enterprise-led social media network platform, that has the dual properties of social network and social media. Internally, it enables employees to communicate more conveniently and fosters collaboration; externally, it facilitates the communication with the external public, customers, and suppliers, ultimately creating value for the enterprise.

2.3 ESM properties

ESM give full play to the properties of social networks and social media. Specifically, from the perspective of social networks, ESM has the visibility, association, centrality and network externalities, from the perspective of social media, ESM is editable, persistent and diverse.

2.3.1 Visibility / Transparency

Visibility, refers to the act of communicating between users using ESM in a transparent state that can be observed. This transparency is divided into three dimensions, including the transparency of communicators identity, communication content and interaction, which correspond to the visibility of the identities, the contents and information exchange activities respectively^[10]. Visibility can be described both internally and externally. Internally, the visibility of the ESM can enhance communication and collaboration among employees, and managers can better understand and manage the employee to assist in business decisions. Externally, enterprises and consumers will communicate more timely and effective, which can be monitored and evaluated by managers, also to strengthen investor judgment on the business.

2.3.2 Persistence

Persistence refers to the fact that users can still communicate in their original mode once the communication is completed on the ESM platform, and because the exchanged information can be preserved for a long time, the others without having to engage in initial communication have the opportunity to learn and discuss. Treem and other scholars believe that persistence affects organizational behavior in three ways, including the availability of content, the "robustness" of communication formats, and the expansion and extensibility of content, which is an important foundation for enterprises to carry out knowledge management^[11].

2.3.3 Editability / Interoperability

Editability refers to ESM users can modify published content, sum up experience, create new knowledge and promote business innovation, there are two meanings, users can develop communication behavior and content, and modify the content released by themselves or others^[12]. The editability of ESM makes information richer and clearer, facilitates the dissemination of information, regulates personal expression, controls the extent and ways in which recipients can access content, and improves the quality of information through information exchange.

2.3.4 Association / Social connectedness

Association, also known as social connectedness, refers to social media can create and maintain the contact between multiple subjects^[13]. Generally, there are two types of connection in ESM. One is the connection between people reflects the properties of the social network. The other is the connection between people and information, which reflects the properties of social media. Association can generally be analyzed from the breadth and depth of contact, and then the social capital in the organization can be portrayed from the perspective of quantity and quality to explore the laws that affect the development of social capital in the organization.

2.3.5 Diversity / Heterogeneity

Diversity describes the differences in the function, the form and the network structure on ESM^[5], also known as heterogeneity. Based on the structure-hole theory, these holes in the social network that have low cohesion and can cross local networks. The diversity of nodes existing in these holes makes it easier to acquire professional knowledge and positively affects job performance. Meanwhile, the relationship between knowledge sharing and performance is more closely related to the diversity of participants in social media, or the structural diversity of the job team.

2.3.6 Decentration / Polycentricity

Centrality is an important indicator of social network analysis, describing "the extent to which one actor is at the network center". Based on social network theory, individuals who occupy a middle ground in social

networks are more likely to achieve creative results, in turn, gain a more favorable position, or even by influencing and shaping others in the network based on their own abilities and interests^[14]. What needs to be emphasized is that in the ESM application scenario, the social networks may show more decentration, or there may be some inconsistencies in network concentricity and departmental organizational structure. This is one of the important directions for ESM future research.

2.3.7 Network externality

Network externality involves the concept of "critical mass," which describes social processes that affect the behavior of individuals. When the value of participating networks increases as more people participate in the network, network externality will appear^[15]. Network externality theory shows that if more people use the same technology, then the technical value of the user will increase, so the continued use of employees is crucial to the success of the ESM..

The above properties of ESM are generally summarized from the perspectives of social media and social network, some of them have been fully recognized, such as visibility and persistence, while polycentricity needs further exploration. Future exploration of these properties will better reveal the value and impact of ESM.

3. ESM RELATED LITERATURES COLLECTION AND BASIC ANALYSIS

3.1 ESM related literatures collection

Our initial idea is to delineate 20 important journals in the IS field and select the "web of science" database for literatures collection. Through the early accumulation, we limited the time scope after 2004 and retrieved the keywords "enterprise social media", "enterprise social network" and "enterprise social software", respectively, and obtained 24 related literatures, which is difficult to reflect the overall situation of ESM research. Therefore, we cross-search using "social media", "social network", "social software" and "performance", "employee", "innovation", "efficiency" to obtain a total of 272 literatures including 24 literatures. Then we have a quick review of the abstract and the full text. We identified 31 literatures related to ESM. Taking into account the limitations of the keywords and the selected journals, we examined 31 literatures and found 55 important literatures in different journals. So far, we have 86 literatures. Finally, we started to read carefully, excluding literatures that were atypical social media studies or were not discussed in the enterprise environment, and confirmed the fuzzy boundaries. Finally, 56 literatures were identified that are closely related to the ESM. At the same time, considering the subjectivity of the manual screening, we divided 56 literatures into 2 copies and invited 4 professionals conducted double coverage readings and compared their screening results to reduce misjudgment, for the controversial literatures, we discuss together, eventually formed a consistent analysis of the results.

3.2 ESM related literatures basic analysis

Through the above methods, we confirmed 56 important ESM literatures. From the time dimension, the ESM research shows a clear growth trend from 2004 to 2015. With the deepening of research questions, ESM research faces some new problems and difficulties, so the number of research papers in the past two years has been reduced. In the following, we will analyze the literatures in detail from the four perspectives of theoretical perspective, research level, research topics and research methods, and elaborate the research status of ESM.

3.2.1 Theoretical perspective analysis

First of all, we analyze the theoretical perspectives of ESM research, in which social network theory is mentioned eight times, with the object being the relationship between people and society^[5]; second, the social capital theory, which mainly discusses the members' interaction influence individuals and teams^[16], and social

capital theory is an interdisciplinary concept, often appears with social space theory or interactive memory theory. In addition, there are cognitive theory based on psychology and study the change caused by experience, the communication theory used for communication visibility, and the media richness theory, computer-mediated communication theory and behavior organization theory. These theories cover various academic fields such as social network science, psychology, computer science and organizational behavior.

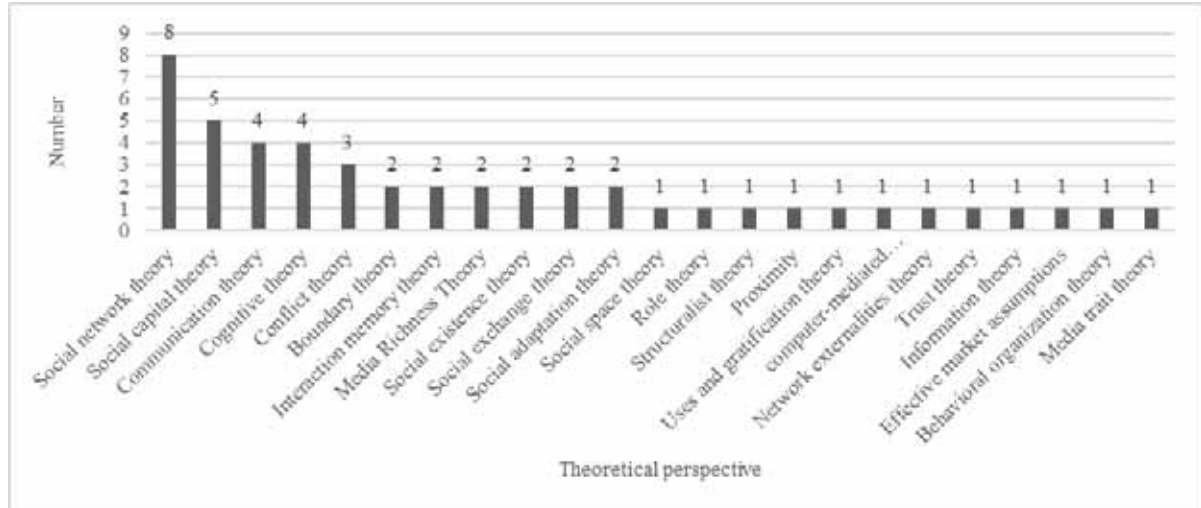


Figure 1. The theoretical perspective analysis of ESM related literatures

The collation of published journals confirms the interdisciplinary nature of ESM research. The final 23 related journals, including the top three journals in the field of IS, as well as journals in the fields related to computer science and behavioral science, we can say that ESM research not only draws attention in the field of IS, but also is considered by other fields as an interesting research topic.

3.2.2 Research level analysis

Ali-Hassan suggested that, ESM is a basic technology that penetrates into the organization through the use of employees, therefore, the research on ESM tends to focus on the individual level ^[2]. Excluding 10 literatures reviews, we classify the remaining 46 literatures on the three levels of individual, group and enterprise, and divide the multi-level research. Not surprisingly, there are a large number of studies at the individual level and relatively few literatures studies at the group level, often because even the analysis within the group focuses on the group members rather than the group structure. At the enterprise level, in 2006, there appeared an

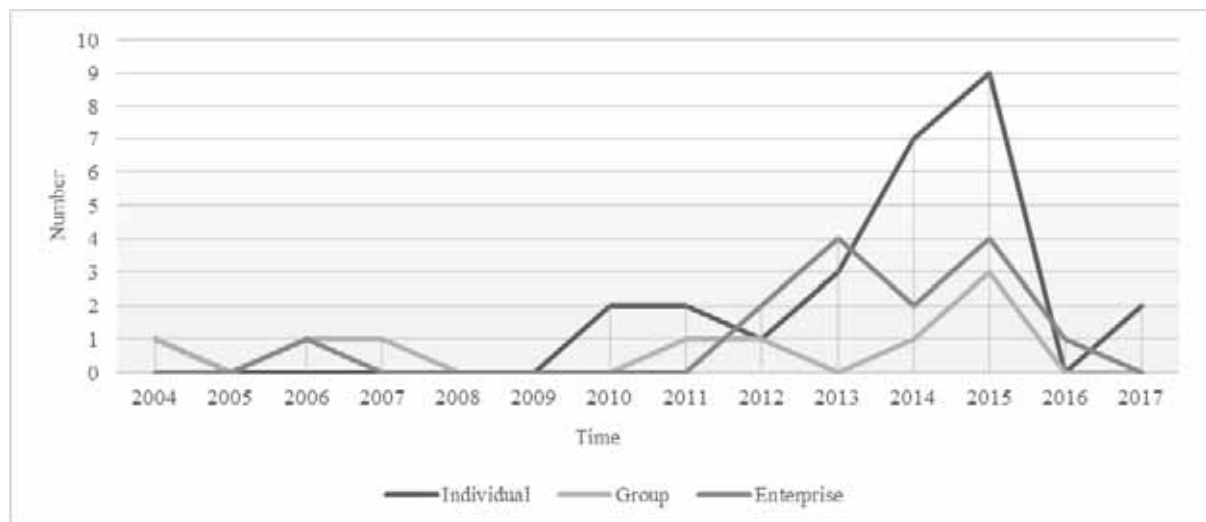


Figure 2. The research level analysis of ESM related literatures

exploration of the management strategies and values of corporate blogs, since 2012, research on enterprise performance, enterprise innovation and enterprise ESM management has been gradually increasing. This confirms Ali-Hassan's view, which means with the depth of individual research, there will be more and more research at the group and enterprise level. Finally, ESM research will give priority to the value of the enterprise and even be targeted at changing the organizational structure.

3.3.3 Research topic analysis

We continue to analyze 46 literatures that include topics such as ESM management, enterprise innovation, relationship management, enterprise performance, group performance, employee performance, employee knowledge management and sharing, and more. After multi-topic repeated counting, there are two main reasons can explain the majority of the topic-- employee knowledge management and sharing. First, the properties of ESM such as visibility, persistence and editability can promote employee knowledge interaction. Second, ESM has the function of enterprise forums and communities, and researchers can easily think about knowledge management issues from a virtual team perspective. The second topic is research on employee performance. According to Sarker, organizations are increasingly focusing on "personal performance" even in a collaborative environment, which leads to an increasing exploration of individual performance^[17]. It can be expected that in the short term, the correlation will still be based on individual knowledge management and performance analysis. However, in the long run, due to the fact that the ESM is ultimately responsible for organizational value, research on enterprise performance and reputation is more likely to be accepted by society.

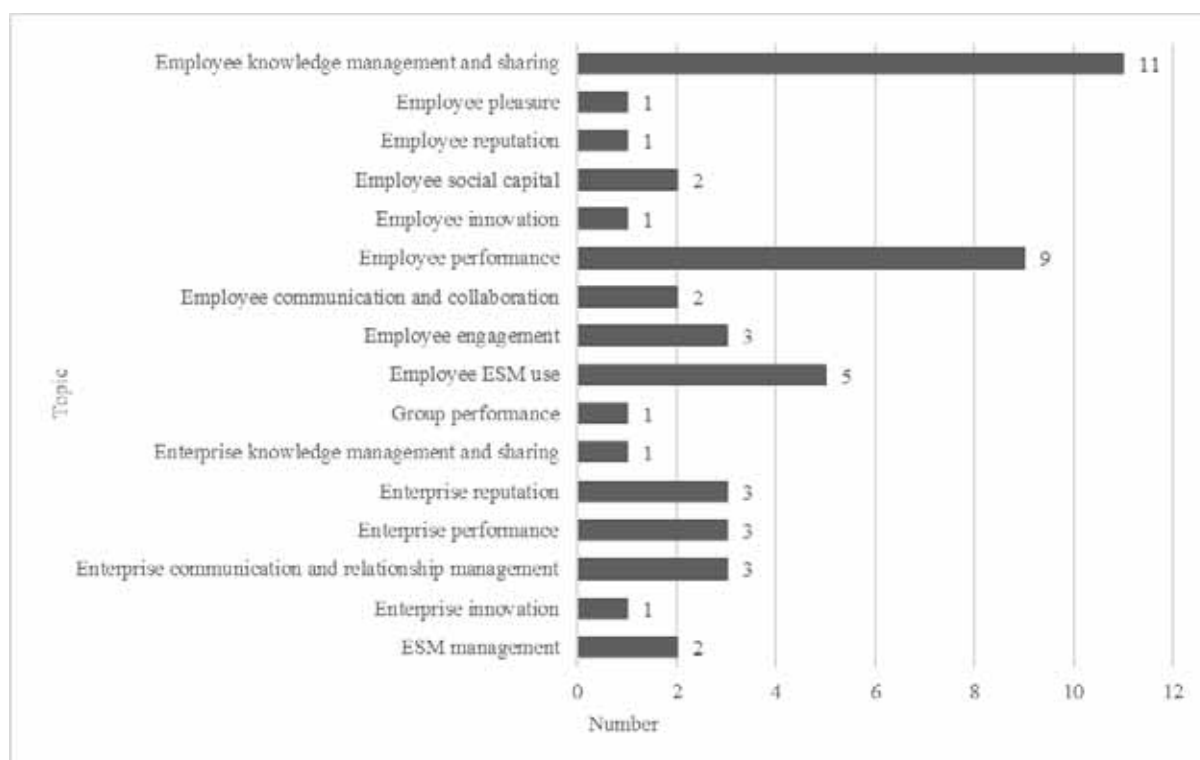


Figure 3. The research topic analysis of ESM related literatures

3.2.4 Research methods analysis

Based on previous research, we think that these 56 literatures mainly use seven kinds of methods, which are case study, secondary data analysis, laboratory experiment, field experiment, questionnaire, descriptive, concepts and conceptual applications. Among them, the most used method is the case study, accounting for 30%. As early as 1989, Straub et al. proposed that the case study is an exploratory study and is mostly used in IS early analysis, and concept and conceptual application categories(36%) are considered as effective means of theory

construction in the initial exploration^[18]. These exploratory methods are the mainstay in the short term, prove ESM research is an emerging field. In addition, scholars' attempts to analyze ESM using secondary data and experimental methods have also been ongoing. At present, we argue that exploratory qualitative research has attracted enough attention. In the near future, non-exploratory quantitative research including secondary data analysis and experiments will receive attention, and the result will play a role of an amendment to qualitative study.

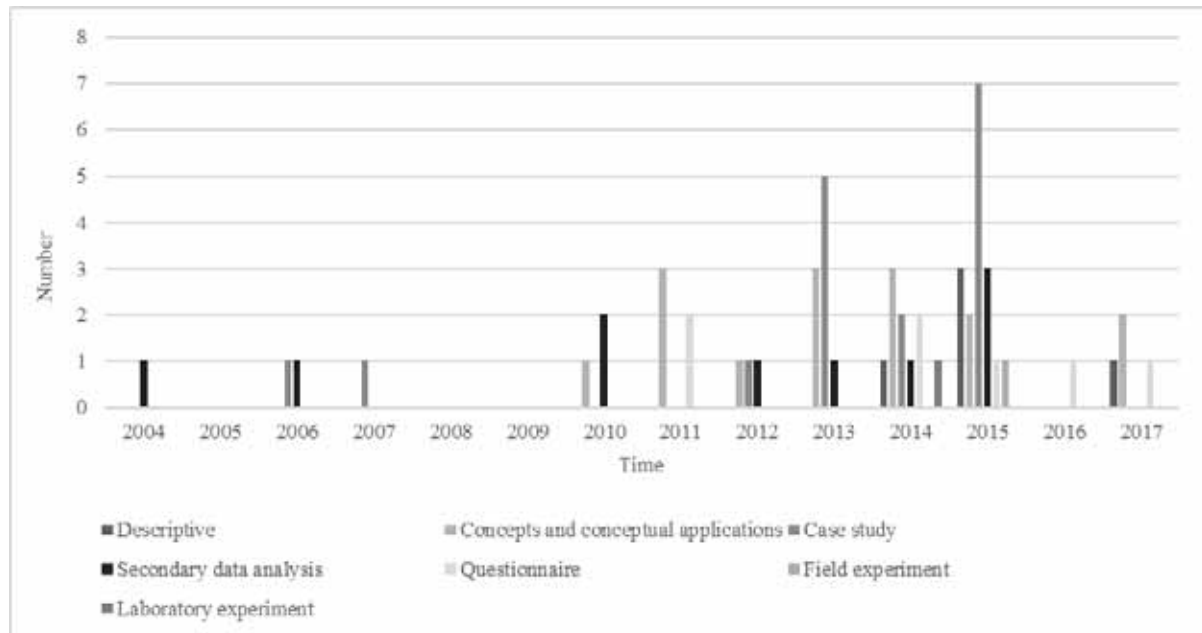


Figure 4. The research methods analysis of ESM related literatures

3.2.5 Basic analysis summary

This chapter mainly illustrates the following three points. First, the arrangement of the research level, the analysis of research topics and methods tells us that ESM research is generally in its infancy, which brings some opportunities for follow-up research. Second, ESM research is an interdisciplinary subject. Finally, the most important focus in the ESM field at present is to explore employee knowledge management and sharing and individual performance, and the focus will continue in the short term. However, with ESM properties and functions' further exploration, the ESM of the future will focus on the enterprise level. Additionally, although exploratory qualitative methods are the mainstream currently, with the development of ESM research, non-exploratory quantitative research will become the focus of scholars' attention and give a feedback to exploratory research.

4. CONCLUSION

Based on the extensive application and importance of ESM, this paper explores the research process of ESM according to the related ESM literatures, and gives the definition of ESM based on enterprise value and internal and external perspectives. And when we considering the duality of ESM. This paper presents seven important properties. Finally, this paper analyzes the basic analysis of 56 extracted literatures. The results show that ESM research is a relatively new field, which has attracted a great deal of attention in various academic fields.

The shortcoming of this paper is that the database selected only contains the foreign periodical literatures, neither domestic research nor the conference essay or the book has been elaborated, which may create some limitations. The next study can consider adding this information. In addition, this paper hopes to combine the

properties and theories in ESM research in the future, explain the impact of ESM adoption from two perspectives within and outside the enterprise, and ultimately accelerate the enterprise organizational change.

REFERENCES

- [1] Aral, S., et al.(2013). Introduction to the Special Issue—Social Media and Business Transformation: A Framework for Research[J]. *Information Systems Research*. 24(1):3-13. <http://pubsonline.informs.org/doi/abs/10.1287/isre.1120.0470>
- [2] Ali-Hassan, H., et al.(2015). Linking dimensions of social media use to job performance: The role of social capital[J]. *The Journal of Strategic Information Systems*. 24:65-89
- [3] Kuegler, M., et al.(2015). What's in IT for employees? Understanding the relationship between use and performance in enterprise social software[J]. *Journal of Strategic Information Systems*. 24(2):90-112
- [4] Treem, J. W. and P. M. Leonardi, *Social Media Use in Organizations: Exploring the Affordances of Visibility, Editability, Persistence, and Association*. 2012, Social Science Research Network: Rochester, NY.
- [5] Wu, L.(2013). Social Network Effects on Productivity and Job Security: Evidence from the Adoption of a Social Networking Tool[J]. *Information Systems Research*. 24:30-51
- [6] Richter, D., et al.(2011). Internet Social Networking: Research State of the Art and Implications for Enterprise 2.0[J]. *Business & Information Systems Engineering*. 3:89-101
- [7] Wehner, B., et al.(2017). Enterprise social networks: A literature review and research agenda[J]. *Computer Networks*. 114:125-142
- [8] Behrendt, S., et al.(2014). Mixed methods analysis of enterprise social networks[J]. *Computer Networks*. 75:560-577
- [9] Beck, R., et al.(2014). Knowledge Exchange and Symbolic Action in Social Media-Enabled Electronic Networks of Practice: a Multilevel Perspective on Knowledge Seekers and Contributors[J]. *MIS Quarterly*. 38:1245-1270
- [10] Stuart, H. C., et al.(2012). Social Transparency in Networked Information Exchange : A Framework and Research Question[J]. *Network Computing*. 451-460
- [11] Treem, J. W. and P. M. Leonardi.(2012). Social Media Use in Organizations: Exploring the Affordances of Visibility, Editability, Persistence, and Association[J]. *Ssrn Electronic Journal*. 36
- [12] Alimam, M., et al.(2017). ITIL perspective on enterprise social media[J]. *International Journal of Information Management*. 37:317-326
- [13] Huang, J., et al.(2013). Reconceptualizing Rhetorical Practices in Organizations: The Impact of Social Media on Internal Communications[J]. *Inf. Manage.* 50:112–124
- [14] Cattani, G. and S. Ferriani.(2008). A Core/Periphery Perspective on Individual Creative Performance: Social Networks and Cinematic Achievements in the Hollywood Film Industry[J]. *Organization Science*. 19:824-844
- [15] Wattal, S., et al.(2010). Network Externalities and Technology Use: A Quantitative Analysis of Intraorganizational Blogs[J]. *J. Manage. Inf. Syst.* 27:145–174
- [16] Li, N., et al.(2015). Reading behavior on intra-organizational blogging systems: A group-level analysis through the lens of social capital theory[J]. *Information & Management*. 52:870-881
- [17] Sarker, S., et al.(2011). The Role of Communication and Trust in Global Virtual Teams: A Social Network Perspective[J]. *Journal of Management Information Systems*. 28:273-310
- [18] Straub, D. W.(1989). Validating instruments in MIS research[J]. *MIS Q.* 13:147-169

Convert Traffic to Purchase: the Impact of Social Network Information on Trust and Purchase Intention in Social Commerce

Yanli Pei^{1*}, Min Zhang¹, Yuqi Zhang¹, Shan Wang²

¹International Business School, Beijing Foreign Studies University, China

²Edwards School of Business, University of Saskatchewan, Canada

Abstract: As social commerce gain popularity, many of them are focused on how to drive huge amount of traffic in social media to online retailer. In this study we adopt experiment method to analyze whether information content and source influence perceived credibility and quality of information, which would then influence consumers' trust and their purchase intention in the social commerce. 203 respondents are recruited and randomly dispatched into 4 treatment groups. The data analysis reveals that experiential information obtains higher source credibility and higher perceived information quality than non-experiential information; user-generated-content (UGC) obtains higher source credibility than marketer-generated-content (MGC), but the correlation between information source and perceived information quality is not significant; both credibility of sources and perceived information quality are positively related to trust and purchase intention in the social commerce. Suggestions have been made for designing social commerce website.

Keywords: social network information, social commerce, traffic conversion, trust

1. INTRODUCTION

Social Commerce is a combination of e-commerce and social media[1], where the users generate content to help others make decisions on purchasing. One form of social commerce is rooted in social network communities, such as Xiaohongshu.com (ref. as XHS) and Mogujie.com (ref. as MGJ) in China. Originally, they are online communities which allow users to share their personal experience and preferences. As social network sites are driving an increasing volume of traffic to retail sites[2], they started to develop their own e-commerce to take advantage of this social influence based on their huge amount of users and user generated content. Then comes to the research questions: How to drive social media traffic to online retailer? What make users trust what others say and what they recommended in the social media and then go to buy on the retail site?

It is reported that XHS is higher than MGJ in terms of the sales volume, the purchase conversion rate, and the repurchase rate. To study why users in XHS could be successfully converted to consumers, we compared the two platforms. The content of MGJ community has an obvious intention of guiding users to purchase and purchase links are attached to the recommended products. Thus community content seems to be generated by marketers. On the contrary, users' comments in XHS are more of users' personal experiences and no purchase links are attached. Moreover, products sold on shopping page are procured by XHS according to users' recommendation rate and not all the recommended products in the community could be bought. Do information type and perceived information source impact on consumers' trust and purchase intention in social commerce?

In this study, experiment method is adopted to analyze whether information content and source influence perceived credibility and quality of information, which would then influence consumers' trust on the social commerce and their intention to purchase. Information type is divided into experiential ones and promotional ones from the perspectives of Experience Consumption Theory, and source of information are categorized into UGC and MGC from the peripheral clues in the Elaboration Likelihood Model. 4 treatment groups (2X2) are designed with descriptions of cosmetics in the social commerce.

2. LITERATURE REVIEW AND THEORETICAL MODEL

Leitner & Grechenig propose that social commerce provide an collaboration platform for consumers, thus making them able to get suggestions, to find appropriate products and to make a purchase decision [3]. Shen & Eder think that social commerce help consumers discovering, integrating and sharing product information, and accordingly making purchase decisions[4].

2.1 Users' community on online shopping

Researchers have great interests in how social network can promote collaboration among users and online sales [5-8]. Uncertainty has been viewed as a severe obstacle for online shopping [9-11]. Some tools such as visualization system and cooperative shopping system can help consumers to deal with the uncertainty [12]. Users' community can also help eliminating uncertainty and support decision making process through sharing product information and providing ratings, interests or personalized recommendations based on purchase history and usage experience. Amblee and Bui prove that social commerce websites use word-of-mouth to affect the reputations of products, and used friends' recommendations, comments and products' ratings to assist purchase decisions [13]. Brich and Holsing analyze the relationship between social networking and online shopping and find that there is a positive correlation between page views, ratings, tags and possibility of click-out (following a link to an online shop) [14]. Therefore, the information on the social commerce website can help eliminating uncertainty, and thus assist purchase decision.

2.2 Trust in the social commerce platform and intention to purchase

In the context of online shopping, uncertainty makes trust a critical factor influencing consumers' intention to buy. Mayer et al. think that trust is the willingness to take the risks rising from uncertainty [15]. Yaobin Lu et al[16] find that building trust among virtual community members is an effective way to establish trust in the C2C website or vendor, and these two types of trust would further influence consumers' intentions to get information and purchase. Hsiao et al[17] find that trust in a website will directly influence consumers' intention to purchase from the website and indirectly through increasing trust in product recommendation in the website and intention to purchase the recommended products.

We can infer that users' trust in the recommendation in the community of the social commerce platform will give rise to the trust in shops of social commerce platform, and thus they will intend to purchase on the social commerce platform. Therefore, we proposed the hypotheses below:

H1: Consumers' trust in the social commerce platform has a positive impact on their intention to purchase.

2.3 Source credibility, perceived information quality and trust in the social commerce platform

What kind of information in users' community can help building consumers' trust in recommended products or websites? According to the Theory of Reasoned Action, rational people will take into account all kinds of information and possible consequences of their behavior before they shape their attitude. Therefore, users' attitudes toward social community platform are mainly based on their judgment on the information in the community, and decide whether to trust in the social commerce and whether to make a purchase.

According to Elaboration Likelihood Mode (ELM) [18], information receivers are impossible to process each piece of information he received elaborately, due to limitation of individual cognition. When an individual is capable of and willing to process information elaborately, the central route of processing information plays the key role and he is more concerned with information quality. When an individual is incapable of and unwilling to process information elaborately, peripheral route, such as the source of the information (source credibility and source similarity), plays the key role [19].

According to Hovland's Communication and Persuasion Theory[20], receivers may understand the information based on their perception and judgement on the sources of information, and a highly credible source is more effective in persuasion. As one dimension to judge the spreaders of information, source credibility make information more influential by enhance its value during the dissemination process [21-24].

Therefore, information source credibility serve as an important peripheral clue that can influence individuals' judgment on information[18, 20]. Particularly in online environment, in order to avoid being confused among huge amount of information, individuals need to rely on peripheral clue such as source credibility to assess information [25, 26]. If people think that the information providers are reliable and professional, they will more trust the information.

H2a Information source credibility has a positive impact on users' trust in social commerce platforms.

According to ELM Theory, in high elaboration mode, persuasion tends to follow the central route and information quality will then become the decisive factor influencing the dissemination of information. Lee finds that higher quality of word of mouth has greater influence [27]. When the information quality within a community is perceived high and helpful, the social commerce platform will be seen as operates from the perspective of consumers. When there is full of useless information, consumer won't develop a favorable impression on the shopping platform. Therefore, perceived quality of information will also influence consumers' trust in the shopping platform. We assume that:

H2b Perceived information quality has a positive impact on users' trust in social commerce platforms.

2.4 Information content and perceived quality, source credibility of information

In the Attribution Theory, Heider[28] pointed out that everyone would try to explain the causes of their own or others' behavior according to some clues. Users in the community are often anonymous, so consumers will form their own perceptions about the source credibility according to information content features. In online consumer community, some contents may describe the real experiences of using products, while some may just copy the marketing notes or functions introduction from the company. The former often contains using experience, feelings, fitness, perceptions of advantages and shortcomings with user's personal tone; the latter often describes ingredients, functions of the products or advocating use efficacy with a promoting tone.

Source credibility usually contains two dimensions: trustworthiness and expertise [20]. Trustworthiness refers to the confidence in the information publisher for providing information in an objective and honest manner. Expertise refers to whether the communicator seems to possess enough knowledge to support his assertions. [29] If the content is about personal experience, consumers will infer that the information publisher has used the product, thus has enough knowledge about the usage of the product. Their suggestions will be thought helpful and objective. Consumers will easily attribute their behavior to internal motivations such as enjoying sharing and believe that such behavior has nothing to do with business. Even if consumers find similar experiences or life backgrounds in the content, they will more likely to trust the source[30]. However, if the content has no personal experience or even has promoting elements, consumers will attribute it to external motivations such as being driven by economic interests. They will question the authenticity and objectivity of the contents, which results in a low level of source credibility. Therefore, we assume that:

H3a Experiential information obtains higher source credibility than non-experiential information

Information quality refers to its accuracy, comprehensive, consistency, etc. Before making decisions, consumers are always eager to know whether it fit me, how does it being used. Others' using experiences contains more details such as aftereffect of use, fitness or compares with similar products. These kind of information are just consistent with consumers' information requirement. In contrast, the non-experiential information only provides product instructions in an official tone. Consumers cannot get enough knowledge on

consequences of using the products. That makes them not confident in stopping searching and making decision. That results in a lower perceived quality of this kind of information. Therefore, we assume that:

H3b Experiential information shows a higher perceived quality than non-experiential information

2.5 Information source and source credibility, perceived information quality

In the field of community in social commerce platform, we focus more on whether the contents come from users or marketers. Some studies have classified community information into user-generated-content (UGC) and marketer-generated-content (MGC), and analyzed their causes, influences as well as the degree of consumers' trust in them [19, 30-32]. While sales personnel often have more knowledge on a product than an inexperienced friend, many consumers doubt salespeople's trustworthiness but would consider a good friend trustworthy [29].

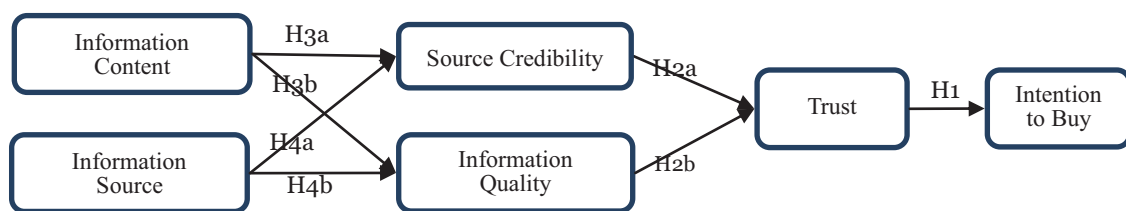
According to Attribution Theory, people infer the communicator's motive based on different situational clues [33, 34]. If the information comes from marketers, consumers may doubt their trustworthiness and consider they are driven by economic interests. But if the information comes from other consumers, people will attribute their motive as enjoying sharing and interaction, etc. The information will easily be trusted because there is no potential conflict of interest. Similarly, under the ELM model, consumers will use peripheral clues—information sources, to understand and process community contents. The credibility of MGC is questionable, for MGC usually emphasizes the positive aspects of products [35]. Therefore, we assume that:

H4a UGC obtains higher source credibility than MGC

Marketers often have more comprehensive information on a product than ordinary consumers [36]. MGC is more capable of providing more relative and professional information thoroughly, so that consumers can spend less time searching and filter information. Different consumers may have different perception or experience with the product. Their opinions are often not consistent or even conflict. UGC are often subjective and has random quality, thus take consumers more time to filter and compare. Therefore, we assume that:

H4b UGC shows a lower perceived information quality than MGC

Based on Communication and Persuasion Theory, ELM and Attribution Theory, we put forward the following model to explore how social network influence online shopping in a social commerce platform.



3. METHODOLOGY

3.1 Treatment Design

In this study, we adopt experiment method to analyze whether information content and source influence perceived credibility and information quality, which would then influence consumers' trust in the social commerce and their intention to purchase. We conduct a pretest and a focus group interview to determine the treatments, experiment materials and objects. Finally, we divide information content into experiential and non-experiential ones based on Experience Consumption Theory, and categorize information source into UGC and MGC from the peripheral clues in ELM model. Thus, 4 treatment cells (2X2) are designed. MGC is defined as content with a click-out button linking to a shopping store, but UGC without that linkage. The content are also examined by pretests and interview to get clearer identification. We finally choose cosmetic products as experiment objects because they are very popular in the social commerce and are high involvement products. In

this research, we recruit 203 respondents, and dispatch almost 50 people to each treatment group randomly.

3.2 Measurement

Table 1 provides an overview of the constructs, their measurements and the sources of the measurements. Each variable is measured by 7 points Likert scale.

Table 1. Measurement

Construct	Items	Sources
Information Content	1. Above content is the publisher's own experience 2. Above content is not the publisher's own experience	Self-developed
Information Source	1. Above content has no shopping chart linking to an online shop, which means the content is generated by users. 2. Above content has a shopping chart linking to an online shop, which means the content is generated by marketers.	Self-developed
Source Credibility	1. I think the content in the above online community is Unreliable Reliable Dishonest Honest Untrustworthy Trustworthy 2. I think the comment in the above online community is knowledgeable. 3. I think the comment in the online community is provided by an expert in the field.	Wu and Shaffer 1987[37][44]
Perceived Information Quality	1. The content in the above community is extensive, and provide comprehensive information about the product. 2. The content in the above community is highly relevant to the product. 3. Content in the above online community that is related to the details of the product is accurate.	Bailey and Pearson 1983[38][45]
Trust	1. I believe the mobile shopping platform will abide by its commitment to customers. 2. I believe this online shopping platform takes customers' best interests close to heart. 3. This online shopping platform is trustworthy. 4. This online shopping platform will not cheat consumers.	Jarvenpaa et al. 2000[39][46]
Intention to buy	1. I would like to buy products or services that I favor on this online shopping platform. 2. I will probably buy products or services that I favor from this platform in the future.	Lim et al. 2006[47]

4. DATA ANALYSIS

4.1 Single dimension test and manipulative check

Table 2 shows the result of single dimension test, in which the minimum value of the initial eigenvalues in principal component analysis of 1.626(bigger than 1), and the maximum value of the second eigenvalues is 0.623(under 1). This means that each observable variable in the group is affected by the same standardized latent variable.

Table2 Single Dimension Test

Variables	Initial Eigenvalues	Second Eigenvalues
Information Content	1.938	0.062
Information Sources	1.961	0.039
Source Credibility	3.499	0.581
Perceived Information Quality	2.102	0.567
Trust	2.868	0.623
Intention to Buy	1.626	0.374

Manipulative check is performed to examine whether subjects exposed to a certain manipulation treatment responded significantly differently. Manipulation checks on information content show that subjects exposed to experiential information indeed agreed that it is publisher's own experience (item1:t=21.6, $p < 0.01$; item2:t=26.4, $p < 0.01$), compared to those subjects exposed to non-experiential information. Manipulation checks on information source show that subjects exposed to content without shopping chart indeed agreed that it is generated by users(item1:t=27.2, $p < 0.01$; item2:t=26.3, $p < 0.01$), compared to those subjects exposed to content with shopping chart.

4.2 Measurement model assessment

We used partial least squares (PLS) to assess the measurement and structure. As shown in Table 3, the composite reliabilities (CR, over 0.6), Cronbach's alphas (over 0.7), and average variances extracted (AVE, over 0.50) by the constructs indicated that they had acceptable levels of reliability and convergent validity. Third, each indicator have a higher loading on its own respective construct than on any other constructs (the result is omitted), and the values of square root of AVE on the diagonal are larger than the correlation below the diagonal (see right part of table 3), which demonstrate convergent and discriminant validity.

Table 3 Reliability test

	Cronbach's Alpha	CR	AVE	Information Content	Information Sources	Source Credibility	Information Quality	trust	Intention to Buy
Information Content	0.968	0.983	0.967	0.983					
Information Sources	0.98	0.99	0.981	0.021	0.99				
Information Quality	0.785	0.875	0.701	0.226	0.247	0.837			
Source Credibility	0.892	0.921	0.7	0.19	0.136	0.75	0.837		
Trust	0.868	0.91	0.717	0.18	0.189	0.811	0.777	0.847	
Intention to Buy	0.77	0.897	0.813	0.223	0.222	0.761	0.749	0.806	0.902

4.3 Structure model assessment and hypothesis test

The structure model is tested using PLS bootstrap. The bootstrapping sample was 500. As shown in Table 4, all the hypothesis are verified except for H4b.

Table 4 Conclusion

Path	Path Coefficient	P-value	Hypothesis	
Trust -> Intention to Buy	0.806	0.000	H1	The consumers' trust has a positive correlation with their intention to purchase. Valid
Source Credibility -> Trust	0.521	0	H2a	Source credibility has a positive correlation with users' trust. Valid
Information Quality -> Trusts	0.386	0	H2b	Perceived information quality has a positive correlation with users' trust Valid
Information Content -> Source Credibility	0.221	0.001	H3a	Experiential information obtains higher source credibility than non-experiential information Valid
Information Content -> Information Quality	0.187	0.002	H3b	Experiential information shows a higher perceived quality than non-experiential information Valid
Information Sources -> Source Credibility	0.242	0	H4a	UGC obtains higher source credibility than MGC Valid
Information Sources -> Information Quality	0.132	0.055	H4b	UGC shows a lower perceived information quality than MGC Invalid

5. CONCLUSIONS AND DISCUSSION

This research study consumers' perceptions of source credibility and quality for different content and different source of information in online communities, and their influence on consumers' trust and purchase intention in social commerce platforms. And we concluded as follows:

Experiential information obtains higher source credibility than non-experiential information, because experiential information seems more expert and customers are more likely to resonate with experiential information and attribute publishers' behavior to some non-economic motivations. Furthermore, experiential information is perceived as higher quality than non-experiential information, because it contains real experience data and is more comprehensive. And the descriptions are more consistent with consumers' requirements of knowing real aftereffect of using.

UGC obtains higher source credibility than MGC, because consumers believe that they are similar with and have no conflict interest with other users in the community. However, the correlation between information source and perceived information quality is not significant. The result shows that although users who generate the content are more trustworthy than marketers, marketers are thought as knowing better about their products.

The quality of the contents generated by users has no significant difference with that of marketers.

Both credibility of sources and perceived quality of information are positively related to trust in the shopping platform. When the source credibility and perceived information quality are high, consumers have stronger trust in the recommendation in social network and then the social commerce platform. That means that consumers' trust improves if they think the publishers are speaking from an unbiased and objective stand. Consumers' trust will in turn improve their willingness to purchase on the social commerce platform.

6. SUGGESTIONS

This research offers managerial implications as follows: First, social commerce platform should encourage customers to post their real experience in the online community and reduce advertisings or promoting contents. These experiential contents can ensure information quality, help consumers making decision and increase user loyalty. Second, social commerce platform should build a healthy connection between social network and online shops in the platforms. In the long run, it is better to make social network relatively independent and prevent retailers take advantage of social network to promote their products. That means that adding click-out button toward online shops to community content is not encouraged.

7. LIMITATION AND FUTURE RESEARCH

This research has several limitations. First, this study did not thoroughly test factors that may influence source credibility and information quality. In future study, we will consider factors such as means of expression and the relation between social network and shops. Second, this study did not observe users for a certain period of time to measure the changes in their perception of information. Third, most sample came from smart phone users. In future studies, reliability can be improved through laboratory experiment. Forth, the empirical findings are from a study in China, so the generalizability of the research is limited.

REFERENCES

- [1] N. Jascanu, V. Jascanu, F. Nicolau. A new approach to e-commerce multi-agent systems [J]. *Annals of Dunarea De Jos*, 2007, 2007(1).
- [2] Y.A. Kim, J. Srivastava. Impact of social influence in e-commerce decision making. the Ninth International Conference on Electronic Commerce: the Wireless World of Electronic Commerce. ACM, University of Minnesota, Minneapolis, Mn, Usa: DBLP, 2007: 293-302
- [3] P. Leitner, T. Grechenig. Community driven commerce: Design of an integrated framework for social shopping. 2007:
- [4] J. Shen, L.B. Eder. Determining Factors in the Acceptance of Social Shopping Websites. Americas Conference on Information Systems, Amcis 2009, San Francisco, California, Usa, August, 2009: 290
- [5] R. Algesheimer, U.M. Dholakia, A. Herrmann. The Social Influence of Brand Community: Evidence from European Car Clubs [J]. *Journal of Marketing*, 2005, 69(3):19-34.
- [6] M. Ma, R. Agarwal. Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities [J]. *Information Systems Research*, 2007, 18(1):42-67.
- [7] P.B. Brandtzæg, J. Heim. User loyalty and online communities: why members of online communities are not faithful. *Intetain*, 2008: 1-10
- [8] F.S.L. Lee, D. Vogel, M. Limayem. Virtual community informatics: A review and research agenda [J]. *Jitta Journal of Information Technology Theory & Application*, 2003, 5(1):47-61.
- [9] A. Dimoka, Y. Hong, P.A. Pavlou. On product uncertainty in online markets: theory and evidence [J]. *Social Science Electronic Publishing*, 2012, 36(2):395-426.
- [10] A. Ghose. Internet Exchanges for Used Goods: An Empirical Analysis of Trade Patterns and Adverse Selection [J].

- Social Science Electronic Publishing,2009,33(2):263-291.
- [11] Y. Kim, R. Krishnan. On Product-Level Uncertainty and Online Purchase Behavior: An Empirical Analysis [J]. *Management Science*,2013,61(10):150402101827007.
 - [12] N. Zmuda. Online shopping goes from solo experience to social interaction [J]. *Advertising Age*,2009DOI.
 - [13] N. Ambler, T. Bui. Harnessing the Influence of Social Proof in Online Shopping: The Effect of Electronic Word of Mouth on Sales of Digital Microproducts [J]. *International Journal of Electronic Commerce*,2011,16(2):91-114.
 - [14] R. Olbrich, C. Holsing. Modeling Consumer Purchasing Behavior in Social Shopping Communities with Clickstream Data [J]. *International Journal of Electronic Commerce*,2011,16(2):15-40.
 - [15] F.D. Schoorman, R.C. Mayer, J.H. Davis. An integrative model of organizational trust: Past, present, and future [J]. *Academy of Management Review*,1995,20(3):709-734.
 - [16] Y. Lu, L. Zhao, B. Wang. From virtual community members to C2C e-commerce buyers: Trust in virtual communities and its effect on consumers' purchase intention[M]. Elsevier Science Publishers B. V.,2010.
 - [17] K.L. Hsiao, J.C.C. Lin, X.Y. Wang, H.P. Lu, H. Yu. Antecedents and consequences of trust in online product recommendations [J]. *Online Information Review*,2013,34(6):935-953.
 - [18] R.E. Petty, J.T. Cacioppo, R. Goldman. Personal involvement as a determinant of argument-based persuasion [J]. *Journal of Personality & Social Psychology*,1981,41(5):847-855.
 - [19] M. Cheung, C. Luo, C. Sia, H. Chen. Credibility of Electronic Word-of-Mouth: Informational and Normative Determinants of On-line Consumer Recommendations [J]. *International Journal of Electronic Commerce*,2009,13(4):9-38.
 - [20] C.I. Hovland, W. Weiss. The Influence of Source Credibility on Communication Effectiveness [J]. *Audiovisual Communication Review*,1953,1(2):142-143.
 - [21] N.H. Anderson. Integration theory and attitude change [J]. *Psychological Review*,1971,78(3):171-206.
 - [22] C.I. Howland, J.L. Irving, H.H. Kelley. Communication and Persuasion: Psychological Studies of Opinion Change [J]. *American Sociological Review*,1954,19(3):355-357.
 - [23] R. Ohanian. Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness [J]. *Journal of Advertising*,1990,19(3):39-52.
 - [24] C. Pompitakpan. The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence [J]. *Journal of Applied Social Psychology*,2004,34(2):243-281.
 - [25] B.J. Fogg, C. Soohoo, D.R. Danielson, L. Marable, J. Stanford, E.R. Tauber, How do users evaluate the credibility of Web sites?: a study with over 2,500 participants[M].2003.
 - [26] M.J. Metzger. Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research [J]. *Journal of the Association for Information Science & Technology*,2007,58(13):2078–2091.
 - [27] J. Lee, D.H. Park, I. Han. The effect of negative online consumer reviews on product attitude: An information processing view [J]. *Electronic Commerce Research & Applications*,2008,7(3):341-352.
 - [28] H. F., The psychology of interpersonal relations[M]. New York, NY:Wiley,1958.
 - [29] R. Ohanian. The impact of celebrity spokespersons' perceived image on consumers' intention to purchase [J]. *Journal of Advertising Research*,1991,31(1):46-54.
 - [30] K.Y. Goh, C.S. Heng, Z. Lin. Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content [J]. *Information Systems Research*,2013,24(1):88-107.
 - [31] C.M.K. Cheung, D.R. Thadani. The impact of electronic word-of-mouth communication: A literature analysis and integrative model [J]. *Decision Support Systems*,2012,54(1):461-470.
 - [32] C. Luo, X. Luo, Y. Xu, M. Warkentin, C.L. Sia. Examining the moderating role of sense of membership in online review evaluations [J]. *Information & Management*,2015,52(3):305-316.
 - [33] E.E. Jones, K.E. Davis. From Acts To Dispositions The Attribution Process In Person Perception 1 [J]. *Advances in*

- Experimental Social Psychology,1966,2(4):219-266.
- [34] H.H. Kelley. The Process of Causal Attribution [J]. American Psychologist,1973,28(2):97-194.
- [35] C.C. Wu, Y.J. Chen, C.J. Wang. Is Persuasive Advertising Always Combative in a Distribution Channel? [J]. INFORMS,2009,28(6):1157-1163.
- [36] K. Ran, I. Simonson. The Effects of Incomplete Information on Consumer Choice [J]. Journal of Marketing Research,2000,37(4):427-448.
- [37] C. Wu, D.R. Shaffer. Susceptibility to persuasive appeals as a function of source credibility and prior experience with the attitude object [J]. Journal of Personality & Social Psychology,1987,52(4):677-688.
- [38] J.E. Bailey, S.W. Pearson. Development of a Tool for Measuring and Analyzing Computer User Satisfaction [J]. Management Science,1983,29(5):530-545.
- [39] S.L. Jarvenpaa, N. Tractinsky, M. Vitale. Consumer trust in an Internet store [J]. Journal of Computer - mediated Communication,2000,1(1-2):45-71.

Exploring Users' Interactive Behaviors in Online Group: A Case

Study of QQ Group “TuanRenTang”

Lei Zhu^{1, 2}, Chunxiang Xue^{1, 2}, Xiuzhi Wu^{1, 2}*

¹School of Economics and Management, Nanjing University of Science and Technology, Nanjing, 210094, China

²Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing, 210094, China

Abstract: The users' interactive behaviors of the online group chat and an accurate identification of users' interaction, which can provide method support for mining user interests and the crowd labeling, was analyzed in this paper. By using social network analysis method, the study took QQ Group “TuanRenTang” as an example to analyze users' interactive behaviors, discover users' interaction relationships, construct interaction networks, and explore the interaction types and community detection. The findings suggested that both explicit and implicit interaction exist in the same topic discussion. Users could be classified into four categories: active interaction, general interaction, passive interaction and lurking interaction based on different user activity. Besides, twenty “experts” and eight communities on the basis of interaction networks had been found out from the sample data of “TuanRenTang” chat records.

Keywords: online discussion group, interactive behavior, interaction, community detection

1. INTRODUCTION

With the rapid development and popularization of mobile internet technology, as real-time interactive and multiple participation communication tools, WeChat Group and QQ Group have gradually replaced the online chat room and BBS to become an important platform for people to get information, and also a major way for companies, agencies and industries to release announcements. E-commerce promotion, fan club and field exchange tend to use the QQ Group, which can hold a larger group of people, and has more functions than the WeChat Group. QQ Group is a place where users can create online relationships with others when they discuss in the group. Nevertheless, the formation of mass online relationships also poses challenges to users. Firstly, the users in QQ Groups are inclined to miss valuable information because of the continuous flow of information into online social networks. Then, the disordered information leads users to be exhausted to view the messages when the QQ Group has a high degree of activity ^[1]. Therefore, it is an urgent problem that how to effectively manage and disseminate QQ Group information. Users' interaction can maintain the generation and transmission of information. Hence, it could be a great value for social network to explore the interactions among users and it is helpful to personalized recommendation, precision marketing and accurate advertising ^[2]. Thus, this paper focuses on the users' interactive behaviors, and their characteristics, the identification of the users' roles of the communities and the experts in QQ Group.

2. RELATED WORK

Interactive behavior is defined as an interactive activity that two or more individuals take part in at the same time ^[3]. The interactive behavior between users is the major route of information dissemination. The booming development of social media (such as Micro-blog, WeChat, QQ, Twitter, etc.) makes the interaction between users more convenient and faster. In the process of generating, accepting, and transferring information,

* Corresponding author. Email: xuechunxiang@njjust.edu.cn(Chunxiang Xue)

there are a lot of interactive relationships among users. But the interactive behaviors vary from one social media to another, such as the behaviors of following, forwarding, commenting, liking, @ and collecting in Micro-blog [2]; the behaviors of posting, replying and reading in BBS [4]; the behaviors of classifying, collecting, commenting and labeling in knowledge community [5]; and the behaviors of retweeting, mentioning and replying in Twitter [6]. User relationships are usually measured by social relationships [7], but social relationships are a static reflection of it. Compared with that, dynamic users' interactive behaviors can directly reveal user relationships.

Online group discussion is a kind of group interaction. Each member plays diverse roles and undertakes different tasks in the group [8]. Chen et al. defined roles from different aspects which included the position of speech, feature of participants and activity degree [9]. Yang divided the subjects into eight different roles according to the types of spreading behaviors, levels of interaction and propagating content [10]. Uddin identified six broad classes of Twitter users by exploiting user's profile and tweeting behavior information [11]. These studies had fully analyzed categories of roles in online social networks. It can help identifying the interaction to explore the users' behaviors and users' roles in the group discussion. Meanwhile, identifying the interaction accurately is the basis of interaction network construction, community detection and key nodes recognition. In terms of requirements of study, this paper made an accurate division of users' roles in QQ group chat according to features of turns.

3. EXPERIMENTAL DESIGN

3.1 Research objectives

The paper conducted an experiment to mine what kinds of potential interactions existed between users in group chat and who were the core users of the group. The exploratory study was conducted on the QQ Group network which was a mobile online social network used by many people from different regions with different ages and occupations. These users created friend links with one another by chatting, sending messages and using various other services. Previous research had shown that chat messaging was the most actively used service in this social network [12]. Thus, the experiment attempted to find users' interaction structure from QQ Group chat messages flow, and then detect user communities and expert users by social network analysis method. The experiment was conducted as the following steps:

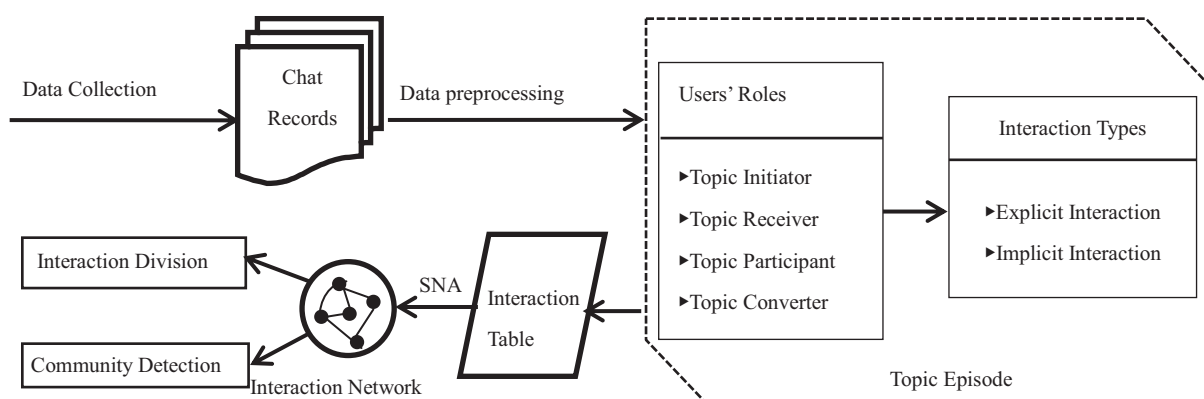


Figure 1. Framework of experiment

3.2 Data collection

This study took QQ Group "Tuan Ren Tang" as the research object, which had high participation and great

activity in the field of library and information science. Most of group members were librarians, college teachers and students, publishing editors and other interested parties in the field. Our dataset covered a wide range of the discussions or topics in politics, travelling, life entertainment, and library activities. Topics of the group involved the submission of papers, development of libraries and academic problems in the information science, which are consistent with the research fields. The sample dataset used in this study was collected from QQ Group chat message manager from February 19 to March 18, 2017. In total, we extracted 4227 messages and 119 participants. The data of chat records was stored in TXT format which included date, time, user name, user ID and conversation (Examples are shown in Table 1).

Table 1. Original date sample table

Date	Time	User Name	User ID	Conversation
2017-02-19	10:49:32	Name1	820134225	请教，人大复印资料，图书类那个刊，全名是什么？
2017-02-19	13:08:21	Name2	1195034327	图书情报资料.
2017-02-19	13:46:39	Name3	1136635921	纠正一下，是叫情报资料工作，抱歉！
2017-02-19	13:57:59	Name3	1136635921	大家收到过这个邮件吗？
2017-02-19	14:22:12	Name4	398371032	没看到这类邮件，是有针对性发的吧.
2017-02-19	14:48:32	Name4	503682783	应该是.

3.3 Data preprocessing

A preprocessing step was performed on the dataset, including date cleaning, date transformation and topic segmentation. Firstly, irrelevant terms or characters were removed (system messages, labels, and blank cells) [13]. After checking the dataset, there existed data loss issue which can increase the complexity of the analysis and cause bias of the results and so on. Thus, the lost data should be fixed refer to chat record from QQ Group chat message manager. Then, in order to facilitate the data processing and data statistics, the data in TXT format was imported into database and exported in the Excel format. All above, topic segmentation was the most important one. The whole record text was cut into pieces manually. One piece is a topic episode, starting with the first message on the topic TS_1 , ending with the TS_1 's sentence that is next to the first sentence of next topic TS_2 . We defined topic episode as TS, regarding that a TS is a topic [14].

If members discuss one topic, it forms interaction relationships between them. Therefore, when they participate in different discussions, their interactions are not visualized easily. How to identify the implicit interaction? The chat messages can reflect the interactive behaviors intuitively. After the topic segmentation, this study made statistics for the time intervals between every two episodes. As is shown in Figure 2, the longest interval can reach 260 minutes, and the shortest interval is 0. That is to say, there is a big gap between the shortest and the longest topic time interval.

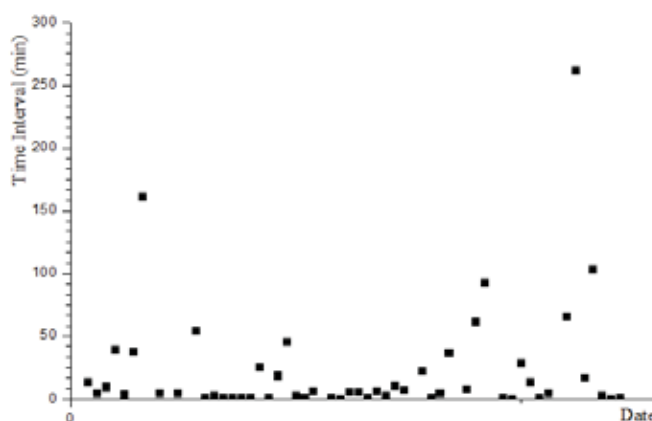


Figure 2. Scatterplot of topic time interval

When the interval is 0, it indicates that two adjoining topics are continuous or crossing. Statistically, 70% intervals are more than 5 minutes. So this paper presented a method for topic segmentation based on a

five-minute interval. That is, if no sentences exist within five minutes after a discussion, we consider this discussion as a topic. ^[15].

4. USERS' INTERACTIVE BEHAVIORS IN ONLINE GROUP DISCUSSION

4.1 Interactive behaviors characteristics

In online group discussion, behavior characteristic is one of the most important user attributes. The users of the same group may have various behaviors. Therefore, this study classified users based on the characteristics of users' interactive behaviors, so as to definite the roles and tasks of users in online group discussion. According to the interactive behaviors of group members in topic discussion, the participants can be divided into four categories: topic initiator, topic receiver, and topic participant and topic converter.

- Topic initiator: the one who brings up a topic and firstly provides a beginning of participants to interact with each other.
- Topic receiver: when topic initiator mentions a member's name or @ a member, this member is regarded as topic receiver.
- Topic participant: all the members in a topic discussion are considered as topic participants. They interact with others by answering questions or giving opinions.
- Topic converter: it is a special kind of role that serves as a connecting link between two adjacent topics. When the depth of original topic discussion has reached the limit, a new topic will be derived ^[16], which usually causes the previous topic to end prematurely or makes the two topics to develop at the same time. The person who proposes a new topic on the basis of original topic is called topic converter.

There are interactions between topic initiator and topic receiver, topic initiator and topic participant, and topic participants. Thus, defining the roles of participants can identify the response relationships. According to knowledge management point of view, interaction comprises explicit interaction and implicit interaction ^[17]. In social network analysis, most studies focus on explicit interaction. For example, replying or retweeting another user's tweet is the explicit interactive behavior in Twitter. In QQ Group, explicit interactive behavior has two forms: @ someone and mention others' name. This paper defined explicit interaction in QQ Group as follows: when $C_i @ C_j (i \neq j)$ or mentions $C_j (i \neq j)$, it can be denoted an explicit interactive relationship between C_i and C_j represented as $E(C_i, C_j)$.

Except explicit interaction, most common interactions are implicit in QQ Group discussion. If C_i and C_j discuss in a topic episode, and there are no @ and mentions, we think that they have implicit interaction represented as $H(C_i, C_j)$. The study generalized that the implicit interaction has the following features:

Table 2. Features of implicit interaction

(1) Linear-feature		(2) Cross-feature	
A L1	• A : Topic initiator	A L1	• A : Topic initiator, and topic participant
B L2	• B, C, D, E, F: Topic participant	B L2	• B, C, D: Topic participant
C L3	They take turns speaking and have implicit	C L3	They speak in no order.
D L4	interactions with A.	B L4	There are implicit interactions between them
E L5	• $H(A, B)$; $H(A, C)$; $H(A, D)$;	A L5	• $H(A, B)$; $H(A, C)$; $H(A, D)$;

F L6 ⋮	H(A, E); H(A, F)	B L6 C L7 D L8 ⋮	H(B, C); H(B, D) H(C, D)
(3) Distinctive-feature			
Form 1		Form 2	
A L1	• A : Topic initiator, and topic participant	A L1	• A : Topic initiator
B L2	• B: Topic participant	B L2	• B, C, D: Topic participant
A L3	Only two users participate in the discussion.	A L3	They talk to A in turn.
B L4	There is implicit interaction between A and B	C L4	There are implicit interaction between A and
A L5	• H(A, B)	A L5	B, A and C, A and D
B L6		D L6	• H(A, B) ; H(A, C); H(A, D)
⋮		⋮	

- Li: each message in a topic episode
- A, B, C, D...: the participants who speak in a topic episode

In fact, explicit interaction and implicit interaction almost both exist in the same topic episode. In observational studies, it can be found that most of the interactions in the topic are implicit. Most of the feature types are cross-features. Each participant speaks freely and has no order. They play a variety of roles according to topics, demands, interests and other factors. Sometimes, a topic initiator may become a topic receiver or topic participant in another topic. It is the reason that we segment the group chat into topic episodes.

4.2 Interaction types division

In order to measure the activity degree and level of interaction, we identified the interaction relationships between users according to the interactive behavior characteristics in QQ group chat, and built a users' interaction table. There were 119 participants and 347 interaction relationships, some of which were shown in Table 3. We counted the number of interactions between each participant with others, and then drew the interaction degree distribution (Shown in Figure 3).

Table 3. Examples of relationships

Source(User ID)	Target(User ID)
1136635921	409143618
1136635921	306554194
1136635921	398371032
1136635921	503682783
1136635921	624685339
1136635921	2020136312
503682783	37825775
503682783	174184576
503682783	2020136312

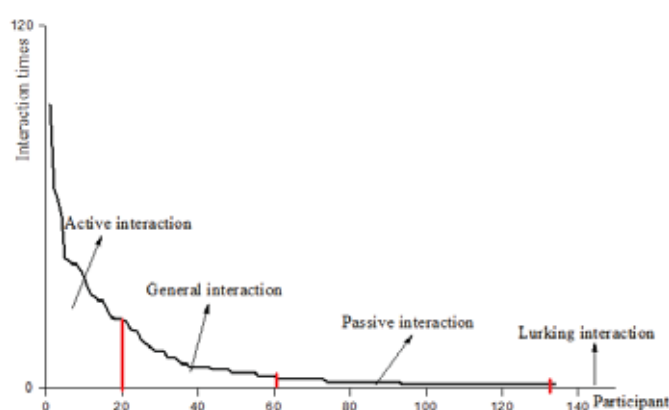


Figure 3. Interaction degree distribution

The study found that there was a large gap among the participants' activity degree. Some users participated in the topic discussions almost every day, while some spoke only once. As shown in the Figure 3, participants can be divided into four categories according to the activity: active interaction, general interaction, negative interaction

and lurking interaction. The members of active interaction represented by “Gaozy” and “Tu Mou(图谋)” are most active in the group chat and interact with others frequently. Therefore, they have stable interaction relationships with someone else and are known as “Experts”. Expert is a person with a high degree of a skill or knowledge of a certain subject^[18]. The members of general interaction don’t have frequent communication with others and they are only involved in the topics they interested such as “State-owned Librarian(国企图书馆员)”. The people like “Xiao Lu(小鹿)” and “Shui Zhongyu(水中鱼)” belongs to negative interaction. Such users don’t take part in the topic discussion actively, besides raising questions and they will leave after getting the answers. Lurking is an activity performed in QQ Group that involves wandering the group chat, but never actually speaking anything in the sample records. With the expansion of time, lurkers may transform into participants when they interact with other members.

4.3 Community detection based on interaction

Community detection is a valuable tool in social media networks. It aims to identify groups of vertices on a graph that are better connected to each other than to the rest of the network^[19]. The interaction network in the online group discussion is a large complex network. Community detection constitutes a significant tool for the analysis of complex networks by enabling the study of network structures and functional characteristics. Therefore, by using social network analysis and visualization tool Gephi, this paper implemented the visual processing to interaction network in online discussion groups, and revealed the depth of relations hidden among them, with steps below. Firstly, the users’ interaction table based on the interactive behaviors in section 4.2 was conducted and saved the table as a CSV file. Secondly, the data of CSV file was imported into Gephi to generate the initial graph. Finally, the result of community detection (Shown in Figure 4) was formed after degree calculation, layout algorithm and modularity.

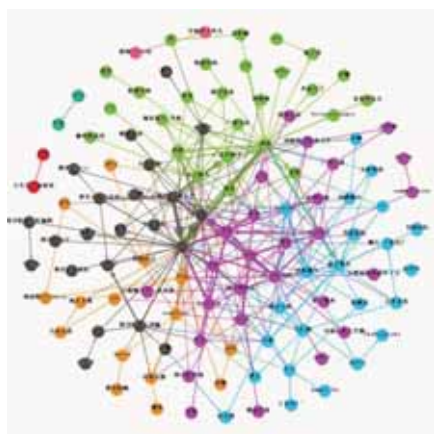


Figure 4. User community

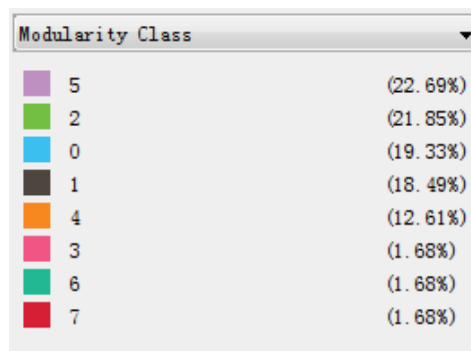


Figure 5. Modularity class

In Figure 4, it is clear that there are eight communities in the network. Figure 5 is the modularity class showing the size of each community. The scale of Community 3, 6 and 7 are very small. The result showed the three communities contain only two members, and the members of them belong to the passive interaction. For example, in a dialogue, the member named “Happy Fish(幸福的大鱼儿)” put forward a question “Hello! Who has digital library resources? (大家好! 请问谁手里有图书馆数字资源?)” “There are lots of digital resources in bidding websites. (数字资源类的单一来源很多招标网站上有很多.)” Lu Dawei(卢大位) answered. The dialogue ended and only encompassed one interaction. So it is true that “Happy Fish(幸福的大鱼儿)” and “Lu Dawei(卢大位)” in Community 3 belong to passive interaction. The sizes of Community 0, 1, 2, 4 and 5 are

almost same holding about twenty to thirty users. The users who belong to these communities are willing to follow and reply to other people in the same community. And the topics they concerned in a community are basically the similar. Therefore, they have closely associated with each. After a series of investigation, it can be concluded that interaction degree doesn't make effect on community division, for this reason, a community may include all three types of interaction (i.e. active, general and passive).

4.4 Summary of main findings

- The roles in online discussion group include four categories: topic initiator, topic receiver, topic participant and topic converter.
- There are two kinds of interactive behaviors: explicit interactive behaviors and implicit interactive behaviors. And most of the interactive behaviors are implicit, which are not easy to be identified.
- Interactions can be divided into four types according to their activity degree: active interaction, general interaction, negative interaction and lurking interaction.
- It can be found that the interaction network in QQ Group has obvious small-world effect. The users of community have strong ties and often participate in the similar topics.

5. CONCLUSIONS AND FUTURE WORK

This study dealt with the problems of expert discovery and community detection by analyzing the users' interactive behaviors and users' interactions in the online group discussion. Community detection can help us to discover the people with similar interests. In terms of that, it can provide personalized service to users efficient and efficiency. Besides, the significance of discovering experts is to construct expertise network. Provided with such expertise network, the newcomers can quickly find out community members with different expertise levels and their relationships. When a user encounters problems, he or she can conveniently consult the right person for the solution.

Although this study focuses on the group chat in the field of library and information science, the research method is also applicable to e-commerce promotion groups and enterprise groups. For the e-commerce promotion group, by analyzing the interactive behaviors of the group members and the content of the topic discussion, it can help find the personal interests and purchase intentions, so that e-commerce companies can label users to advertise accurately, and identify potential customers to conduct personalized recommendation and achieve precision marketing. In addition, for enterprise group, it is benefit to explore organizational structure, and it is also a practical way to boost communication between colleagues, facilitate knowledge sharing, and bring efficiency to work.

However, there are still some problems and deficiencies need to be improved for further study. The study collected only small sample data of the group chat. The data processing is artificial, cockamamie and complicated, which can't avoid the influence of artificial factors. As future work, we intend to incorporate in machine algorithm supporting automatic analysis of large sample data. Meanwhile, chat content is also an important factor in interaction. We intend to combine interests' content with users' interaction, through which it could significantly reveal more insights and ultimately strengthen accuracy of the classification. This paper, as the research-in-progress papers, introduces the preliminary research ideas, but it needs to be optimized in many aspects. We will keep on exploring potential work based we achieved in our experimentation.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China (Grant No.71503126, No.71303111) and the Social Science Foundation of Jiangsu Province (Grant No. 14TQB10).

REFERENCES

- [1] Di Xiaoqiang, Qiu Jin, Li Jinqing, et al. (2017). Analysis of the group chat behavior based on LDA model. *Information Science*, 35(12): 45-49 (in Chinese)
- [2] Xu Jianmin, Li Tengfei, Wu Shufang. (2016). A micro-blogging community discovery method based on user's interactions. *Journal of Hebei University (Natural Science Edition)*, 36(2): 189-196(in Chinese)
- [3] Li Xujun. (2016). Research on information dissemination in social network based on interactive behavior characteristics. Ph D Thesis. Hefei: Hefei University of Technology (in Chinese)
- [4] Si Xiameng, Liu Yun. (2011). Empirical analysis of interpersonal interacting behavior in virtual community. *Acta Physica Sinica*, 60(7): 859-866 (in Chinese)
- [5] Deng Shengli, Hu Jiming. (2010). Analysis of knowledge innovation based on group interaction learning in the Web 2.0 interaction environment. *Information Studies: Theory & Application*, 33(2): 17-20 (in Chinese)
- [6] Omodei E, Domenico M D, Arenas A. (2015). Characterizing interactions in online social networks during exceptional events [J]. *Frontiers in Physics*, 3(15): 59.
- [7] Zhang Shuai. (2016). The user-interactive behavior-based influence maximization in the social network. Ms D Thesis. Kunming: Yunnan University (in Chinese)
- [8] Wang Ning. (2014). Multi-agent group interaction research based on game leaning. Ph D Thesis. Xin Xiang: Henan Normal University (in Chinese)
- [9] Chen Ran. (2012). Research on social network of active group in Network Forum. *Press Circles*, (18): 51-55 (in Chinese)
- [10] Wang Fang. (2011). Research on sharing of tacit knowledge in the virtual community----Taking a QQ Community as an example. Ms D Thesis. Shang Hai: East China Normal University
- [11] Uddin M M, Imran M, Sajjad H. (2014). Understanding types of users on Twitter. *Eprint Arxiv*, 162(12): 1722.
- [12] Lim E P, Correa D, Lo D, et al. (2013). Reviving dormant ties in an online social network experiment. *Science*, 329(5996):1194-1197
- [13] Nugroho R, Yang J, Zhong Y, et al. (2015). Deriving Topics in Twitter by Exploiting Tweet Interactions. *IEEE International Congress on Big Data*. IEEE, 87-94.
- [14] Zhang Wei, Cao Xianbin, Yin Hongzhang. (2009). Chat room social network mining based on multi-features fusion. *Journal University of Science and Technology of China*, 39(5): 540-546 (in Chinese)
- [15] Chen Bikun. (2013). Study the enterprise knowledge sharing in Enterprise Social Network. *Information studies: Theory & Application*, 36(11): 82-86 (in Chinese)
- [16] Yin Xicheng, Zhu Hengmin, Ma Jing, et al. (2015). Coupling network model of the diffusion process of micro-blog topic. *Information studies: Theory & Application*, 38(11): 82-86 (in Chinese)
- [17] Zhang Yu. (2015). Research on interaction behavior influencing factors in informal learning microblog community. Ms D Thesis. Qufu: Qufu Normal University (in Chinese)
- [18] Li Yanyan, Ma Shaoqian, Zhang Yonghe, et al. (2012). Expertise network discovery via topic and link analysis in online communities. *IEEE International Conference on Advanced Learning Technologies*, 311-315.
- [19] Papadopoulos S, Kompatsiaris Y, Vakali A, et al. (2012). Community detection in social media. *Data Mining & Knowledge Discovery*, 24(3): 515-554.

The Evaluation on Home Improvement Website's Efficiency Based on SBM-DEA

Jiangping Wan^{1*}, Jiawen Huang²

^{1,2}School of Business Administration, South China University of Technology, Guangzhou, China

Abstract: We built the efficiency evaluation index system of home improvement website's inputs and outputs. The SBM - DEA model is applied to evaluate the efficiency of 30 home improvement websites and the stability of relatively effective websites. Finally, we put forward some specific website optimization suggestions with the results. The purposes of this paper are to study the efficiency of home improvement websites and provide webmasters with optimization suggestions under limited input resources.

Keywords: SBM-DEA, home improvement website, efficiency evaluation, index system

1. INTRODUCTION

With the accelerated pace of urbanization in China in recent years, the demand for urban housing has increased year by year, which has brought a wealth of demand for home improvement industry. Many new O2O home improvement enterprises were born in the home improvement market in the Internet plus era. At the same time, many traditional home improvement enterprises began to establish e-commerce platform and joined the tide of the Internet home improvement. It is the problem that webmasters need to solve that how to optimize the home improvement website to get a leading position in the fierce competition. Therefore, it is well worth studying the efficiency of home improvement websites and the rationality of website design.

In the literature, various approaches have been proposed to evaluate the website's efficiency. For instance, Kang et al. applied E-S-QUAL for assessing e-commerce website service quality effectively and proposed a fuzzy hierarchical TOPSIS based on E-S-QUAL for evaluating e-commerce [1]. Yi et al. applied the system comprehensive evaluation method to evaluate 40 agricultural enterprise websites in Guangdong Province [2]. Tang et al. built a website user efficiency indicator system based on user behavior model and information construction theory, and selected 12 representative government websites for empirical research [3]. Li et al. evaluated 72 domestic e-commerce websites using factor analysis method and correspondence analysis method [4]. ERTUĞRUL et al. applied MACBETH and PROMETHEE methods to compare performances of online books shopping websites in Turkey [5].

There are many references evaluating website efficiency based on DEA method. Yang et al. measured returns to scale and total factor productivity in e-commerce based on DEA [6]. Yuan et al. applied SBM-DEA model and super-efficiency SBM-DEA model to evaluate the efficiency of 13 domestic insurance e-commerce websites and analyzed the stability of relatively effective websites [7]. Yang et al. applied DEA models and traditional production theory to account for how scale affects efficiency in pure e-commerce firms [8]. Zhang evaluated 24 e-commerce websites based on CCR model [9]. Huang et al. used CCR model, BCC model and super-efficiency model compare the differences of efficiency between provincial archives websites [10].

The purposes of this paper are to study the efficiency of home improvement websites, provide webmasters with optimization suggestions based on results and guide them to improve the efficiency of website under limited input resources. We selected 30 domestic home improvement websites as research sample and built the efficiency evaluation index system of home improvement website's inputs and outputs. Then, we applied the SBM-DEA model to evaluate the efficiency and stability of sample websites. Finally, we put forward some

* Corresponding author. Email: csjpw@scut.edu.cn (Jiangping Wan) , jiawen.huang@hotmail.com (Jiawen Huang)

specific website optimization suggestions with the results.

This paper is organized as follows: section 2 is research method, section 3 is index system construction and data acquisition, section 4 is result and analysis, and section 5 is conclusions and suggestions.

2. RESEARCH METHOD

DEA model was developed by Charnes et al. in 1978^[11]. It is a well-established nonparametric methodology for evaluating the relative efficiency of a set of comparable entities called decision making units (DMUs) with multiple inputs and outputs.

Traditional DEA models, such as CCR model and BBC model, evaluate the efficiency of decision-making units only from the ratio of input to output, but do not consider the problem of slack variables, resulting in the slackness of input-output and the deviation of measured results^[12]. In order to study the efficiency of home improvement websites, we applied the output-oriented SBM model based on laxity variables and the super-efficiency SBM-DEA model to evaluate the website resource input-output efficiency and website stability.

3. INDEX SYSTEM CONSTRUCTION AND DATA ACQUISITION

3.1 Index system construction

We selected input categories and output categories to build the basic index system of home improvement website. Referring to other literature^{[12]-[14]}, we chose the following input and output categories as needed (Figure 1). Besides, we selected 6 input indicators and 3 output indicators to design the efficiency evaluation index system of home improvement website (Table 1).

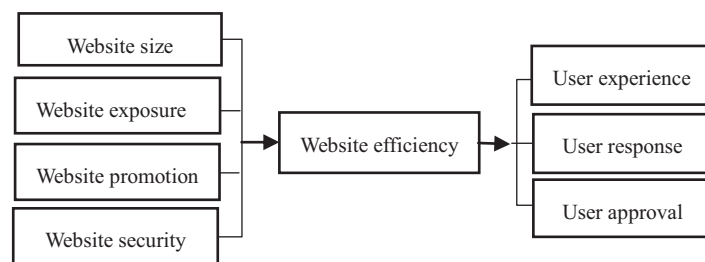


Figure 1. The basic index system of home improvement

Table 1. The efficiency evaluation index system of home improvement website

Evaluation Category		Evaluation indicator
Input indicator	Website size	X ₁ : Total number of web pages X ₂ : Average page size
	Website exposure	X ₃ : Number of distinct foreign sites referenced
	Website promotion	X ₄ : Total number of meta tags X ₅ : Total number of foreign sites references
	Website security	X ₆ : Website security test score
	User experience	Y ₁ : Website access speed
Output indicator	User response	Y ₂ : Daily page view
	User approval	Y ₃ : Baidu index

The input and output indicators in Table 1 are explained as follows:

(1) The input indicators

The total number of web pages (X₁) refers to the total number of static pages in the root directory and subdirectory. The total number of website pages reflects the overall scale of website construction. A large number of pages means a large overall scale and high cost of maintaining the website^[7].

The average page size (X₂) refers to the average size of all pages in the website (calculated by KB). The larger the average page size, the more abundant the content of the website.

The number of distinct foreign sites referenced (X₃) refers to the total number of links to different external websites on this website. To some extent, this indicator reflects the cost of maintaining the stability of the

content and the internal friend links ^[7].

The total number of meta tags (X_4) refers to the total number of meta tags used throughout the website, which reflects the input of website promotion.

The total number of foreign sites references (X_5), also called the total number of backlink, refers to the number of different external websites pointing to the website. The broad exposure can attract search engine crawlers to crawl the website quickly, which can enhance the speed of web content collection and improve the website experience. Therefore, this indicator also reflects the input of website promotion.

Security test score (X_6) refers to a score of vulnerability, tampering or fraudulent content. This indicator reflects the input of website safety and has a direct impact on the outputs.

(2) The output indicators

Website access speed (Y_1) reflects the user experience in the process of visiting the website. A page with a slow load speed can leave a bad impression on users. Thus, we used the reciprocal of website access speed as one of the output indicators.

Daily page view (Y_2) reflects the effect of user response. We can learn if a website has a good content readability and a high conversion rate from this indicator.

Baidu index (Y_3) refers to the index of a website in keywords ranking, traffic and user experience. This indicator reflects the comprehensive optimization of a website. We chose this indicator because better website optimization means a higher level of user approval.

3.2 Data acquisition

(1) Sample website selection

We selected the top 30 home improvement websites as the DMUs from 2016 annual list of the best Internet home improvement enterprises ^[15] (Table 2).

Table 2. The sample websites

No.	Name	URL	No.	Name	URL
1	Tubatu	http://www.to8to.com	16	Wanshifu	http://www.wanshifu.com
2	Maelline	https://jz.mmall.com	17	Boluoni	http://www.boloni.cn
3	Qijia	http://www.jia.com	18	Meilele	http://www.meilele.com
4	Liba	http://www.liba.com	19	Baianju	http://www.bnq.com.cn
5	Kujiale	http://www.kujiale.com	20	Aikongjian	http://www.ikongjian.com
6	Qianggongzhang	http://www.7gz.com	21	Huizhuang	http://www.huizhuang.com
7	Oupai	http://www.oppein.cn	22	Longfa	http://www.bjlongfa.cn
8	Juranzhijia	http://www.juran.com.cn	23	Mingjia	http://www.mjdec.com
9	Aifuwo	http://www.fuwo.com	24	Xingyi	http://www.xydec.com.cn
10	Taipingyang	http://www.pchouse.com.cn	25	Xinju	http://www.homekoo.com
11	Woaiwojia	http://www.525j.com.cn	26	Hongmayi	http://www.hmyzs.com
12	Tubashu	http://www.tobosu.com	27	Jikemeijia	http://www.mj100.com
13	Jintanglang	http://www.jtljia.com	28	Pingguo	http://www.apple2003.com
14	Dongyirisheng	http://www.dyrs.com.cn	29	Yuanzhou	http://www.yz-china.com
15	Wangzhu	http://www.lcds365.com	30	City family	http://www.csrj.com.cn

(2) Data acquisition

After confirming the sample websites, we used Maxamine Web Analyst software to collect the total number of web pages, the average page size, the number of distinct foreign sites referenced, the total number of meta tags and the total number of foreign sites references. Then, we used Chinaz Tool (<http://tool.chinaz.com>) to collect website security test score, website access speed and Baidu index. Finally, we used Alexa Tool to collect daily page view. Our acquisition work was completed within 2 hours from 9:00 am to 11:00 am in order to make the collected data comparable and reliable. And the acquisition work lasted for 15 days from May 1 to May 15

in 2017. Finally, we calculated the average number of each indicator within 15 days as the model data (Table 3).

Table 3. The collected model data

DMU	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	Y ₁	Y ₂	Y ₃
1	59.7	483	518	180	3,889	96.0	3,248,000	0.240	8
2	251.3	561	22	60	564	100.0	72,100	0.857	4
3	55.8	627	356	32	1,341	96.0	178,000	0.581	6
4	89.0	23	32	13	301	100.0	195,500	0.255	6
5	13.0	1,745	56	4	72	100.0	592,000	0.948	6
6	54.0	1,557	926	184	2,057	97.0	511,000	0.492	4
7	119.9	573	171	26	536	95.0	48,500	0.978	6
8	968.3	149	198	2	2,587	97.0	50,000	1.567	4
9	26.5	634	214	79	990	100.0	440,500	0.676	4
10	13.9	769	239	21	1,352	98.0	326,500	0.333	7
11	126.9	558	681	140	2,692	100.0	38,900	0.584	6
12	71.7	626	923	91	4,479	100.0	170,000	0.437	7
13	149.6	34	96	76	488	100.0	30,200	0.634	2
14	43.5	574	302	35	1,230	99.0	70,100	0.391	4
15	294.5	551	109	86	631	100.0	22,300	0.700	1
16	12.0	3,657	40	6	109	34.0	32,000	0.728	4
17	205.6	159	85	235	937	50.0	21,500	0.643	1
18	54.6	1,011	45	4	70	53.0	344,500	0.392	8
19	515.8	154	18	129	415	100.0	10,400	1.007	4
20	221.3	236	535	407	3,634	100.0	19,100	1.245	2
21	24.7	28	32	135	285	100.0	14,750	0.911	1
22	7.5	598	30	14	182	100.0	21,200	0.641	3
23	281.4	533	31	284	1,882	95.0	9,750	0.323	3
24	1,709.7	176	78	348	1,383	100.0	30,300	0.413	4
25	52.2	564	154	33	1,323	100.0	163,000	0.349	6
26	926.5	477	47	182	756	100.0	10,550	0.639	2
27	74.8	418	10	514	1,411	39.5	8,750	1.340	1
28	1,728.0	595	1,864	98	2,077	100.0	22,800	4.739	4
29	239.8	513	398	289	3,571	90.5	12,400	0.941	3
30	1,479.7	451	1,836	480	9,883	100.0	9,570	1.058	4

4. RESULT AND ANALYSIS

4.1 Website efficiency analysis

We used MaxDEA 6.0 software to calculate the technical efficiency (TE) and pure technical efficiency (PTE) with SBM-O-V model and SBM-O-C model. Then, we calculated the scale efficiency (SE) and returns to scale (RTS) according to $SE = TE/PTE$ (Table 4).

Table 4. The result of 30 home improvement websites based on SBM-DEA

DMU	TE	PTE	SE	RTS	DMU	TE	PTE	SE	RTS
1	1	1	1	Constant	17	0.1531	1	0.1531	Increasing
2	1	1	1	Constant	18	1	1	1	Constant
3	0.6786	0.6880	0.9862	Decreasing	19	1	1	1	Constant
4	1	1	1	Constant	20	0.0661	0.0800	0.8257	Decreasing
5	1	1	1	Constant	21	1	1	1	Constant
6	0.4152	0.4403	0.9431	Decreasing	22	1	1	1	Constant
7	1	1	1	Constant	23	0.1006	0.1092	0.9214	Increasing
8	1	1	1	Constant	24	0.1610	0.1626	0.9902	Decreasing
9	0.7615	0.7627	0.9983	Decreasing	25	0.4723	0.4838	0.9762	Increasing
10	1	1	1	Constant	26	0.0848	0.0866	0.9799	Decreasing
11	0.0507	0.0508	0.9989	Decreasing	27	1	1	1	Constant
12	0.2459	0.2478	0.9926	Decreasing	28	1	1	1	Constant
13	0.4369	0.4379	0.9979	Decreasing	29	0.0178	0.0206	0.8646	Decreasing
14	0.2284	0.2587	0.8829	Increasing	30	0.0120	0.0121	0.9922	Decreasing
15	0.0952	0.0967	0.9843	Decreasing	Mean	0.5993	0.6313	0.9496	
16	1	1	1	Constant	S.D.	0.4035	0.3981	0.1518	

Table 4 illustrates the TE, PTE, SE and RTS of the sample based on SBM-DEA model. The technical efficiency of 14 websites are equal to 1, indicating that these websites are in the production frontier of SBM model with relatively effective outputs. There are 11 websites with efficiency values between 0 and 0.4. It is obviously that over one-third of sample websites have low output efficiency. Thus, it is necessary for these websites to improve resource utilization and adjust inputs according to the shadow prices. Moreover, the average value of TE is 0.5993, which suggests 40.07% of website resources are ineffective in website outputs. The standard deviation of TE is 0.4035, which indicates that there is a considerable efficiency gap between the sample websites. It is illustrated that a polarization phenomenon exists in the input efficiency of home improvement websites in China.

As for scale efficiency, 14 websites whose SE value are equal to 1 are of scale effective, and 15 websites whose SE value are between 0.8 and 1 are of light scale non-effective. The mean SE value is 0.9495 and the standard deviation of SE is 0.1518 indicating that most of the sample websites have reached the best scale, or have a small difference with the best scale. Thus, the overall scale of sample websites is relatively stable. In terms of RTS, 14 scale effective websites are in the CRS stage, which indicates that the scale of each website has a reasonable correspondence with the inputs and outputs. Thus, these websites can maintain their current scale and develop steadily. 12 scale non-effective websites are in the DRS stage, so they should properly reduce website scale and input resources to cut down the cost. 4 scale non-effective websites are in the IRS stage, indicating that their output growth rate is higher than the input growth rate. Therefore, these websites should increase the input and scale to make full use of scale effect for a CRS stage.

As for PTE, the mean PTE value is 0.6312 indicating that the overall technology of the sample websites is at a medium level with much improvement space. The standard deviation of PTE is 0.3981, which suggests huge technical gap might exists among some websites.

The mean PTE value (0.6312) is smaller than mean SE value (0.9495). As PTE is mainly influenced by the management and technical level of websites and SE is mainly influenced by the scale of websites, this clearly illustrates that the efficiency of input resources is lower than that of website scale construction among home improvement websites. Therefore, home improvement websites need to increase their technical input according to the stage of their own scale returns, thereby enhancing the overall technical efficiency.

4.2 Super efficiency analysis of effective websites

Based on super-efficiency SBM-O-V model, we calculated the super efficiency values and analyzed the stability of 14 effective websites by means of DEA-SOLVER Pro software (Figure 3).

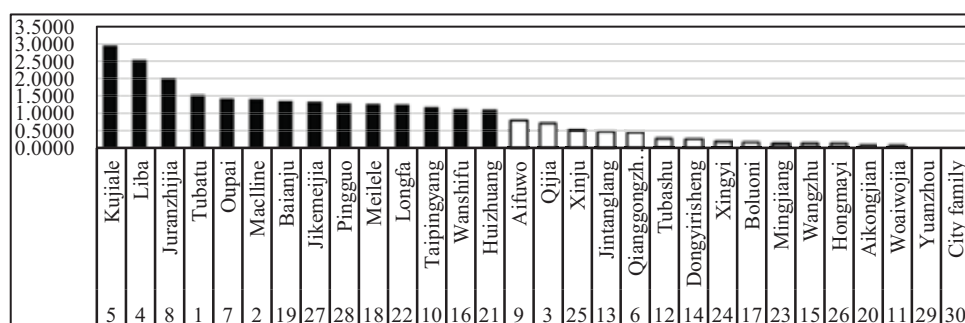


Figure 3. Super efficiency ranks of home improvement websites

Figure 3 presents the super efficiency value of the sample websites, illustrating that the super efficiency of the 16 non-effective websites are the same as those of SBM-DEA, but the super efficiency of the 14 effective websites are quite different from those of the original SBM-DEA. It indicates the relatively effective websites remain effective in different stability. The higher the super efficiency, the higher the stability when a website

increasing input resources. Take Kujiale(5) and Macline (2) for example. Kujiale's super efficiency value is 2.9091 and Macline's super efficiency value is 1.3987. As illustrated in Figure 4, 2 websites are effective when increasing their input resources by 1.3987 times. However, when increasing input resources by 2.9091 times, Kujiale still stay relatively effective while Macline becomes non-effective website because of high inputs ratio.

Compared with 2016 annual list of the best Internet home improvement enterprises, the relatively effective websites in the top 10 of the list still ranks top in terms of stability maintenance. For instance, Kujiale ranks No.1 in terms of stability and ranks No.5 in the annual list, Tubatu ranks No.1 in the annual list and ranks No.4 in terms of stability, indicating that the home improvement websites with strong strength can maintain relatively effective and obtain more ideal outputs when increasing input resources. However, top 10 to 30 of annual list varies widely in terms of stability. Jikemeijia(27) and Pingguo(28), at the end of the annual list, have higher stability than other websites which are in the front rank. It suggests that the lower-ranking websites can increase their input resources and site's scale appropriately under the condition of effectively maintaining their stability.

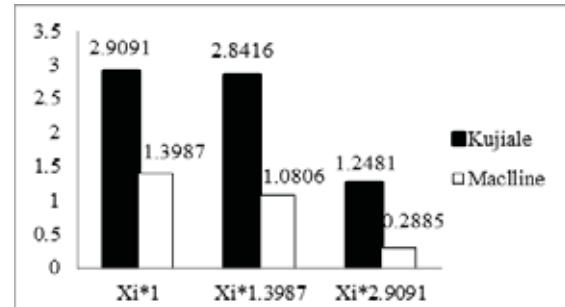


Figure 4. The super efficiency values of input changes

4.3 Input redundancy and output insufficiency analysis of non-effective websites

Based on SBM-I-V model and SBM-O-V model, we used MaxDEA 6.0 software to calculate the projection of input and output indicators of 16 non-effective websites. Then we calculated input redundancy rate and output insufficiency rate so as to put forward some suggestions. The formulas of input redundancy rate and output insufficiency rate are as follows:

$$\text{Input redundancy rate} = \frac{\text{original value} - \text{projection value}}{\text{original value}} \quad (1)$$

$$\text{Output insufficiency rate} = \frac{\text{projection value} - \text{original value}}{\text{original value}} \quad (2)$$

It can be seen from Table 5 that all 9 indicators illustrate the results of input redundancy and output insufficiency. The result of SBM-I-V model illustrates that non-effective websites have a high input redundancy rate in both website exposure and website promotion with an average rate of 50%, and the rate of 68% in the total number of foreign sites references (X_5). It illustrates that most of the home improvement webmasters prefer to website promotion inputs and tend to use META tags and links for SEO.

As for every single website, it can be found that City family (30) with the lowest TE has a high degree of redundancy in all input indicators except the security test score. And its redundancy rate of total number of foreign sites references is 95% and the insufficiency rate of Baidu index and daily PV are 64% and 24517%, which indicates there is a phenomenon that the current input of website is larger than the input of ideal website's scale and the output is extremely scarce. This is partly because the website pays too much attention to the accumulation of resources, but ignores the effective allocation of resources, resulting in high redundancy, page browsing shortage and poor responsiveness. Therefore, City family (30) should reduce foreign links and META tags in the promotion and keep the high quality links.

Xingyi (24) with low TE has a small number of input resources. For example, the total number of web pages and the number of distinct foreign sites referenced are only 176 and 78, while the average size of pages is as high as 1709.7KB, which reflects the resource allocation of the website is unreasonable, such as the average web page size is too large, the problem of slow page loading speed etc. Therefore, Xingyi (24) should reorganize its page layout, reduce the size of displayed images and add some service content, thereby enhancing its user experience.

Even if the input redundancy rate is close to zero, it does not mean that the resources of the website is well allocated, which may be the illusion caused by insufficient input. For instance, Boluoni (17) has no input redundancy or output insufficiency. However, when combine with its security test score of only 50 points, we can find Boloni (17) has not been repaired a number of high-risk vulnerabilities. Moreover, the Baidu index is closely related with traffic and keyword rankings. But the Baidu index of Boluoni (17) is only 1 point, which means the website has not been well invested in efficient resources to optimize the search engine. Therefore, Boluoni (17) should add security resources, repair web page vulnerabilities, improve website security system construction, and carry out necessary search engine optimization.

Table 5. The input redundancy rate and output insufficiency rate of non-effective websites

DMU	TE	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	Y ₁	Y ₂	Y ₃
		Input redundancy rate						Output insufficiency rate		
		Original value						Original value		
3	0.6786	0.00	0.00	0.58	0.40	0.47	0.02	1.36	0.00	0.00
		55.8	627	356	32	1341	96	178000	0.581	6
6	0.4152	0.67	0.61	0.89	0.80	0.65	0.00	2.82	0.24	0.76
		54	1557	926	184	2057	97	511000	0.492	4
9	0.7615	0.09	0.00	0.62	0.26	0.46	0.07	0.88	0.00	0.05
		26.5	634	214	79	990	100.0	440500	0.676	4
11	0.0507	0.19	0.51	0.86	0.86	0.85	0.02	55.98	0.00	0.10
		126.9	558	681	140	2692	100	38900	0.584	6
12	0.2459	0.00	0.00	0.92	0.87	0.93	0.23	9.11	0.00	0.00
		71.7	626	923	91	4479	100	170000	0.437	7
13	0.4369	0.45	0.14	0.62	0.00	0.26	0.00	3.15	0.00	0.70
		149.6	34	96	76	488	100	30200	0.634	2
14	0.2284	0.45	0.00	0.89	0.64	0.85	0.05	8.02	0.00	0.57
		43.5	574	302	35	1230	99	70100	0.391	4
15	0.0952	0.96	0.18	0.72	0.50	0.67	0.00	24.43	0.00	3.59
		294.5	551	109	86	631	100	22300	0.700	1
17	0.1531	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		205.6	159	85	235	937	50	21500	0.643	1
20	0.0661	0.00	0.62	0.56	0.72	0.87	0.00	34.11	0.00	0.37
		221.3	236	535	407	3634	100	19100	1.245	2
23	0.1006	0.69	0.90	0.03	0.81	0.79	0.00	22.70	1.27	0.51
		281.4	533	31	284	1882	95	9750	0.323	3
24	0.1610	0.96	0.86	0.59	0.88	0.79	0.00	15.06	0.00	0.39
		1709.7	176	78	348	1383	100	30300	0.413	4
25	0.4723	0.00	0.00	0.71	0.69	0.83	0.17	2.94	0.26	0.00
		52.2	564	154	33	1323	100	163000	0.349	6
26	0.0848	0.94	0.95	0.32	0.54	0.61	0.00	30.18	0.00	1.47
		926.5	477	47	182	756	100	10550	0.639	2
29	0.0178	0.74	0.43	0.81	0.57	0.87	0.00	142.15	0.00	0.70
		239.8	513	398	289	3571	90.5	12400	0.941	3
30	0.0120	0.81	0.78	0.85	0.86	0.95	0.00	245.17	0.00	0.64
		1479.7	451	1836	480	9883	100	9570	1.058	4
Mean	—	0.44	0.37	0.62	0.59	0.68	0.03	37.38	0.11	0.62
		371.17	516.88	423.19	186.31	2329.81	95.47	108573.13	0.63	3.69

5. CONCLUSIONS AND SUGGESTIONS

The conclusions are as follows: (1) Chinese home improvement websites have a medium overall efficiency and a stable overall scale. (2) The website scale construction is more effective than the input resources in home improvement websites. The medium standard deviation of PTE reflects huge technical gap might exists among some websites. (3) Compared with 2016 annual list of the best Internet home improvement enterprises, the relatively effective websites in the top 10 of the list still ranks the top in terms of stability, while the top 10 to 30 of the list varies widely in terms of stability. (4) Most of the webmasters prefer to website promotion inputs and tend to use META tags and links for SEO, which illustrates a lack of website promotion accuracy.

The following suggestions are put forward: (1) Optimize website resource allocation. Webmasters need to redistribute and optimize the input resources, such as allocating the extra input to Baidu index, PV and access speed so as to improve the website's efficiency. (2) Pay attention to core service content. Some home improvement websites have beautiful page design only in introduction of enterprise and service. But the content of purchase and user interaction, which can bring profits, are designed in online customer service module. Therefore, the page design of home improvement website should be optimized not only in the beauty of the page but also in the purchase content to enhance the quality of content construction and attract and keep users. (3) Improve website's promotion quality. The greater the external impact of backlinks, the higher the cost of maintaining the stability of the website. Therefore, the websites with high redundancy should appropriately reduce the number of backlinks and META tags, pay attention to the quality of the links and the reliability of the link promotion platform. (4) Optimize search engine. A high search engine rankings means high probability of users accessing the website and a big number of page views. Thus, some redundant resources should be allocated to the website's search engine optimization. Improving the quality of links and keyword can improve the website's efficiency. (5) Improve the website scale according to the current situation. The relatively effective websites can combine their super efficiency values and enterprises' current situation to increase the input resources of their own website and core competence.

REFERENCES

- [1] Daekook Kang, Wooseok Jang, Yongtae Park. (2016). Evaluation of e-commerce websites using fuzzy hierarchical TOPSIS based on E-S-QUAL. *Applied Soft Computing*, 42(5): 63-65.
- [2] Yi Famin, Rong Xianmin. (2010). Cluster analysis of agricultural e-commerce websites in Guangdong province. *Forum on Science and Technology in China*, 3: 129-133(in Chinese).
- [3] Tang Zhiwei, Ai Qiqi, SongLi. (2014). Empirical study of government website evaluation based on user performance. *Journal of Intelligence*, 33(05): 198-202+152(in Chinese).
- [4] Li Junjun, Chen Famin. (2008). Ecommerce website evaluation based on factor analysis and correspondence analysis. *Information Science*, 26(08): 1252-1256(in Chinese).
- [5] ERTUĞRUL İ, ÖZTAŞ G. (2016). Performance Analysis of Online Bookstores by Using MACBETH and PROMETHEE Methods. *Journal of Internet Applications & Management*, 7(2): 21-38.
- [6] Zhuofan Yang, Yong Shi, HongYan. (2017). Analysis on pure e-commerce congestion effect, productivity effect and profitability in China. *Socio-Economic Planning Sciences*, 57(3): 35-49.
- [7] Yuan Feng, Chen Junting, Shao Xiangli. (2015). The evaluation on insurance e-commerce website's efficiency based on SBM-DEA. *Insurance Study*, 3:36-45(in Chinese).
- [8] Zhuofan Yang, Yong Shi. (2016). Scale Expansion and Operations Inefficiency in Pure E-commerce Firms. 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops, Omaha: IEEE, 54-57.
- [9] Zhang Qitong. (2014). B2C E-Commerce Websites Evaluation and Strategy Research. Ms D Thesis. Tianjin: Tianjin University,
- [10] Huang Lixia, Zou Chunlong. (2016). Research on efficiency evaluation of provincial archives website based on DEA method. *Archives Management*, 6: 34-36(in Chinese).
- [11] Cheng Gang. (2014). Data Envelopment Analysis and MaxDEA Software. Beijing: Intellectual Property Press, 4
- [12] Tan Hongmei. (2010). DEA Model Based on Information Technology in China Case Study of Performance Evaluation of Listed Companies. Ms D Thesis. Chengdu: Xihua University,
- [13] S Kaur, K Kaur, P Kaur. (2016). An empirical performance evaluation of universities website. *International Journal of Computer Applications*, 146(15): 10-16.
- [14] Ping Sun, David A. Cárdenas, Rich Harrill. (2016). Chinese Customers' Evaluation of Travel Website Quality: A Decision-Tree Analysis. *Journal of Hospitality Marketing & Management*, 25(4): 476-497.
- [15] Ai Yi. (2017). 2016 Internet home improvement business rankings. *China Internet Weekly*, 2: 58-60(in Chinese).

Seller Channel Choice and Optimal Pricing on Heterogeneous Platforms under Online Price Comparison System

Qingfang Sang

¹School of Business and Administration,
Zhongnan University of Economics and Law
Wuhan, 430073, China

Abstract: This paper studies the channel choice and pricing strategy for sellers who are facing heterogeneous e-commerce platforms under the price comparison system. In such a system, consumers can see the sales information of the same product on different platforms, so they will compare and choose between different sellers and platforms. This paper portrays the behaviors of consumer's price comparing and builds a decision model of sellers in different channel selection modes which are based on the demand functions of consumer utility and Hotelling model. The optimal pricing and maximum profit of the sellers in different selection patterns will be obtained by solving the model and the final results can provide a decision-making reference for sellers who are in the face of similar situations.

Keywords: price comparison system, channel choice, online pricing, heterogeneous platforms

1. INTRODUCTION

With the rapid development of electronic commerce and Internet new technology, online shopping has become a trend and the various e-commerce platforms have also swarmed. These platforms provide a complete network infrastructure and online services to enable both parties to complete the entire transaction. In fact, different platforms also have different operating modes, they can be simply divided into two categories: one is the open platform, which does not charge the seller and the buyer's any fees, such as Taobao. The other is the for-profit platform, which will charge buyer and seller's a certain transaction costs in order to achieve better transaction protection and the perfect customer service, such as Tmall, Amazon^[1]. In addition, consumers tend to have different degrees of acceptance of the platform, this can be comprehended from the following two aspects: (1) In terms of enterprise scale, reputation and sales, the entrance threshold of for-profit platform is higher than that of open e-commerce platform, so consumers tend to choose the for-profit platform that can provide higher quality goods. (2) The for-profit platform often provides a more perfect consumer protection mechanism than open e-commerce platform, which enables consumers to enjoy a more perfect shopping experience. At the same time, the changes in consumer shopping habits promote them to find products with good quality and low price, and there are a lot of online price comparison plug-ins now. These browser plug-ins can help consumers to observe the price of multiple online platforms with the same series of products, including the recent price movements, so consumers can select the most suitable items. For sellers, they are not independent of the sales process, the existence of these plug-ins makes the market competition more intense. The platform selection mode and pricing strategy of competitors will affect other sellers' decisions, so this paper focuses on the following questions: (1) How does the seller choose the channel when they are facing the heterogeneous e-commerce platform? (2) How does the competitors' decision affect other sellers' decisions? (3) How does the consumer's price comparison behavior affect the seller's decision? (4) How to determine the optimal pricing in the established selected pattern?

¹ Qingfang, Sang. Email: sangqingfang@foxmail.com
Corresponding Author: Xuan, Jiang. Email: judyx.jiang@foxmail.com

The issue of channel choice and price strategy has always been the focus of academic research at home and abroad. In fact, some scholars focus on online channel, because the open platform and for-profit platform also have their own characteristics, we can see the heterogeneous platforms as two channels^{[2]-[3]}. With the popularity of online shopping, more and more sellers prefer to set up shop on the high-quality e-commerce platforms^[1]. Bilateral markets are also should be paid attention to it, Armstrong focuses on the competition in the bilateral market and discussed the equilibrium pricing under the different choice modes between the monopoly platform and the competitive platform^[4]. In addition, many scholars also pay attention to the sellers' pricing strategy under the dual channel, Chiang, Dumrongsiri and Cai follows closely about the optimal pricing of the various stakeholders in the supply chain when both online and traditional physical stores exist^{[5]-[7]}. In addition, some scholars also concern the pricing issues under the state of competition, Lin(2006) investigates the impact of value-added services on the sellers' optimal prices and their own needs, he also extends the model into many competitive sellers^[8]. Under the background of stochastic demand, Cao considers the pricing and coordination problems of closed loop supply chain when two sellers are competing, and made specific analysis in two cases of centralized decision and decentralized decision respectively^[9]. In the study process, the influence of the heterogeneity of consumers on sellers' decision also can't be ignored. Chen divides the consumers into strategic and short-sighted types, and analyzes the dynamic game pricing in the process of multiple sellers' price war^[10]. Wang also studies the impact of strategic consumers and short-sighted consumers on retailer pricing and inventory strategy^[11]. Besides, Zhang shows the optimal strategy of digital product supply chain based on consumer preference^[12]. Anyway, the research on network price comparison has received more attention in recent years. Bai analyzes the influence of the proportion of price comparison to consumers of manufacturers and retailers in a two-level supply chain system that composed by a manufacturer and a seller^[13]. Shi considers the effect of retailer risk aversion on manufacturer channel selection under the network price comparison^[14]. And Li discusses the pricing problem of online sellers under the price comparison behavior of strategic consumers and gave the optimal pricing of online sellers under dynamic pricing and price promises respectively^[15]. Shen considers the pricing and coordination strategy of dual channel under the behavior of network price comparison, and use game theory to discuss the decision model of single channel and dual channel supply chain^[16]. Deng explores the pricing strategy of sellers when adopting dual channel and showed that consumers' price comparison behavior can be reduced and the price perception can be improved in a specific way^[17].

Actually, there are few studies based on consumer network price behavior, and current studies mainly focus on dual channels. The channel choice and pricing of sellers when are facing two heterogeneous platforms is rarely studied, this paper depicts the price behavior of consumers and shows the decision-making model of two competing sellers. The optimal pricing and profit of the sellers in different decision-making modes will be obtained after analysis, which will provide a reference for sellers.

2. BASIC MODEL

2.1 Model description

There are two heterogeneous e-commerce platforms: A, B , A is the for-profit platform and charges f_s, f_b to seller and consumer respectively, the fees are mainly used for the maintenance of the platform, the guarantee of the products and the protection of consumers' rights. B is the open platform and charges no fees. The various mechanisms are provided by the A platform, which make it easier and more secure for consumers to buy. Therefore, it is assumed that the basic purchasing utility is v in the A platform and θv in the B platform for consumers. v is the consumer valuation and is uniformly distributed within the consumer population from 0 to 1,

with a density of 1, and θ is called the customer acceptance of the open platform. The model in this paper is developed for those products with $\theta \in [0, 1]$.

Sellers can enter the open platform at any time because of the low entrance threshold and it's assumed that two sellers, s_1, s_2 already exist on the open platform before. But the for-profit platform A adopts the way of directional investment, only those who have successfully entered the platform brand pool can apply for entry. The aim of this measure is to cooperate with good brands, control product quality indirectly and provide good services for consumers. Assuming that platform A appears on the market now, s_1 has successfully entered the platform brand pool through the application, but s_2 has failed to enter the brand pool for its own reasons, so s_1, s_2 can choose the modes as:

- (1) s_1, s_2 only settle on the B platform;
- (2) s_1 settles on the A and B platform at the same time, and s_2 settles only on the B platform..

Besides, assuming that A and B platform have its loyal consumers respectively, and the proportion of loyal customers is N_A, N_B accordingly. Customers often search platform products through the keywords, and be ready to pay after comparison between different sellers. At this time, the price plug-ins will show the price information of the same good on the other website, then consumers will decide whether to jump to another website to purchase, so the decision order of sellers and customers is as follows:

- (1) The seller sets the selling price under different choices;
- (2) Consumers first choose between different products, then choose between different platforms after observing the price information.

Using the Hotelling model to describe the consumer's choice between sellers on the same platform, and the parameter t represents the costs of consumers searching online, so the market share when sellers are on the A

platform is $q_{iA} = \frac{1}{2} + \frac{p_{jA} - p_{iA}}{2t}, i \in \{1, 2\}$, the demand of sellers on the B platform at the same time is similar.

The parameters description are as follows:

Table 1. Parameters description	
Parameter	Implication
$p_{iA}, p_{iB}, i \in \{1, 2\}$	The pricing of s_1, s_2 on the A, B platform, respectively
θ	Customers' acceptance of the open platform
t	The search cost for consumers accessing the Internet
f_s, f_b	Fees charged to the sellers and the consumers by the for-profit platform, respectively
$Q_{iA}, Q_{iB}, q_{iA}, q_{iB}$ $i \in \{1, 2\}$	The demand of s_1, s_2 on the A, B platform and the market share of s_1, s_2 on A or B
Π_{s1}, Π_{s2}	The total profit of s_1, s_2 respectively

2.2 Scenarios discussion

2.2.1 Scenario 1: s_1, s_2 sell product on the B platform only

When the loyal consumers on the B platform decide to buy product from $s_i, i \in \{1, 2\}$, the basic utility needs to

satisfy $\theta v - p_{iB} \geq 0, i \in \{1, 2\}$, v is the consumption value. According to the consumer choice model, the number of consumers buying $s_i, i \in \{1, 2\}$ product on the B platform is $N_B(1 - \frac{p_{iB}}{\theta}), i \in \{1, 2\}$, that is:

$$Q_{iB} = N_B(1 - \frac{p_{iB}}{\theta})$$

Consumers will choose between the competing $s_i, i \in \{1, 2\}$ on the B platform and the Hotelling model can be used to get the market share of the sellers, so the number of consumers from B platform is willing to choose $s_i, s_j, i, j \in \{1, 2\}$ is:

$$q_{iB} = \frac{1}{2} + \frac{p_{jB} - p_{iB}}{2t}, i, j \in \{1, 2\}, i \neq j$$

The profit of $s_i, i \in \{1, 2\}$ selling goods on the B platform is:

$$\begin{aligned} \Pi_{iB} &= Q_{iB}(p_{iB} - c_i) \\ &= N_B(1 - \frac{p_{iB}}{\theta})(\frac{1}{2} + \frac{p_{jB} - p_{iB}}{2t})(p_{iB} - c_i) \end{aligned}$$

Assuming that $c_i = 0$, after the partial derivative and analysis, the optimal pricing and profits are as follows,

$$\begin{aligned} p_{1B} &= p_{2B} = t + \frac{\theta}{2} - \sqrt{t^2 + \frac{\theta^2}{4}} \\ \Pi_{iB} &= \frac{N_B(1 - \frac{t + \frac{\theta}{2} - \sqrt{t^2 + \frac{\theta^2}{4}}}{\theta})(t + \frac{\theta}{2} - \sqrt{t^2 + \frac{\theta^2}{4}})}{2}, i = 1, 2 \end{aligned}$$

Proposition 1 The optimal pricing and profit for sellers in scenario 1 is: $p_{1B} = p_{2B} = t + \frac{\theta}{2} - \sqrt{t^2 + \frac{\theta^2}{4}}$,

$$\Pi_{iB} = \frac{N_B(1 - \frac{t + \frac{\theta}{2} - \sqrt{t^2 + \frac{\theta^2}{4}}}{\theta})(t + \frac{\theta}{2} - \sqrt{t^2 + \frac{\theta^2}{4}})}{2}, i = 1, 2.$$

Proposition 1 shows that when s_1, s_2 sell on the open platform only, the optimal pricing and profit of them is related to consumers' acceptance of the open platform and the online search costs and is positive correlated with both.

2.2.2 Scenario 2: s_1 sells product on the A and B platform, s_2 sells product on the B platform only

According to the idea of backward induction, the consumer's decision on the platform choice should be first expressed. Because s_1 sells product on the A and B platform, s_2 sells product on the B platform only, the loyal consumers of the A platform can only buy s_1 goods on the A platform, and there is no other brand choice. They just have to decide to buy on A or B platform, if they decide to buy on the A platform, the following conditions should be satisfied:

$$\begin{cases} v - p_{1A} \geq 0 \\ v - p_{1A} \geq \theta v - p_{1B} \end{cases} \quad (1)$$

If they decide to buy on the A platform, the following conditions should be satisfied:

$$\begin{cases} \theta v - p_{1B} \geq 0 \\ v - p_{1A} \leq \theta v - p_{1B} \end{cases} \quad (2)$$

After the analysis, the number of loyal customers from A platform decides to buy from s_1 on A is:

$$Q_{As1}^A = \begin{cases} 1 - \frac{p_{1A} - p_{1B}}{1 - \theta}, & \text{if } \frac{p_{1B}}{\theta} \leq p_{1A} \\ 1 - p_{1A}, & \text{else} \end{cases} \quad (3)$$

The number of loyal customers from platform A decides to buy from s_1 on B is:

$$Q_{As1}^B = \begin{cases} \frac{\theta p_{1A} - p_{1B}}{\theta(1 - \theta)}, & \text{if } \frac{p_{1B}}{\theta} \leq p_{1A} \\ 0, & \text{else} \end{cases} \quad (4)$$

If consumers from B want to choose s_1 , they will see the price information of s_1 product on the A, B platform through the price comparison system and then decide to buy from A or B . So if consumers are willing to buy the product of s_1 on A , the condition should satisfy like (1), the demand Q_{BS1}^A is similar to (3), and if consumers are

willing to buy the product of s_1 on B , the condition should satisfy like (2), the demand Q_{BS1}^B is similar to (4).

Besides, since s_2 sells product on the platform B only, consumers who want to buy from s_2 can conclude s_2 on the platform B through the price comparison system and will go to the platform B totally.

According to the idea of backward induction, now it is necessary to determine the consumers' choice from platform A and B on the product brands s_1 and s_2 . As s_1, s_2 sell on the platform B in scenario 1, the ratio of consumers willing to choose $S_i, S_j, i, j \in \{1, 2\}, i \neq j$ is $q_{iB} = \frac{1}{2} + \frac{p_{jB} - p_{iB}}{2t}$, and because there is only s_1 on the platform A , customers on platform A will buy from s_1 totally, so they just need to make a decision on which platform to buy the product.

In brief, the profits brought to s_1 by consumers buying goods through A and B platforms are as follows:

$$\Pi_{s1} = \begin{cases} [(1 - \frac{p_{1A} - p_{1B}}{1 - \theta})(p_{1A} - f_s) + \frac{\theta p_{1A} - p_{1B}}{\theta(1 - \theta)} p_{1B}] [N_A + (\frac{1}{2} + \frac{p_{2B} - p_{1B}}{2t}) N_B], & \text{if } \frac{p_{1B}}{\theta} \leq p_{1A} \\ (p_{1A} - f_s)(1 - p_{1A}) [N_A + (\frac{1}{2} + \frac{p_{2B} - p_{1B}}{2t}) N_B], & \text{else} \end{cases}$$

The profits brought to s_2 by consumers buying goods through the B platform are as follows:

$$\pi_{s2} = p_{2B} N_B \left(\frac{1}{2} + \frac{p_{1B} - p_{2B}}{2t} \right)$$

When $\frac{p_{1B}}{\theta} > p_{1A}$, the final profit of s_1 can be solved as $\Pi_{s1} = 0$, this is the case that s_1 does not want to see, so

we don't discuss this condition. And when $\frac{p_{1B}}{\theta} \leq p_{1A}$, the final pricing can be obtained after solving:

$$\begin{cases} \partial \pi_{S1} / \partial p_{1A} = 0 \\ \partial \pi_{S1} / \partial p_{1B} = 0, \\ \partial \pi_{S2} / \partial p_{2B} = 0 \end{cases}$$

And the results are as follows:

$$\begin{aligned} p_{2B} &= \frac{p_{1B} + t}{2} \\ p_{2A} &= p_{2B} + \frac{f_s - \theta + 1}{2} \\ p_{1B} &= \frac{(4N_A / N_B + 3)t}{4} + \frac{3\theta - M}{8} \end{aligned}$$

$$\text{And } M = \sqrt{4t(3 + \frac{4N_A}{N_B})[t(3 + \frac{4N_A}{N_B}) - q] + \frac{q[8(1 - f_s)^2 + q(16f_s - q - 7)]}{1 - q}} \quad (M \text{ is meaningful})$$

The equilibrium solutions under this model need to satisfy the following conditions:

$$\frac{p_{1B}}{\theta} \leq p_{1A}, 0 \leq p_{1A} \leq 1, 0 \leq p_{iB} \leq \theta, i \in \{1, 2\}$$

So the parameters need to satisfy the following conditions in order to make the solutions exist:

$$0 \leq f_s \leq 1 + \theta, 0 \leq t \leq 2\theta, 0 \leq p_{1B} \leq \min \left\{ \frac{\theta(f_s - \theta + 1)}{2(1 - \theta)}, \theta, 2\theta - t, \frac{1 - f_s + \theta}{2} \right\}$$

In order to simplify the calculation, we select a part of the scope to discuss: When $0 \leq t \leq \theta, 0 \leq f_s \leq 1 - \theta$,

$$\theta \leq t(w + 3) \leq \frac{5}{2}\theta, w = \frac{4N_A}{N_B}, p_{1B} \text{ should satisfy: } 0 \leq p_{1B} \leq \frac{\theta(1 - \theta + f_s)}{2(1 - \theta)}$$

After calculation, the final results are as follows:

$$(1) \text{ if } 0 \leq \theta \leq \frac{1}{6}, 0 \leq t \leq \theta, \theta \leq t(w + 3) \leq \frac{5}{2}\theta,$$

$$\begin{cases} \text{if } 0 \leq f_s \leq 1 - \theta - \sqrt{2t(w + 3)(1 - \theta)}, p_{1B} = 0; \\ 1 - \theta - \sqrt{2t(w + 3)(1 - \theta)} \leq f_s \leq 1 - \theta, p_{1B} = \frac{(w + 3)t}{4} + \frac{3\theta - M}{8} \end{cases}$$

$$(2) \text{ if } \frac{1}{6} < \theta < \frac{1}{3}, 0 \leq t \leq \theta$$

$$\begin{cases} \text{if } \theta \leq t(w + 3) \leq \frac{1 - \theta}{2}, \begin{cases} \text{if } 0 \leq f_s \leq 1 - \theta - \sqrt{2t(w + 3)(1 - \theta)}, p_{1B} = 0; \\ 1 - \theta - \sqrt{2t(w + 3)(1 - \theta)} \leq f_s \leq 1 - \theta, p_{1B} = \frac{(w + 3)t}{4} + \frac{3\theta - M}{8} \end{cases} \\ \frac{1 - \theta}{2} \leq t(w + 3) \leq \frac{5}{2}\theta, p_{1B} = \frac{(w + 3)t}{4} + \frac{3\theta - M}{8} \end{cases}$$

$$(3) \text{ if } \frac{1}{3} < \theta < 1, 0 \leq t \leq \theta, 0 \leq f_s \leq 1 - \theta, \theta \leq t(w + 3) \leq \frac{5}{2}\theta$$

$$p_{1B} = \frac{(w + 3)t}{4} + \frac{3\theta - M}{8}$$

So the final optimal price and profit of S_1, S_2 are:

if $p_{1B} = 0$

$$\begin{aligned}
 p_{1A} &= \frac{f_s - \theta + 1}{2}, p_{2B} = \frac{t}{2}, \\
 \text{and } \begin{cases} \Pi_{s1} = N_B \frac{(w+3)(f_s + \theta - 1)^2}{16(1-\theta)} \\ \Pi_{s1} = N_B \frac{t}{8} \end{cases} \\
 \text{if } p_{1B} &= \frac{(w+3)t}{4} + \frac{3\theta - M}{8} \\
 p_{1A} &= \frac{4f_s + 4 - \theta - M}{8}, p_{2B} = \frac{3\theta - M}{16} + \frac{(w+7)t}{8}, \\
 \text{and } \begin{cases} \Pi_{s1} = N_B \left[\frac{w+3}{4} + \frac{u}{2t} \right] \left[\frac{(4u + f_s + \theta - 1)}{4(1-\theta)} + \frac{4u^2(\theta-1) - u\theta(f_s - \theta + 1)}{\theta(1-\theta)} \right] \\ \Pi_{s1} = N_B \frac{(t-2u)^2}{8t} \end{cases} \\
 \text{And } w &= \frac{4N_A}{N_B}, u = \frac{M}{16} - \frac{3\theta}{16} - \frac{t(w+3)}{8}
 \end{aligned}$$

Proposition 2 The optimal price and profit of sellers in scenario 2 is influenced by various parameters, different range of parameters will determine the final pricing and profit.

3. CONCLUSIONS

This paper studies the channel choice and pricing strategy of sellers facing heterogeneous platforms under the price comparison system. We show the optimal pricing strategy and the maximum profit of s_1 and s_2 when they sell goods on the B platform as well as s_1 on the A, B platform and s_2 on B platform. The research shows that when s_1 and s_2 only sell goods on the B platform, the optimal pricing strategy and profit of s_1, s_2 are affected by two factors, that is the acceptance of consumers to B platform and the online search cost of consumers. And when s_1 is on the A, B platform at the same time and s_2 is on the B platform, the optimal pricing decision of the sellers is different under different parameters range. The future research needs to compare the profits of the two sellers in two scenarios and get the optimal decision combination of them through numerical analysis.

ACKNOWLEDGMENT

Thanks the support from the research foundation of Zhongnan University of Economics and Law Under Grant 2016GS03.

REFERENCES

- [1] Jiang X, Deng S.(2014). Seller Investment Incentive on Heterogeneous Platforms Under Competition. International Journal of u-and e-Service, Science and Technology, 7(5): 39-50.
- [2] Hagiu A.(2007). Proprietary vs. open two-sided platforms and social efficiency. Division of Research, Harvard Business School
- [3] Nocke V, Peitz M, Stahl K.(2007). Platform ownership. Journal of the European Economic Association, 5-6: 1130-1160.
- [4] Armstrong M. (2006). Competition in two-sided markets. The RAND Journal of Economics, 37(3): 668-691.

- [5] Chiang W Y K, Chhajed D, Hess J D.(2003).Direct Marketing, Indirect Profits: A Strategic Analysis of Dual Channel Supply Chain Design. *Management Science*, 49(1):1-20.
- [6] Dumrongsiri A, Fan M, Jain A, Moinzadeh, K.(2008).A supply chain model with direct and retail channels.*European Journal of Operational Research*, 187(3): 691-718.
- [7] Cai G, Zhang Z G, Zhang M.(2009).Game theoretical perspectives on dual-channel supply chain competition with price discounts and pricing schemes. *International Journal of Production Economics*,117(1):80-96.
- [8] Lin Zhibing, Cai Chen, Xu Baoguang.(2006).Pricing analysis in retailers competing Mode.*Chinese Management Science*,(05):87-90(in Chinese).
- [9] Cao Xiaogang, Wenhui.(2015).Pricing and coordination of closed loop supply chain with retailer competition under stochastic demand.*Operations and Management*,24(01):34-39(in Chinese).
- [10] Chen Xiaohong,Tan Yunqiang.(2015). Dynamic game pricing of multiple retailers considering double factor of strategic consumers.*Journal of Management Engineering*,29(04):178-185.
- [11] Wang Xiayang,Fuke,Yubin.(2016).Retailer pricing and inventory decision considering the change of residual value and customer heterogeneity.*Chinese Management Science*,24(1):97-106.
- [12] Zhang Yanli,Hu Xiaojian,Yang Haihong.(2017).The pricing model of green products under the preference of consumer heterogeneity and environmental protection.*Industrial. Statistics and Decision*,(06):40-45(in Chinese).
- [13] Bai Shizhen,Chen Xiaoli.(2017).Research on pricing decision of dual channel supply chain based on customer price behavior.*Industrial Technology Economy*,36(01):68-74(in Chinese).
- [14] Shi Chunlai,Ni Jiajia,Mu Yanfeng.(2016).The influence of risk aversion on the choice of manufacturer channel under the network price.*Industrial Engineering and Management*,21(06):53-59-67(in Chinese).
- [15] Lijuan,PuYang Xiaojuan.(2017).Pricing strategy of e-retailers with discount selling period. *Journal of Management Engineering*,31(3):149-154.
- [16] Shen Chenran,Xiong Zhongkai,Yanwei.(2014).Research on dual channel pricing and coordination strategy under network price comparison behavior.*Chinese Management Science*,22(1):84-93(in Chinese).
- [17] Deng Xidong,Zhubin,Ding Xuefen.(2016).Dual channel decision making considering inter channel price behavior. *Industrial Engineering and management*,21(1):23-28(in Chinese).

Quantifying the Effect of Mobile Channel Visits on Firm Revenue

Yingnan Zhao, Quansheng Wang, Peijian Song, Fan Feng*

School of Business, Nanjing University, Nanjing, 210093, China

Abstract: The explosive penetration of mobile devices is one of the most prominent trends in e-business. Although the importance of mobile channel has prompted growing literature, little is known about the revenue implications of customer visit toward mobile channel. This study examines (1) the differential effect of mobile visits in affecting firm revenue (i.e. mobile vs. desktop visits), and (2) which type of mobile visits are more effective (i.e., direct vs. search engine and referral traffic; visits for high vs. low involvement products). We collect an unique objective daily data from a leading online travel agency in China. With a vector autoregressive (VAR) method, we find that, compared with desktop channel, mobile channel visits have shorter carryover effect, but larger short-term effect on firm revenues. Further, mobile channel has larger short-term effect on firm revenues for search engine traffic and lower involvement products. Our findings provide important theoretical contributions and notable implications for mobile commerce strategy.

Keywords: mobile channel, mobile commerce, firm revenue, vector autoregressive method

1. RESEARCH QUESTION

The explosive penetration of mobile devices is one of the most prominent trends in customer-firm interactions. According to a survey by Forrester Research, 91% online retailers in the U.S. have a mobile strategy in place or in development, 48% retailers have a mobile website. Despite the emerging trends of mobile commerce, marketers' beliefs about the value of mobile channel seem to be at best mixed, if not negative. The importance of mobile channel has prompted growing literature focused on the mobile channel transactions. However, little is known about the transaction revenue implications of customer visit toward mobile channel, and what is the difference between mobile channel and desktop channel in value creation. Furthermore, traffic source characteristics were largely overlooked in the literature. Hence, we explicitly quantify different value mobile channel traffic monetized, and explore how different traffic sources and products might differentially contribute to transactions.

2. MAJOR RESEARCH FINDINGS

We analyze characteristics between mobile and desktop channel including convenience, targeting, search cost and payment risk. Based on information process theory and decision-making literature, we compare their effects in short- and long- term. On the one hand, mobile channel can support time-critical activities and facilitate immediate transactions because of convenience. Marketers can also target location-sensitive promotional offers to mobile devices users^[1]. When customers receive the timely location-based promotion, they will transact immediately if they are interested in. Therefore, the whole duration between need recognition and transaction is short in mobile channel, and mobile visits have larger effect on short-term revenues than desktop visits. On the other hand, low search cost makes it feasible for customers to acquire full knowledge of product information in the search stage. And because of the advantage in transaction risk, customers will transact high risk products in desktop channel and engage in extensive searches which further increases the duration between need recognition and transaction. Therefore, compared with mobile channel visits, the desktop channel visits have longer carryover effect on firm revenues. Considering the mobile visits mainly have short-term effects on firm revenue, we differentiate different types of mobile visits: three sources of traffic and

* Corresponding author. Email: songpeijian@nju.edu.cn (Peijian Song), zhaoyingnan@smail.nju.edu.cn (Yingnan Zhao), wangqs@nju.edu.cn (Quansheng Wang)

products with different level of involvement.

We choose online travel agency as our empirical setting. We estimated a VAR model, where endogenous variables are traffic and transaction revenue in both mobile and desktop channel. We also included a vector of the exogenous variables such as, an intercept C ; a deterministic-trend variable T , which captured the impact of the omitted, gradually changing trend of the variables; indicator for holidays H (holidays as 1 and non-holidays as 0). The VAR specification is given by Equation 1:

$$\begin{bmatrix} \ln TM_t \\ \ln TD_t \\ \ln RM_t \\ \ln RD_t \end{bmatrix} = \begin{bmatrix} C_{TM} \\ C_{TD} \\ C_{RM} \\ C_{RD} \end{bmatrix} + \begin{bmatrix} \delta_{TM} \\ \delta_{TD} \\ \delta_{RM} \\ \delta_{RD} \end{bmatrix} \times T + \begin{bmatrix} \gamma_{TM} \\ \gamma_{TD} \\ \gamma_{RM} \\ \gamma_{RD} \end{bmatrix} \times H \\ + \sum_{j=1}^J \begin{bmatrix} \varphi_{11}^j & \varphi_{12}^j & \varphi_{13}^j & \varphi_{14}^j \\ \varphi_{21}^j & \varphi_{22}^j & \varphi_{23}^j & \varphi_{24}^j \\ \varphi_{31}^j & \varphi_{32}^j & \varphi_{33}^j & \varphi_{34}^j \\ \varphi_{41}^j & \varphi_{42}^j & \varphi_{43}^j & \varphi_{44}^j \end{bmatrix} \begin{bmatrix} TM_{t-j} \\ TD_{t-j} \\ RM_{t-j} \\ RD_{t-j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{TM,t} \\ \varepsilon_{TD,t} \\ \varepsilon_{RM,t} \\ \varepsilon_{RD,t} \end{bmatrix} \quad (1)$$

where TM is traffic in mobile channel, TD is traffic in desktop channel, RM is transaction revenue in mobile channel, RD is transaction revenue in desktop channel, t refers to days, j equals the number of lags included, and ε is white-noise disturbances distributed as $N(0, \Sigma)$. Both δ , γ and φ are parameters that need to be estimated. Following Dekimpe and Hanssens^[2], we used Generalized IRFs to ensure that the order of variables in the system did not affect the results. Standard errors were derived by simulating the fitted VAR model by Monte Carlo simulation with 1,000 runs to test the statistical significance of parameters ($p = 0.10$). Duration of carryover effects were measured by wear-out time, that is the lag number of periods it takes after the peak impact before transaction revenue effects die out. We determined the duration of impact as equal to the last period in which the IRF value had a $|t|$ -statistic greater than 1.96.

We find that mobile channel and desktop visits show a significant positive relationship with firm revenue (0.19 and 0.12 basis points respectively, $p < 0.01$) for short-term. This finding suggests that mobile channel visits are more effective in affecting firm revenue. Desktop channel visits demonstrate a longer wear-out time (10 days) than mobile channel visits (2 days), although insignificant in the long-term. In addition, search engine visits have a significant relationship with firm revenue (0.12 basis points, $p < 0.10$) in the short-term, while referral visits are not significantly related to firm revenue (0.06 basis points, $p > 0.10$). Direct visits show an insignificant effect in the short-term (0.00 basis points, $p > 0.10$), and the coefficient of direct visits is smaller than that of search engine and referral visits. Referral traffic does not achieve significant effect that may result from the methods diverting referral traffic. This kind of traffic are without explicit shopping goal. In addition, due to some of the significant technical limitations display advertisements face in mobile, some literatures suggest that this kind of ads may not be effective. Mobile visits for high involvement products are positively associated with firm revenue (0.16 basis points, $p < 0.01$) in the short-term, and mobile visits for low involvement products are also significantly but more effectively associated with firm revenue (0.23 basis points, $p < 0.05$). Thus, compared with high involvement products, mobile visits of low involvement products have a larger short-term effect on firm revenues. Moreover, results that mobile visits for both high and low involvement products are not significantly associated with firm revenue in the long run.

3. CONCLUSIONS

Given the increasing penetration of mobile devices among consumers, a better understanding of mobile channel value creation is needed. This study finds that, compared with desktop channel, mobile channel visits have shorter carryover effect, but larger short-term effect on firm revenues. Further, mobile channel has larger short-term effect on firm revenues for search engine traffic and lower involvement products.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71472083, 71372035.

REFERENCES

- [1] Hui, S. K., Inman, J. J., Huang, Y., & Suher, J. 2013. The effect of in-store travel distance on unplanned spending: Applications to mobile promotion strategies. *Journal of Marketing*, 77(2): 1-16.
- [2] Dekimpe, M. G. & Hanssens, D. M. 1999. Sustained spending and persistent response: A new look at long-term marketing profitability. *Journal of Marketing Research*, 36(4): 397-412.

Research on Spillover Effect of Paid Search Advertising Channels

Chaofan Yang¹, Li Li², Ruibo Yao^{3}*

^{1,2}School of Economics and Management, Nanjing University of Science and Technology, Nanjing, 210094, China

³Focus Technology Co., Ltd., 210061

Abstract: With the diversification of paid search advertising channels, e-commerce enterprises are paying more and more attention on how to evaluate the effectiveness of different paid search advertising channels correctly and accurately to choose the optimal advertising channel or channels. We develop a multivariate time series model to investigate the spillover effect of paid search advertising channels based on the ad click-through rate and conversion rate, and calibrate the model using an e-commerce site's web log data. We determine the long-term equilibrium relationship between each channel's advertisement clicks through the co-integration test and evaluate the effect of short-term fluctuations in the interaction between each channel advertisement clicks through the vector error correction model. Based on the empirical results, this paper puts forward suggestions on the advertising strategy of this e-commerce website.

Keywords: spillover effect, paid search advertising channel, co-integration, vector error correction model, time series model

1. INTRODUCTION

With the development of search engines, paid search advertising, which is an important economic basis for search engines, is getting more and more attention. By the end of the first half of 2017, paid search advertising accounted for 26.35% of all online advertisements in our country, ranking the second largest form of internet advertising. Paid search ads can offer more opportunities for businesses to promote their own websites, which make paid search ads have become one of the main venues where companies competing for customers^[1]. Therefore, how to improve advertising effectiveness and stimulate more consumers to generate clicks is becoming the focus of every e-commerce enterprise.

Paid search advertising is one of the most successful business patterns of Internet advertising. Generally, the enterprises purchase the specific keywords related to its business, when consumer use the keyword search for products or services, search engine will display ads involved with the keywords the company has bought in the sponsored link module of search results page^[2]. Different companies bidding on keywords, mainly based on bidding keyword technology, rank these advertising links. In order to provide useful advice to businesses on the choice of keywords and improve the enterprises economic benefits by increasing the audience of advertisements, most of the current researches on how to improve the effectiveness of paid search ads focus on how to optimize the keywords of paid search ads. In addition, there are also some studies focusing on the impact of the design of advertising content on the purchase conversion effect. Nevertheless, there are few studies on paid search advertising channels. Moreover, the research for paid search advertising channels are more concentrated in evaluating the effectiveness of advertising channel by measuring the immediate effects of advertising in a specific search channel.

With the continuous expansion of the search engine market, businesses are increasingly becoming more selective in choosing channels to run paid search advertising. For example, enterprises can bid on keyword search ads on multiple search engines (e.g. Baidu, 360 etc.) by evaluating the immediate effect of specific advertising channels (the impact of Paid Search Advertising on consumer's current ad click or conversion),

* Corresponding author. Email: lily691111@126.com(Li Li)

making the enterprises have possibility of choosing wrong advertising channels. Considering the immediate effect reflects purchase conversion resulting from the recently click on an ad, yet consumers will not generate click on ads or purchase conversion through just one search (the search channel does not have an immediate effect). Under the environment of multiple search channels, consumers are usually not limited to one single search channel, and the ad click behavior of consumers in a particular search channel may have an impact on their subsequent search or advertisement click behavior. The interaction between paid search ads channels is the spillover effect of paid search advertising channels. This spillover effect is particularly important for enterprises to choose the right channel or channel combination to improve their advertising effectiveness. Therefore, we develop a multivariate time series model to investigate the spillover effect of paid search advertising channels according to both ad click-through rate and conversion rate based on the actual web log data from an e-commerce website. Firstly, we determine the long-term equilibrium relationship between each channel's advertisement clicks through the co-integration test. Secondly, we evaluate the effect of short-term fluctuations in the interaction between each channel advertisement clicks through the vector error correction model. Finally, we use the impulse response functions to reveal the spillover effect between paid search advertising channels intuitively. Based on the empirical results, this paper puts forward suggestions on the advertising strategy of this e-commerce website.

The remainder of this article is organized as follows. In Chapter 2, we mainly review the relevant literature. In Chapter 3, we mainly present the data, including the data extraction process and the main features of the data. In Chapter 4, we will present our modeling methodology and analyze the modeling results. In Chapter 5, we will do summary, discusses the shortcomings of our study, and promote the direction of future research.

2. LITERATURE REVIEW

The literature review mainly introduces two aspects, which provides a theoretical basis for our study. On the one hand, we will introduce the researches on paid search advertising; on the other hand, we will introduce the spillover effect and the application of spillover effect research.

2.1 Paid search advertising

Paid search advertising is a service offered by Internet search engines, based on keyword bidding technology. When a user searches for those keywords, the corresponding ads will be displayed in the sponsored section of the search results page. If companies do not select the right keywords, they may show the product to the improper consumers, eventually leading to poor ad performance [3]. Therefore leveraging bidding keywords to achieve the match between paid search advertising and search terms of consumers is very important for companies. Numerous research efforts has been devoted to optimizing keyword suggestions to improve the effectiveness of paid search advertising [4, 5]. Massoudi et al. selected the related terms co-occurring with the original query to augment the query keywords through the conditional probabilities [6]. Sarmiento et al. through analyzed the correlation between query content and click-link behavior, got the keywords that are directly related to the business value of the enterprise [7]. In addition, there are some studies discussing how to improve the ads' click-through rate and ads' conversion rate based on the advertising content design [8]. Such as Jansen et al. proved content description and ad title are the most important part of evaluating paid search ads [9]. Rutz, Trusov et al. believed that the brightness and density of paid search advertising content have a significant impact on effectiveness of advertising [5].

We can find that the current researches on paid search advertising mainly focus on the keyword suggestion method and advertisement content design, while the researches on the paid search-advertising channel (search engine) are scarce. At present, the search channels for advertising on the Chinese market mainly include Baidu

search engine, Sougou search engine, 360-search engine, Ali cloud search and so on. With the optimization of keyword suggestion method and bidding mechanism, the cost of paid search advertising has increased. Meantime, the emergence of the diversified channels of paid search advertising have brought some challenges to enterprises on how to choose suitable advertising channels, making companies pay more attention to how to choose advertising channels or channel combinations to maximize the advertising effectiveness. Therefore, our study attempts to analyze the interaction between different paid search advertising channels and provide evidence for enterprises to choose the appropriate paid search advertising channels or channel combinations. The sample data for our study came from an insurance e-commerce website. The mainly paid search advertising channels of this enterprise are Baidu, Sougou and 360 search. We will primarily model and analyze the benefits of these three paid search advertising channels.

2.2 Spillover effect

The spillover means overfill and flow-out. In economics, spillover specifically represents activity externality. The concept of spillover is the effect of one agent on the welfare of other agents, originally proposed by Marshall in his book *Principles of Economics*. Stiglitz argued that the additional costs and benefits that are not included in the market transactions are spillover effects. This laid the theoretical foundation for the study of spillover effects, after which scholars began to pay attention to the spillover effect on economic growth and productivity. The researches on spillover effect mainly focused on the risk assessment or price changes in the financial field. For example, Farooq et al. analyzed the conduction mechanism of the volatility period of the oil price in the U.S. and Gulf stock markets based on the multivariate GARCH model ^[10]. Diebold and Yiimaz studied the two-way financial market spillover through the VAR framework ^[11].

With the development of online advertising, scholars have introduced spillover effects into the field of online advertising research. Naik and Peters proposed a hierarchical model to uncover the spillover effect between offline and online ads ^[12]. Based on the performance of specific keywords, Ghose and Yang explored the spillover effect between natural search and paid search ^[13]. Rutz and Bucklin proposed a dynamic linear model to study the potential spillover effect of paid search from generic keywords to brand keywords ^[14]. Florian Nottorf et al. further extended the method of Rutz et al. and use DLM model to study the spillover effects in cross-industry paid search advertising ^[15]. Pavel Kireyev et al. modeling the spillover effect between display and paid search ads, based on multivariable time series model ^[16].

We can find that the existing researches on the spillover effect of search ads mainly focused on specific search engines, lacking of in-depth exploration on multiple search engines (search-advertising channels). However, in fact, consumers will not generate click on ads or purchase conversion through just one search. Assuming that a consumer searches for a particular product keyword at a time, he may not react to the ads (ad click or ad purchase conversion) of this channel immediately. He will make a purchase after a series of searches at subsequent times. In the subsequent search process, the search channels used by consumers may change due to the search environment or personal preferences. Based on the click-through behavior of consumers on multiple paid search advertising channels, this paper explores the interactive effect of the effectiveness of paid search advertising channels, namely, the spillover effect. Considering the specific business process and data characteristics of our research object, we finally chooses the multivariable time series modeling method to analyze the spillover effect between paid search advertising channels. The specific content will be introduced in later chapters.

3. DATA DESCRIPTION

We use data from an insurance e-commerce website. In the year 2016, the website provided us the annual

server log data on the website's online marketing. In the log data, the server records each click of the user to form the user's click stream data. We need to clean and convert the raw data to get the relevant data variables that describe the paid search ads and their advertising channels. Specific data extraction process is as follows.

Firstly, we remove the invalid data such as reptile data and get the data about user IP, access time, website agent, source webpage, and cookie during the process of obtaining the original data. Secondly, we do the user identification based on cookies. When the user disables the cookie, we use IP address in conjunction with the user agent to identify the user. Thirdly, we need to do the session segmentation on the user's access because those users may complete multiple business when they visit a page. In our study, we do not adopt the traditional method of session segmentation by setting the time threshold value, we segment session based on the users' purchase behavior. We determine whether the user has made a purchase, and treat each purchase as a session cut point. The click behavior before the purchase as a single session process. Then we extract variables from the process of session, and identify the user's click behavior based on the keyword contained in a specific URL. Finally, we obtain the data of paid search advertising clicks and purchases conversions completed by consumers in different search channels.

The mainly paid search advertising channels of this insurance website are Baidu search, 360 search and Sougou search engines. We aggregate our data of clicks and purchase conversions to the week level (52 weeks a year) to avoid over-parameterization. Table 1 presents the specific meaning of the variables used in this paper.

Table 1. Variable description

Variable	Description (weekly)
Baidu_C	Baidu paid search ad clicks
Search360_C	360 search paid search ad clicks
Sougou_C	Sougou paid search ad clicks
Baidu_P	Purchase conversions through Baidu paid search ad
Search360_P	Purchase conversions through 360 search paid search ad
Sougou_P	Purchase conversions through Sougou paid search ad

Table 2 provides summary statistics of our data.

Table2. Summary statistics (per week)

Variable	Baidu_C	Search360_C	Sougou_C	Baidu_P	Search360_P	Sougou_P
Mean	6839.08	7412.00	2895.08	192.04	147.15	79.50
Median	6519.00	6854.00	2530.50	179.00	149.00	82.00
Maximum	11453.00	19660.00	7838.00	320.00	264.00	134.00
Minimum	1925.00	45.00	752.00	51.00	5.00	15.00
Std Dev	2193.97	3590.70	1326.44	69.81	65.03	28.69

4. MODELING AND ANALYSIS

Our study uses the time series modeling approach to measure the spillover effect between paid search advertising channels. The specific research process is as follows.

4.1 Modeling and parameter estimates

Persistence modeling of multivariate time series methods involves several steps.

The time series model requires that each variable be a stationary sequence to avoid the spurious regression. Therefore, we conduct Augmented-Dickey-Fuller (ADF) unit root test to determine if the variables are evolving or stationary. Table 3 summarizes the resulting statistics of the unit root tests. The result of the unit root test shows that all variables are evolving. And the result of the unit root test of variables after One Order Difference

shows that Baidu_C, Search360_C, Sougou_C and Sougou_P are integrated of order 1.

Table3. Summary of results of unit root test

Test/variable	Baidu_C	Search360_C	Sougou_C	Baidu_P	Search360_P	Sougou_P
Original sequence	-3.6568	-2.1648	-1.865	-2.2019	-3.0157	-2.0163
Differenced sequence	-5.279	-5.4368	-5.9187	-3.6893	-3.7344	-4.972

Note: Bold numbers indicate significant evidence of non-stationarity.

The co-integration relationship test is then used to determine whether the linear combination of variables has a co-integration relationship, which indicates the long-term equilibrium relationship between the two variable sequences. Co-integration test the basic steps are as follows:

(1) Confirm the single integer order of variables. Through unit root test, we know that Baidu_C, Search360_C, Sougou_C and Sougou_P are first-order integrated series. Because the co-integration test requires that the variables are integrated of same order, so the follow study excludes variables Baidu_P and Search360_P, and uses the variables Baidu_C, Search360_C, Sougou_C and Sougou_P for analysis.

(2) Estimate the long-run equilibrium relationship. We use Eq. (1) to estimate long-run equilibrium relationship.

$$Y_t = \alpha + \beta X_t + \mu_t \quad (1)$$

If the variables are co-integrated, the estimates of the co-integration coefficients α and β are obtained by regression. In order to determine if the covariates exist between the variables, we need to do unit root test for the residuals. If the residuals are stationary, then we can tell that the co-integration relationship is real. Eq. (2) represents the residual sequence model.

$$\hat{\mu}_t = Y_t - (\hat{\alpha} + \hat{\beta}X_t) \quad (2)$$

(3) Residual sequence stationary test. The residual ADF test show that the sequence is stationary, indicating that this long-term equilibrium relationship indeed exists among the variables. Eq. (3) represents the estimation of the long-run equilibrium relationship between variables.

$$\begin{bmatrix} \text{Search360_C} \\ \text{Search360_C} \\ \text{Search360_C} \\ \text{Baidu_C} \\ \text{Baidu_C} \\ \text{Baidu_C} \\ \text{Sougou_C} \\ \text{Sougou_C} \\ \text{Sougou_C} \\ \text{Sougou_P} \\ \text{Sougou_P} \\ \text{Sougou_P} \end{bmatrix} = \begin{bmatrix} -3.55 \\ 2.11 \\ 2.06 \\ 7.52 \\ 3.86 \\ 6.81 \\ -0.13 \\ 2.78 \\ 5.71 \\ -3.67 \\ 1.54 \\ -1.14 \end{bmatrix} + \begin{bmatrix} 1.39\text{Baidu_C} \\ 0.83\text{Sougou_C} \\ 1.54\text{Sougou_P} \\ 0.14\text{Search360_C} \\ 0.62\text{Sougou_C} \\ 0.48\text{Sougou_P} \\ 0.91\text{Search360_C} \\ 0.13\text{Baidu_C} \\ 0.51\text{Sougou_P} \\ 0.91\text{Search360_C} \\ 0.32\text{Baidu_C} \\ 0.67\text{Sougou_C} \end{bmatrix} + \mu_t \quad (3)$$

Based on the outcomes of the unit root and co-integration tests, we specify a vector error correction model (VEC). The general form of the VEC model with K lags is given by Eq. (4):

$$\Delta Y_t = \gamma_0 D_t + \sum_{k=1}^K \gamma_k \Delta Y_{t-k} + \alpha e_{t-1} + \mu_t \quad (4)$$

In Eq.(4), Y_t is a vector of endogenous variables at time t , D_t is a vector of deterministic components (eg, intercept, trend), e_{t-1} is a matrix of co-integration relations, γ_k and α are The parameter matrix to be estimated, Σ is the covariance matrix of the error term μ_t . The coefficients in γ_k account for the effect of past changes in endogenous variables on their current bias. The coefficients in α reflect the speed of adjustment of the endogenous variables to the equilibrium co-ordination relationship. The Bayesian Information Criterion identifies a lag-length of 3 as optimal, and the resulting VEC model is indicated in Eq. (5)

$$\begin{bmatrix} \Delta BaiduC_t \\ \Delta Search360C_t \\ \Delta SougouC_t \\ \Delta SougouP_t \end{bmatrix} = \begin{bmatrix} 1127.07 \\ 550.9 \\ -220.3 \\ 18.6 \end{bmatrix} + \begin{bmatrix} -0.5470 & 0.0982 & -0.0993 & -20.51 \\ -0.6419 & -0.0619 & -0.0061 & -12.14 \\ 0.2369 & -0.0280 & -0.5078 & -18.75 \\ -0.0018 & 0.0025 & -0.0004 & -0.6087 \end{bmatrix} \begin{bmatrix} \Delta BaiduC_{t-1} \\ \Delta Search360C_{t-1} \\ \Delta SougouC_{t-1} \\ \Delta SougouP_{t-1} \end{bmatrix} + \begin{bmatrix} -0.5327 & 0.2322 & 0.0510 & 10.5216 \\ -0.1033 & 0.0460 & 0.0242 & 1.2969 \\ 0.2624 & 0.0674 & -0.7414 & -11.0359 \\ -0.0014 & 0.0025 & -0.0004 & -0.3045 \end{bmatrix} \begin{bmatrix} BaiduC_{t-1} \\ Search360C_{t-1} \\ SougouC_{t-1} \\ SougouP_{t-1} \end{bmatrix} + \begin{bmatrix} \mu_{BaiduC,t} \\ \mu_{Search360C,t} \\ \mu_{SougouC,t} \\ \mu_{SougouP,t} \end{bmatrix} \quad (5)$$

Through the co-integration test and the vector error correction model (Eq.3 and Eq.5), we obtain the interrelationship between paid search ads clicks in three search channels:

(1) Baidu paid search ad clicks. In the short term, changes in previous Baidu paid search ad clicks have a negative impact on the current 360 paid search ad clicks(-0.6419), and a positive impact on Sougou paid search ad clicks(0.2369). In the long run, Baidu paid search ad clicks have positive impact both on 360 paid search ad clicks and Sougou paid search ad clicks. It is worth mentioning that Baidu paid search ad clicks has the greater effect on the 360 paid search ad clicks (1.39).

(2) 360 paid search ad clicks. In the short term, changes in previous 360 paid search ad clicks have a positive impact on the current Baidu paid search ad clicks(0.0982), and a negative impact on Sougou paid search ad clicks(-0.0280). In the long run, 360 paid search ad clicks have positive impact both on Baidu paid search ad clicks (0.14) and Sougou paid search ad clicks (0.91).

(3) Sougou paid search ad clicks. In the short term, changes in previous Sougou paid search ad clicks have negative impact both on Baidu paid search ad clicks (-0.0993) and 360 paid search ad clicks (-0.0061). In the long run, Sougou paid search ad clicks have positive impact both on Baidu paid search ad clicks and 360 paid search ad clicks, and the impact on 360 paid search ad clicks is greater (0.83).

4.2 Impulse response analysis

In order to reveal the spillover effect intuitively, we still use impulse response functions to analyze the impact of paid search ads clicks on different search channels, the results are as follows.

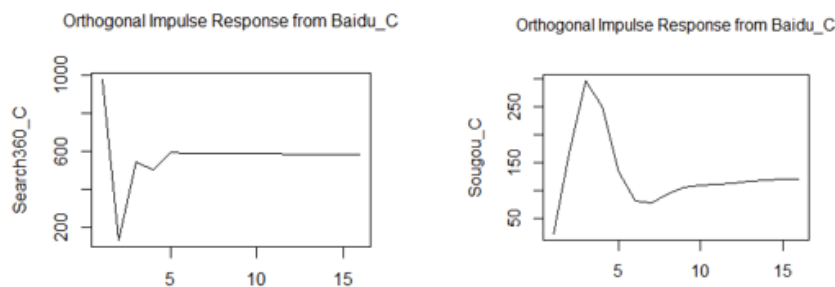


Figure1. Impact of Baidu paid search ad clicks on others

Fig. 1 shows the effect of Baidu paid search ad clicks on other two paid search ad channels are both powerful. In the short term, a shock of Baidu paid search ad clicks generates about 1000 360 paid search ad clicks immediately. After five weeks, 360 paid search ad clicks dip and then stabilize at about 600 clicks in the long run. In contrast, Baidu paid search ad clicks has a weaker effect on Sougou paid search ad clicks, but still positive. Therefore, when enterprises advertise in the 360 search engine or Sougou search engine, they can refer

to the situation of Baidu paid search ad clicks and choose the advertising channel or channel combination that can maximize the paid search advertising benefit.

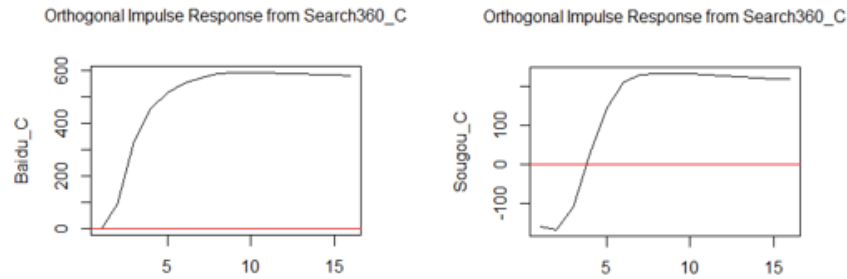


Figure2. Impact of 360 paid search ad clicks on others

Fig.2 shows the impact of 360 paid search ad clicks on other two paid search ad channels. The original shock of 360 paid search ad clicks generates few clicks of Baidu paid search ad clicks and Sougou paid search ad clicks immediately. However, Baidu paid search ad clicks increase then stabilize at about 600 clicks and Sougou paid search ad clicks increase then stabilize at about 150 clicks in the long run.

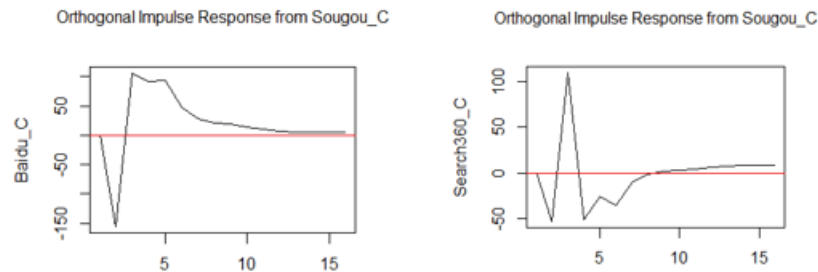


Figure3. Impact of Sougou paid search ad clicks on others

Fig.3 shows the impact of Sougou paid search ad clicks on other two paid search ad channels. In the short term, the effect of Sougou paid search ad clicks on other two paid search ad channels are both significant and fluctuating. Nevertheless, in the long run, Sougou paid search ad clicks has no significant impact on the Baidu paid search ad clicks and 360 paid search ad clicks.

In sum, the impulse response analysis suggests that Baidu paid search ad clicks may drive 360 paid search ad clicks and Sougou paid search ad clicks. With the same, 360 paid search ad clicks may drive Baidu paid search ad clicks and Sougou paid search ad clicks too. The spillover effect can be determined among three search channels in both short term and long run. Therefore, enterprise can dynamically adjust advertising channels based on business changes and advertising historical performance. When the clicks on a channel's paid search ads change, the company can adjust their investment strategies for other channels to choose different channel combinations to achieve higher return on investment.

5. CONCLUSIONS

Our study analyzed the spillover effect of different paid search advertising channels to make e-commerce websites can choose paid search advertising channels more reasonable. We developed a multivariate time series model based on the ad click-through rate and conversion rate, and calibrated the model using an e-commerce site's web logs data. We determined the long-term equilibrium relationship between each channel's ads clicks through the co-integration test and evaluated the effect of short-term fluctuations in the interaction between each channel ads clicks through the vector error correction model. Besides, we also used impulse response function to analyze the spillover effect intuitively. Through the modeling analysis, we found that the interaction between Baidu paid search ads, 360 paid search ads and Sougou paid search ads is verified, which validates the spillover

effect proposed by the research institute. Our findings can provide evidence for the company to choose the appropriate paid search advertising channels or channel combinations and have some guiding significance to this website. In the future research, we hope that on the basis of this study, we will continue to expand the relevant data such as advertising expenditures and advertising investment budgets so as to enhance the significance of the research on the actual business management. In addition, considering the heterogeneity of consumers, use data at the individual level will lead to conclusions that are more abundant.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71771122 and 71271115.

REFERENCES

- [1] A. Ghose, S. Yang, An empirical analysis of search engine advertising: sponsored search in electronic markets, *Manage. Sci.* 55 (10) (2009) 1605–1622.
- [2] Y. Chen, G. Xue, Y. Yu, Advertising keyword suggestion based on concept hierarchy, *Proceedings of the 2008 International Conference on Web Search and Data Mining*, Stanford, CA, USA, 2008, pp. 251–260.
- [3] Jansen, J. *Understanding Sponsored Search: Core Elements of Keyword Advertising*. New York: Cambridge University Press, 2011.
- [4] Z. Da, J. Engelberg, P. Gao, In search of attention, *J. Finance* 66 (5) (2011) 1461–1499.
- [5] J.B. Kim, P. Albuquerque, B.J. Bronnenberg, Mapping online consumer search, *Journal of Marketing Research* 48 (1) (2011) 13–27.
- [6] K. Massoudi, M. Tsagkias, M. de Rijke, W. Weerkamp, Incorporating query expansion and quality indicators in searching microblog posts, *Advances in Information Retrieval*, Springer, 2011, pp. 362–367.
- [7] L. Sarmento, P. Trezentos, J.P. Gonçalves, E. Oliveira, Inferring local synonyms for improving keyword suggestion in an on-line advertisement system, *Proceedings of the 3rd International Workshop on Data Mining and Audience Intelligence for Advertising* (2009) 37–45.
- [8] Haans H, Raassens N, Hout R V. Search engine advertisements: The impact of advertising statements on click-through and conversion rates[J]. *Marketing Letters*, 2013, 24(2):151-163.
- [9] Jansen, B. J. (2007). The comparative effectiveness of sponsored and non-sponsored links for Web Ecommerce queries. *ACM Transactions on the Web*, 1(1), 1–25.
- [10] Malik, F. , Hammoudeh, S. Shock and volatility transmission in the oil, US and Gulf equity markets[J]. *International Review of Economics&Finance*, 2007, 16(3): 357—368.
- [11] Diebold, F. X. , Yilmaz, K. Trans—atlantic equity volatility connectedness:U.S.and European financial institutions, 2004-2014[J]. *Journal of Financial Econometrics*, 2015, 14(1).
- [12] Naik, P.A., & Peters, K. (2009). A hierarchical marketing communications model of online and offline media synergies. *Journal of Interactive Marketing*, 23(4), 288–299.
- [13] Ghose, A., & Yang, S. (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science*, 55(10), 1605–1620.
- [14] Rutz, O.J., & Bucklin, R.E. (2011). From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research*, 48(1), 87–102.
- [15] Nottorf F, Funk B. A cross-industry analysis of the spillover effect in paid search advertising[J]. *Electronic Markets*, 2013, 23(3):205-216.
- [16] Kireyev P, Pauwels K, Gupta S. Do display ads influence search? Attribution and dynamics in online advertising [J]. *International Journal of Research in Marketing*, 2015, 33(3):475-490.

A Research on the Influential Factors of Listing Sales Based on Online Information in Short Rental Markets

Hong Peng^{1}, Shuang Tao²*

¹ School of Business Administration, Zhongnan University of Economics and Law, China

² School of Business Administration, Zhongnan University of Economics and Law, China

Abstract: Recent years, sharing economy as a new business model has developed very well in the whole world and the online marketplace for peer-to-peer accommodation rental services such as XiaoZhu change people's life a lot. With the development of social media and information technology, consumers can not only publish text comments but also can share the photos taken by them on the platform as a supplement. In order to study the impact of online information on the sales volume of houses under the background of sharing-economy, we collect data from XiaoZhu and use the correlation analysis and regression analysis. This research provides a new perspective for the sharing research and focuses on China's online short-term rental market. The result shows that the total number of online reviews, the number of picture comments and the price have significant effects on home sales.

Keywords: sharing economy, short rental markets, online information, home sales

1. INTRODUCTION

As an important branch of the shared economy, online short-term rental service has brought great changes to people's life. Compared to the standard hotel accommodation, online short-term rental services make use of idle resources to provide multiple houses to meet different needs of tenants such as economic needs and social needs^[1]. This new business model has its own unique features and strong competitiveness in recent years.

Online short-term rental service is a new concept which is different from traditional tourist accommodation industry in a shared economic environment. Traditional hotel accommodation industry has been unable to fully meet the needs of individual users while online short-term rental service fills the gap in this market. Online short-term rental industry originated in the United States and the huge success of HomeAway and Airbnb indicates a bright future of the short rental industry.

China has become an important market in global sharing economy and people are increasingly accepting the concept of sharing. XiaoZhu is a peer-to-peer platform. Levitt (1981) said that this marketplace offered intangible experienced goods and the sellers are the coproducers of the service experience^[2]. Consumers cannot judge the real quality of renting an apartment on XiaoZhu before they start using the service^[3].

Online information offers enough help to consumers when they need to make purchase decisions. They can browse the web page to get useful messages such as price and location of the house. With the development of the review system, online user reviews have become an important source of information for consumers too^[4]. They will be affected by the comments posted online about the consumption experience. Given the important impact of online user reviews on potential consumer shopping decisions, XiaoZhu's online feedback system encourages consumers to comment and communicate. Especially, they can upload actual photos to increase the authenticity and credibility of their words.

The realization of the picture comments not only enhances the dimension of the comment system, but also provides consumers with further decision-making information so that the online comments can add more value

* Corresponding author. Email: hpeng520@126.com

to potential customers^[5]. At the same time, due to the picture comments to further restore the real products and service quality of merchants, it has played a supervisory and supervisory role in the uneven quality of products and services available to online merchants.

This paper will investigate the impact of online information upon buyers' behavior use the data collected from XiaoZhu and provide useful information and practical suggestions for the short-term rental services like XiaoZhu.

2. LITERATURE REVIEW AND HYPOTHESIS

2.1 Short rental markets

The most crucial factor that motivates people to start participating in the shared economy is cost. People may consider the cost factor when they book a room on short-term rental platform and researchers have made a great of researches on this issue. Huang(2009) said that when consumers purchase products especially practical products, the price of products played a role in influencing the purchasing decision^[6]. Wang & Nicolau(2017) said that a key dimension of the hospitality industry is pricing^[7]. They identified the price determinants of sharing economy based accommodation offers in the digital marketplace.

H1: Housing price has a significant negative impact on home sales.

XiaoZhu as a typical representative of the C2C model provides an experiential product. When consumers purchase experiential service products in the network environment, the impact of information on decision-making will be even more significant^[8]. XiaoZhu allows tenants to evaluate the overall experience of the property to reflect the user satisfaction with the availability. Zhang et.al(2013) found that the scores were significantly related to future sales by researching sales information on digital cameras on Amazon^[9]. Zhang(2014), who reviewed the online data of restaurants on the mass audience, found that restaurant ratings had a significant impact on consumers' choice and consumption^[10].

H2: Housing overall grade has a significant positive impact on home sales.

2.2 Online reviews and picture comments

In order to serve consumers better and enable consumers to capture useful information within a short period of time, the network evaluation system has developed a wide range of ratings in addition to the text message function so that visitors can be more fully evaluated by the customers who purchase the products information of products. Anderson(1998) found that extremely satisfied and extremely dissatisfied customers were more likely to become spontaneous word-of-mouth communicators^[11]. Online consumer reviews are both information providers and product recommenders and offer important information to consumers when shopping online. Zervas(2015) made a research on the impact of online reviews' timidity and the number of reviews on the sales of new products from the perspective of information providers^[12]. Many scholars also studied from the attributes of comments which included usefulness, credibility, informational and persuasiveness. Mudambi and Schuff(2010) examined the factors that influence the usefulness of a comment when consumers make purchasing decisions^[13]. Based on the theory of information economics, the experimental conclusion shows that the product category can adjust the process of the extreme comments. With the development of internet technology, online comments system perfect the functions and consumers can make picture comments on the products. Herring and Dainas(2017) further concluded that simple picture comments on products and text comments have no difference on the perception of customers across different product categories^[14]. This conclusion was closely related to the selection of the picture comments and the conclusion does not have extensive practicability so the impact of image comments has yet to be researched.

H3: The number of online reviews has a significant positive impact on home sales.

H4: The number of picture comments has a significant positive impact on home sales.

H5: Sunburn rate has a significant positive impact on home sales.

3. RESERACH DESIGN

3.1 Research object

XiaoZhu was established in 2012. It is a typical C2C paltform in China which provides costomers with different humane,family-friendly,diversified and cost-effective accommodation options by displaying the idle resources of the landlord on the online platform.Various types of houses,including family apartments,special folk,holiday villas and other personalized short rental accommdation products are offered to five million active costomers in more than 300 cities across the country.

As a platform offered by information exhibition and online booking between consumers and landlords, XiaoZhu gradually improves the real-name authentication mechanism, personal credit supervision mechanism and property protection in oder to fully and effectively utilize the idle resources and provide bilateral service for platform users. Consumers can browse the homepage's basic structure, ratings and location information through the selected homepage and settled time of checking in while the detail information can be viewed by clicking on the detailed homepage^[15]. Simiarly to other online platforms, XiaoZhu provides with the information feedback system so that consumers can score accoriding to the accommodation experience on the aspects of cleanliness, safety, cost performance, description of compliance, traffic convenience and so on, and make an object evaluation. At the same time, consumers can also upload actual photos as supplementary. By online comments, cosumers can get more realistic housing information before they make the purchase decision.

3.2 Data collection

This research selected housing source data of XiaoZhu as the research object and collected houding data of five cities like Hangzhou, Shanghai, Guangzhou,Xi'an and Chengdu until November 2017. This reaserach mainly focus on the impact of sales in the aspects of the number of online reviews, the number of picture comments, sunburn rate, price and the ovearall grade.

On the paltform of XiaoZhu, there are mainly three kinds of housing resouce including complete set, single room and shared rooms while the number of shared rooms is about only fourty suits and take a tiny part. So this reserach chosse the complete set and single room these two main product type. And in XiaoZhu, a landlord may have mutiple listings but the total orders showed on the host profile did not distinguish between each room inventory. In the same period of time, multisource landlord may have much more oders coompared to those single-source landlord and may have impact on the final reserach results^[16]. The length of the registration time will also affect the sales data and the housing source may have more orders if they are earlier registered^[17].Combine the above data processing steps, there are 477 listings crawled by availabilty of Python programs and processed by the tool of SPSS.Table1 shows the description and descriptive statistics of the data.

Table 1. Variable description and descriptive statistics

Variable name	Variable description	Average	Maximum	Minimum	Std
total_order	Total number of house orders	83	996	7	100.47
num_com	Total number of online comments	32	209	4	31.96
num_pic	Total number of online picture comments	7	56	0	7.95
grade	The overall scores	4.94	5	4.5	0.10
price	Present house price	271	1288	66	147.31
sunburnn rate	The rate of picture comments on total comments	0.23	0.91	0	0.16

4. ANALYSIS

4.1 Correlation analysis

From the results of the correlation analysis, there is a significant positive correlation between the total number of comments and total orders. The correlation between the total picture comments and total orders and the correlation between house price and total orders are also show significant positive. Overall score and baskets rate do not have significant correlation with total orders. The results are shown as the following Table2.

Table 2. Correlation analysis

	total_order	num_com	num_pic	grade	price	sunburn rate
total_order	1					
num_com	0.781**	1				
num_pic	0.504**	0.592**	1			
grade	-0.045	-0.015	0.431*	1		
price	0.052	0.006	0.110	0.148*	1	
sunburn rate	-0.205**	0.007	0.082	0.122*	-0.045	1

4.2 Regression analysis

In order to further explore the impact on house selling, we use regression test to check the relationship among data. Before the regression analysis, we first conducted a multi-collinearity test on the variables and the results show that the VIF values (variance expansion factors) of all the variables are far less than 10, and there is no multicollinearity in the model. The established regression model is suitable and effective. Regression analysis results are shown in the following Table3:

Table3. Regression analysis

Variable	coefficient	Standard deviation	T value	Sig	VIF
num_com	0.463	0.058	5.357	0.000	2.926
num_pic	0.326	0.043	3.554	0.000	3.296
grade	-0.140	0.066	-1.203	0.129	1.578
price	-0.045	0.068	-0.870	0.386	1.047
sunburn rate	-0.062	0.041	-2.975	0.018	1.037

The number of comments ($\beta = 0.463$, $p = 0.000$) and the number of picture comments ($\beta = 0.326$, $p = 0.000$) have a significant impact on the sales volume, that is, the larger the the number of comments and the number of picture comments, landlords get much more house orders. This results support the hypotheses H3 and H4. The house price ($\beta = -0.062$, $p = 0.018$) also have a significant negative impact on the sales volume. That is, the higher the price of the house, the fewer orders the landlord can get, which is consistent with the law of actual market economy. The price and demand in a negative correlation so that the hypothesis H1 is supported. From the view of sunburn rate ($\beta = -0.140$, $p > 0.05$), sunburn rate which means that the rate of picture comments in total online comments does not have a significant impact on sales growth so that the hypothesis H5 is not supported. Consumers may pay more attention to the number of comments and wonder is there any picture comments about products but they have less attention to the proportion of the number of picture comments. While the overall grade ($\beta = -0.045$, $p > 0.05$) did not pass the significance test so that hypothesis H2 is not supported. Through the analysis of actual data, consumers can upload their evaluation in the form of grades according to the consumption experience. The higher the score means the better the consumption experience, but in the reality, consumers prefer give great grades. Regarding the research data, the average of score is 4.94 and the minimum is 4.5 while the full marks is 5. This shows that on the platform of XiaoZhu, consumers' experience scores do

not make a big difference and in this condition, other consumers can not rely solely on scoring data to make purchasing decisions.

5. GENERAL DISCUSSION

This article chooses XiaoZhu as the research object which is the represent of short-rent sharing platform in China to explore the impact of online information such as reviews on the sales volume of houses. Research shows that the review data has the most significant impact. Housing orders will be much more if the total numbers of comments and picture comments are higher. At the same time, the price will also have an impact on sales while the overall score does not significantly affect it.

The specific discussion and guidance on the management of accomodation and the management of Internet platforms can be explored from the following three aspects:

Firstly, platforms and merchants need to place a high value on the number of online reviews in order to attract potential users. The number of reviews is the most intuitive indicator of the popularity of the product and the greater the cumulative number of reviews, the more popular among customers. Previous studies also show that the cumulative amount of comments on products have a positive impact on its' sales. As a rapidly developing new industry in sharing economy, XiaoZhu should pay more attention to consumer comment mechanism and provide potential users with more realistic and valuable information. From the perspective of the landlord, they should pay more attention to communication with users and encourage users to timely upload housing evaluation. On the one hand, this can solve the existing problems in housing more intensively, on the another hand, this can attract more potential users to increase the housing income.

Secondly, the results of this study shows that the picture comments have a great influence on the sales of houses. Users can upload the actual taken photos at the same time when they make the comments and grade the service so that they can give more vivid description of the literal expression. Potential users can make the judgement between picture comments and description of the house from owners and get more highly trust on the housing resource. Therefore, platforms and landlords should encourage consumers to upload reviews while upload more actual taken photos to increase the rate of unboxing to attract potential users.

Thirdly, housing price is also an important part of consideration when consumers choose short-term rental house. Previous studies show that travelers' participation in accommodation is mainly due to cost savings(economic attractiveness) and longing for local social relations(social attractive). According to the data of XiaoZhu, users are more sensitive to the price and when they have to make choices, they prefer to consider the impact of housing price. Therefore, landlord should pay more attention to market feedback when they make pricing decisions and provide more attractive listing price on the foundation of high quality to expand consumer groups.

6. CONCLUSIONS

This article did not select the more familiar and more mature representative in online short-term rental industry like Airbnb but combined with the actual situation of China's domestic economy and the development of online short-term rental industry to choose XiaoZhu whose number of consumers and the condition of deal sum are on the top lines in China. The result shows that the total number of online reviews, the number of picture comments and the price have significant effects on home sales. This research make up for the blank of domestic short-term online industry to help researchers have deeper understanding of Chinese consumers' consumption choices. But the downside is that in the process of capturing and using panel data, only the data of several cities are selected and the reserach data of urban data of other more prosperous regions are lacking. This article only studies the impact of sales volume on the online short-term rental market form total

number of reviews, picture comments and onther quantitiative aspects, in the futher studies should consider the in-depth impact of image quality and graphic description consistency on sales volume.

REFERENCES

- [1] Edelman BG, Luca M.(2014) .Digital Discrimination: The Case of Airbnb.com. Harvard Business School Working,60.
- [2] Levitt, T. (1981). Marketing intangible products and product intangibles. Harvard Business Review, May-June,123.
- [3] Fagerström A, Pawar S, Sigurdsson V, et al. (2017).That personal profile image might jeopardize your rental opportunity! On the relative impact of the seller's facial expressions upon buying behavior on Airbnb. Computers in Human Behavior,72:123-131.
- [4] Duan W, Gu B.,(2008).Whinston A B. Do Online Reviews Matter? – An Empirical Investigation of Panel Data. Decision Support Systems, 45(4):1007-1016.
- [5] Mudambi S M, Schuff D.(2010).What makes a helpful online review? a study of customer reviews on amazon.com. Society for Information Management and The Management Information Systems Research Center,20.
- [6] Huang P, Lurie N H, Mitra S.(2009). Searching for Experience on the Web: An Empirical Examination of Consumer Behavior for Search and Experience Goods. Journal of Marketing, 73(2):55-69.
- [7] Wang D, Nicolau J L. (2017).Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. International Journal of Hospitality Management, 62:120-131.
- [8] Li Huiying.(2013).Impact of Online Comments on Consumer Perception and Sales of Corporate Products . Harbin Institute of Technology.(in Chinese)
- [9] Zhang L, Ma B.,(2013). Cartwright D K. The impact of online user reviews on cameras sales[J]. European Journal of Marketing, 47(7):1115-1128.
- [10] Zhang H.Y, Zhou R.T, Yan,H.(2014). The Influence of Internet Word of Mouth on Consumers' Online Behavior. Management World, (3):178-179.(in Chinese)
- [11] Anderson E. (1998).Customer Satisfaction and Word of Mouth[J]. Journal of Service Research, 1(1):5-17.
- [12] Zervas G, Proserpio D, Byers J. (2015).A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average. Social Science Electronic Publishing,197.
- [13] Mudambi S M, Schuff D. (2010).What makes a helpful online review? a study of customer reviews on amazon.com. Society for Information Management and The Management Information Systems Research Center,233.
- [14] Herring S, Dainas A.(2017). 'Nice Picture Comment!' Graphicons in Facebook Comment Threads Hawaii International Conference on System Sciences,39.
- [15] Mingming Cheng.(2016). Sharing economy: A review and agenda for future research. International Journal of Hospitality Management, 57:60-70.
- [16] Stors N, Kagermeier A. (2017).Airbnb-hosts as players in new urban tourism: motivation in participating from the supply-side perspective. Geographische Zeitschrift,24-29.
- [17] Hammer B.(2014). A Picture is Worth a Thousand Words, but Are They the Words That Matter? -- An Analysis of the Effects of Image Comments on Social Networking Sites. Dissertations & Theses - Gradworks, 33(5):873-888.

Dual-Channel Supply Chain Network Equilibrium Model with Consumer-Driven

Jun Ma^{a,b}, Ding Zhang^c, Ding Feng^d, Yiliu Tu^{a,*}

a. Department of Mechanical and Manufacturing Engineering, University of Calgary, Calgary, AB, T2N 1N4, Canada

b. School of Management, Shenyang University of Technology, Shenyang, LN, 110870, China

c. School of Business, State University of New York, Oswego, NY, 13126, USA

d. School of Mechanical Engineering, Yangtze University, Jingzhou, HB, 434020, China

Abstract: In this paper, we study designing and managing effective dual-channel supply chain network equilibrium model to optimize the profit of each node in dual-channel supply chain network and satisfy seamlessly customer demand. The customer demand in each channel is driven by the heterogeneous consumer characteristic attributes. In our proposed model, Multinomial Logit (MNL) function is used to make a purchase decision for customers considering selling price, operation time and retail services. Furthermore, the Variational Inequality is used to express the equilibrium solution in dual-channel supply chain network. A numerical example in dual-channel supply chain network is presented to show the MNL function can be a good replacement for the demand function when customers are heterogeneous and the proposed model can be helpful to avoid time trap.

Keywords: Dual-channel supply chain network, Network Equilibrium Model, Variational Inequalities, Modified Projection Method

1. INTRODUCTION

Avoiding loss affected by “channel conflict” in dual-channel supply chain (SC) and dual-channel supply chain network (SCN) is challenging for firms who need to optimize selling price, processing time and service in dual-channel markets with heterogeneous consumer characteristic attributes. The operations of dual-channel SCN not only affect people’s lives, but also the firms’ profitability in SCN. Some manufacturers even have switched into single-channel SC from dual-channel SC in order to avoid this kind of channel conflict^[1], while others have tried to implement differentiation strategy in dual-channel SC.

Dual-channel SC has increasingly attracted interest in recent literature. Pricing strategy in dual-channel SC is one of the key issues that previous research focus on. Considering channel control, Chiang et al.^[2] construct a price-setting game between a manufacturer and its independent retailer. Chun and Kim^[3] analyze the price differences between a retailer channel and a direct online channel. Chen et al.^[4] examine a manufacturer’s pricing strategies in a dual-channel SC and illustrate how a contract enables both the manufacturer and the retailer to be a win-win. Moreover, pricing strategies model of dual-channel SC has been developed for various characteristics and types SC, including risk management^[5], retail service^[6], price discounts^[7], brand management^[8], information sharing^[9], inventory management^[10], lead time management^{[11],[12]}, and quality management^[13]. However, the aforementioned papers just considered the pricing issue and did not address the importance of customer choice based on the heterogeneous characteristic attributes. In our model, different choices from heterogenous customers in the characteristics of products is the driving force of forming dual-channel SC.

As noted in Nagurney et al.^[14], time plays a critical role in time-sensitive products SC. Time is a

fundamental element in dual-channel SC which customers care. In many papers, operation time is an significant factor that impact product quality^{[15],[16]}, waste of products^[17] and lost value^[18]. These studies focused on the importance of operation time in SC and the benefit followed operation time control. For the various type of products in our model, customers care different operation time combinations, including production time, delivery time and retail time in practice.

Our study builds upon earlier SCN equilibrium research and makes some new contributions. First, we develop a network equilibrium model for dual-channel SCN. On the one hand, because most of the models focused on intra-SC competition, there is little work that studies the competition among dual-channel SCs. On the other hand, competition is no longer between stand-alone companies, but rather SC against SC^[19]. The strategies in a dual-channel SC were impacted by competing and cooperating SC. Second, selling prices, operation time and retail service are used as the key characteristics of products in dual-channel SCN. Consumer choices on selling price, operation time and retail service are driven by the consumer characteristic attributes. By accounting for consumer-driven time and service competition, we can address a deeper understanding of time and service competition in dual-channel SCN, which allows managers to evaluate the time and service competition decision from consumer's perspective in dual-channel SCN and implement a discrete choice model to support their decentralized decisions. Next, ignoring consumers' preferences and needs, Stalk and Webber^[20] reported many firms found themselves caught in a time trap. MNL function, as a discrete choice model, is developed to understand consumer purchase decisions among selling price, operation time and retail service based on the consumer characteristic attributes in dual-channel SCN. Finally, shoes were used as a representative example in dual-channel SCN.

2. DUAL-CHANNEL SCN EQUILIBRIUM MODEL

In this section, we develop a dual-channel SC network equilibrium model for manufacturers, retailers and customers in markets including the costs associated with the production, transportation, marketing, storage, and retail service. The decision-makers in the SCN, specifically the manufacturers and retailers, are assumed to seek profit maximization considering the operation time and retail service. In view of consumer characteristic attributes among price time and retail service, the consumers in the market make their own decisions in accordance with their preference and utility. A typical dual-channel SC structure is shown in Figure 1.

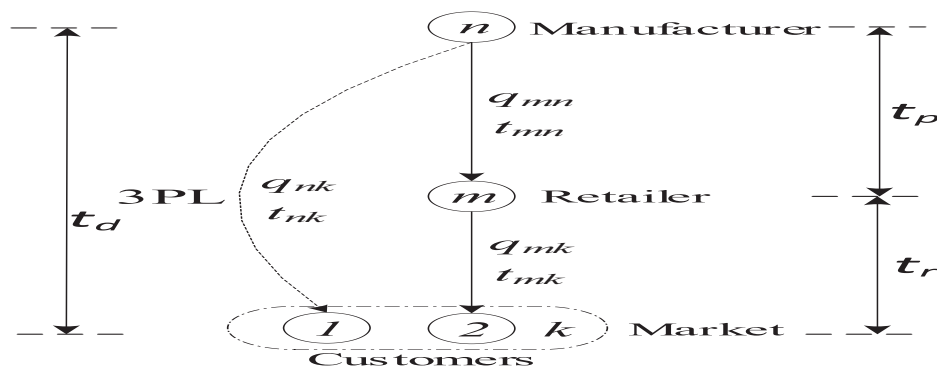


Figure 1. A typical dual-channel SC structure.

More than two dual-channel SCs form a dual-channel SCN. We designate N product manufacturers, a typical manufacturer is called n who transports its products to K demand markets by direct channel, a typical market is called k (Figure 2). In the traditional channel, manufacturer n will sell and delivery part of products to retailer m at first. Then, retailer m will sell the products to demand markets in the end and provide additional retail service for customers using traditional channel. In this scenario, according to consumer preference between selling price, time and retail service, there are two options for consumers to select: direct channel and traditional channel.

Consider a general network with time attribute $H = [G, L, T]$, where G denotes the set of nodes in the network, L denotes the set of directed links and T denotes the set of operation time on directed links. The links between node n and node m can be production and transaction from manufacturer n to retailer m . The links between node m and node k can be the transaction from retailer m to market k . The links between node n and node k can be the production and transaction from manufacturer n to market k by direct channel or 3PL service.

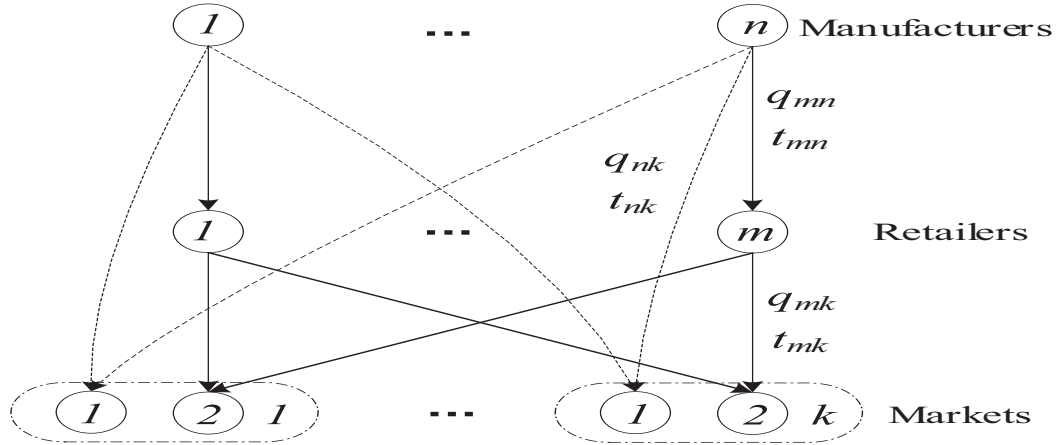


Figure 2. A typical dual-channel SCN structure.

In this paper, we assume the process of production, transportation and marketing is assumed to be acyclic. We suppose that the firms in SC compete and cooperate in a selling season. Different from the definition of a general operation time on links directly, similar to that in^[21], we divide the operation time in dual-channel SC into 3 types, (t_{nm}, t_{mk}, t_{nk}) according to customer preference and product characteristics. The operation time (t_{nm}) between manufacturers and retailers includes production time, storage time and delivery time so on. The operation time (t_{mk}) between retailers and demand markets includes storage time and delivery time so on. The operation time (t_{nk}) between manufacturers and demand markets includes production time, storage time and delivery time so on. Operation time in dual-channel SC is a key factor that customers care. For example, In a fashion clothing or a book dual-channel SCN, customers balance between t_{nk} and t_{mk} . In a customized products dual-channel SCN, customers trade-off between t_{nk} and $t_{nm} + t_{mk}$. Usually the operation time T between a manufacturer and a marketplace is easy for consumers to get from the external packing of part products in supermarkets or to percept after making an order.

Definition I: The variational inequality problem

The variational inequality problem. The finite-dimensional variational inequality problem, $VIP(U, K)$, is to determine a vector, $x^* \in \Omega \subset R^n$, such that

$$\langle \nabla U_s(X^*), (X_s - X_s^*) \rangle \geq 0, \forall X \in \Omega. \quad (1)$$

Where $\Omega = \left\{ x_{a_{ij}} \geq 0, p_{ij} \geq 0: N_k \cdot Pr_j(\vec{V}_j) = \sum_{s \in S} x_{s_{ij}} \right\}$ is the feasible set of chain variables for the SCN in our model. U is a given continuous function from K to R^n , K is a given closed convex set, and $\langle \cdot, \cdot \rangle$ denotes the inner product in n dimensional Euclidean space. X^* is a network economic equilibrium if and only if X^* satisfies (1)^[22].

2.1 The equilibrium conditions for manufacturers

The transaction between manufacturer n and retailer m is a solid arrow line in Fig. 2. The transaction

between manufacturer n and demand market k is denoted by dotted arrow lines in Fig. 2. The production output of manufacturer n satisfies the conservation of product flow equation (2). The production output of manufacturer n is equal to the sum of the quantities sent from manufacturer n to all retailers and all customers through all delivery methods. Here we assume that each manufacturer has a production cost function f_n .

$$Q_n = \sum_{m=1}^M q_{nm} + \sum_{k=1}^K q_{nk} \quad (2)$$

$$f_n = f_n(Q_n), \forall n. \quad (3)$$

The transaction cost between manufacturer n and retailer m , as well as the transaction cost between manufacturer n and consumers in market k , include the transportation cost and insurance cost which depends on the volume of product flow between the two nodes. The transaction costs are given by:

$$c_n = c_{nm}(q_{nm}) + c_{nk}(q_{nk}), \forall n, m, k. \quad (4)$$

The operation time t_{nm} between manufacturer n and retailer m , including production time and transportation time are related to the relevant transaction cost. In our model, manufacturer n provides w operation time options associated with different cost functions.

$$t_{nm} \rightarrow c_{nm}, \forall n, m. \quad (5)$$

Every manufacturer goes for profit maximization. So, the profit of manufacturer n is equal to the price p_{nm}^1 that manufacturer n charges for the product multiplied by the volume of product shipping to all retailers, with the addition of the price p_{nk}^1 that manufacturer n charges for the products multiplied by the volume of product shipping to all consumers in markets, minus the production cost and transaction cost. The function of profit maximization for manufacturer n can be expressed as:

$$\max \sum_{m=1}^M p_{nm}^1 \cdot q_{nm} + \sum_{k=1}^K p_{nk}^1 \cdot q_{nk} - f_n(Q_n) - \sum_{m=1}^M c_{nm}(q_{nm}) - \sum_{k=1}^K c_{nk}(q_{nk}) \quad (6)$$

$$s. t. q_{nm} \geq 0, q_{nk} \geq 0; \forall n, m, k. \quad (7)$$

2.2 The equilibrium conditions for retailers

The transaction cost between retailer m and customers in market k may include transportation, display, and storage costs associated with these products. There is only a delivery way between retailers and customers in markets, which is a solid arrow line in Figure 2. Let c_{mk} and c_{mk}^s respectively denote a transaction cost function and retail service cost between retailers m and customers in market k . c_{mk} is a function of q_{mk} and c_{mk}^s is a function of retail service s_{mk} . The transaction cost function and retail service cost function, then, can be expressed as:

$$c_{mk} = c_{mk}(q_{mk}), \forall m, k. \quad (8)$$

$$c_{mk}^s = c_{mk}^s(s_{mk}), \forall m, k. \quad (9)$$

The operation time t_{mk} between retailers m and customers in market k is related to the relevant transaction cost, too. Assuming that every retailer is a profit-maximizer, the profit of retailer m is equal to the price p_{mk}^2 that retailer m charges for the products multiplied by the volume of the products shipping to all customers, minus the transaction cost and retail service cost. The function of profit maximization for retailer m can be expressed as:

$$\max \sum_{k=1}^K p_{mk}^2 \cdot q_{mk} - \sum_{k=1}^K c_{mk}(q_{mk}) - \sum_{k=1}^K c_{mk}^s(s_{mk}) - \sum_{n=1}^N p_{nm}^1 \cdot q_{nm} \quad (10)$$

$$s. t. \sum_{n=1}^N q_{nm} \geq \sum_{k=1}^K q_{mk}, \forall n, m, k. \quad (11)$$

2.3 The behavior of consumers in markets

Suppose that there are j products in dual-channel SCN. Let p_k^{2j} denote the price of product j on market k . We assume that the prices of all market K are transparent for all consumers, because the Internet has changed the way to obtain price information^[23]. In point of price, the consumers take the price charged by a retailer p_k^2 , or the price charged by a manufacturer p_k^1 , into consideration. The price will not exceed the price that the consumers in markets are willing to pay for the product. The equilibrium conditions in markets take the following forms:

For all manufacturers $n; n = 1, 2, \dots, N$:

$$\left. \begin{aligned} p_k^{1j*} &\leq \xi_k^{2j*}, \text{ if } q_{nk}^{j*} > 0 \\ p_k^{1j*} &> \xi_k^{2j*}, \text{ if } q_{nk}^{j*} = 0 \end{aligned} \right\} \quad (12a)$$

For all retailers $m; m = 1, 2, \dots, M$:

$$\left. \begin{aligned} p_k^{2j*} &\leq \xi_k^{2j*}, \text{ if } q_{mk}^{j*} > 0 \\ p_k^{2j*} &> \xi_k^{2j*}, \text{ if } q_{mk}^{j*} = 0 \end{aligned} \right\} \quad (12b)$$

Customers in our model are assumed to be heterogeneous. Some consumers care more about operation time or retail service of goods than price, others vice versa. Suppose that customers are statistically identical and independent. In our study, we view the same product with distinctive characteristics attribute as variants. Let Pr_{ij} denote the probability of the consumers choosing variant $i: i = 1, \dots, I$. The utility function that customers choose variant i is a linear random utility model (LRUM)^[24].

$$U_j(i) = \vec{\theta}_j \cdot \vec{V}_j + \varepsilon_j, \forall i, j \in J. \quad (13)$$

Where $\vec{V}_j(i)$ is the product's characteristics attributes vector, including consumer's valuation, selling price, operation time and retail service of variant, denoted respectively by C, P, T, S . In this expression $\varepsilon_{j,w}$ is called idiosyncratic taste differences of consumers and uncertainty which cannot be observed. $\varepsilon_{j,w}$ are *i.i.d.* and follow a double exponential distribution with mean zero. The exponential, the double exponential and the Gumbel distributions were verified well to satisfy the requirements^{[25],[26]}.

Let $\vec{\theta}_j(\alpha, \beta, \gamma)$ denote the customer's sensitivity coefficients that are nonnegative absolutely continuous random variables expressing the consumer's preference^[24], including price sensitivity coefficient, time sensitivity coefficient and retail service sensitivity coefficient computed by using maximum-likelihood estimation (MLE), respectively. The probability of choosing alternative (i, j) out of a total alternative of is:

$$Pr_{ij}(\vec{V}_j) = \frac{e^{(U_j(i))}}{1 + \sum_{i=1}^I e^{(U_j(i))}}, i = 1, \dots, I; j = 1, \dots, J. \quad (14)$$

The probability of a consumer rejecting all variants options is $1 - Pr_{ij}(\vec{V}_j)$. If $\vec{\theta}_j$ are estimated according to customer choices in i variants, the likelihood (\mathcal{L}) is given as

$$\mathcal{L}(\vec{\theta}_j) = \left[1 - \sum_{i=1}^I Pr_{ij}(\vec{V}_j) \right]^{N_0} \prod_{i=1}^I Pr_{ij}(\vec{V}_j)^{N_i}, \quad (15)$$

Then we solve the optimal $\vec{\theta}_j$ for $j \in [1, \dots, J]$ which maximizes $\ln \mathcal{L}(\vec{\theta}_j)$.

If the equilibrium price the consumers are willing to pay is positive, the shipments from retailers and manufacturers must be equal to the demand for the products of consumers.

$$N_k \cdot Pr_j(\vec{V}_j) \begin{cases} = \sum_{m=1}^M q_{mk}^* + \sum_{n=1}^N q_{nk}^*, \text{ if } \xi_k^{2(j)*} > 0 \\ < \sum_{m=1}^M q_{mk}^* + \sum_{n=1}^N q_{nk}^*, \text{ if } \xi_k^{2(j)*} = 0 \end{cases} \quad (16)$$

2.4 The equilibrium conditions of the dual-channel SCN

The manufacturers and retailers are assumed to compete in a noncooperative fashion. It is also assumed that the production cost functions and the transaction cost functions for each manufacturer and retailer are continuous and convex. The optimization function (6) for all manufacturers, the optimization function (10) subject to (11) for all retailers and functions (12), (16) can be expressed as the following variational inequality^{[26],[28]-[31],[29],[30],[31]}.

$$\sum_{n=1}^N \left[\frac{\partial f_n(q_n^*)}{\partial q_n} \right] \times [q_n - q_n^*] + \sum_{n=1}^N \sum_{m=1}^M \left[\frac{\partial c_{nm}(q_{nm}^*)}{\partial q_{nm}} - \xi_m^1 \right] \times [q_{nm} - q_{nm}^*] + \sum_{n=1}^N \sum_{k=1}^K \left[\frac{\partial c_{nk}(q_{nk}^*)}{\partial q_{nk}} - \xi_k^2 \right] \times [q_{nk} - q_{nk}^*] + \sum_{m=1}^M \sum_{k=1}^K \left[\frac{\partial c_{mk}(q_{mk}^*)}{\partial q_{mk}} + \xi_m^1 - \xi_k^2 \right] \times [q_{mk} - q_{mk}^*] + \sum_{m=1}^M \sum_{k=1}^K \left[\frac{\partial c_{mk}^s(s_{mk}^*)}{\partial s_{mk}} \right] \times [s_{mk} - s_{mk}^*] + \sum_{m=1}^M \left[\sum_{k=1}^K q_{mk} - \sum_{n=1}^N q_{nm} \right] \times [\xi_m^1 - \xi_m^{1*}] + \sum_{k=1}^K \left[\sum_{m=1}^M q_{mk} - N_k \cdot Pr_j(\vec{V}_j) \right] \times [\xi_k^2 - \xi_k^{2*}] + \sum_{k=1}^K \left[\sum_{n=1}^N q_{nk} - N_k \cdot Pr_j(\vec{V}_j) \right] \times [\xi_k^2 - \xi_k^{2*}] \geq 0 \quad \forall q, \xi^1, \xi^2 \in R_+^{NM+MK+NKJ}. \quad (17)$$

The term ξ_m^1 is the Lagrange multiplier associated with constraint (11) for retailer m . The term ξ_k^2 is the Lagrange multiplier for customers in market k . The production quantity and shipments that the manufacturers send to the retailers must be equal to the production quantity and shipments that the retailers accept from the manufacturers. The amount of product purchased by the customers must be equal to the shipments that the customers accept from the retailers and manufacturers. We state the equilibrium conditions of the whole dual-channel SCN as a variational inequality formulation.

3. ALGORITHM AND NUMERICAL EXAMPLE

In this section, the algorithm that we use for the computation of the product equilibrium pattern satisfying variational inequality (17) is the Projection Method^{[32],[31]}. Then we present one example of dual-channel SCN in disparate geographic areas, specifically, SCN for shoes.

The example is developed for shoes dual-channel SCN. There are two retailers in two markets who purchase shoes shipped from two manufacturers. Two manufacturers ship shoes to these retailers and customers in two demand markets, in which the populations are $N_1 = N_2 = 10000$, respectively (Figure 3). The two manufacturers are designed to offer two different strategies to satisfy heterogeneous customers with different requirements: online direct channel and traditional retail channel.

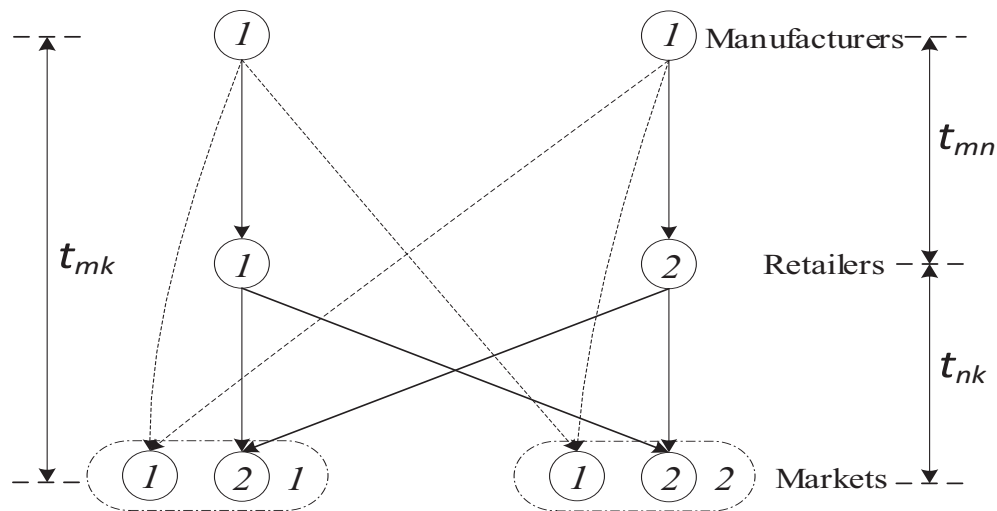


Figure 3. The shoe SCN structure.

The relative data of example are given in table 1. The retailer service cost function is given: $c_{mk}^s =$

$\frac{1}{2}\eta s_{mk}^2, \forall m, k^{[33]}$. The algorithm is implemented in MATLAB R2017a (step length $\rho = 0.001$, convergence precision $\varepsilon = 0.002$, $\eta = 0.5$).

Table 1. Cost functions and time required in this shoes dual-channel SCN.

SC 1		SC 2	
Time	Cost	Time	Cost
$t_{p_1} = 2$	$f_1 = 4q_1$	$t_{p_2} = 2$	$f_2 = 4q_2$
$t_{d_{11}} = 2.5$	$c_{nk} = 2q_{nk} + 1$	$t_{d_{21}} = 2.5$	$c_{nk} = 2.5q_{nk} + 1$
$t_{d_{12}} = 3$	$c_{nk} = 2q_{nk} + 1$	$t_{d_{22}} = 3$	$c_{nk} = 2.5q_{nk} + 1$
$t_{r_{11}} = 1$	$c_{mk} = 3q_{mk}$	$t_{r_{21}} = 1$	$c_{mk} = 3q_{mk} + 1$
$t_{r_{12}} = 1$	$c_{mk} = 3q_{mk} + 1$	$t_{r_{22}} = 1$	$c_{mk} = 3q_{mk}$

Considering 2 different cases, we respectively generate the sales record of 400 heterogeneous customers to estimate the consumer's valuation, C_{ij} , price sensitivity coefficient α_{ij} , time sensitivity coefficient β_{ij} and retail service sensitivity coefficient γ_{ij} for each demand market. We test 2 cases where the customers are heterogeneous follow the under distributions. Each case has a special consumer's valuation, or high price sensitivity or high retail service sensitivity.

Case 1: The customers are heterogeneous in α , β and γ , i.e., $C = 1$, $\alpha = N(0.07, 0.002^2)$, $\beta = N(0.05, 0.001^2)$ and $\gamma = N(0.05, 0.001^2)$.

Case 2: The customers are heterogeneous in α , β and γ , i.e., $C = \mathcal{U}(0.9, 1.1)$, $\alpha = N(0.05, 0.002^2)$, $\beta = N(0.05, 0.001^2)$ and $\gamma = N(0.07, 0.001^2)$.

These two cases are used to simulate the implement of our dual-channel network equilibrium model when the firms face heterogeneous customers in practice. At first, we use the method introduced in section 2.3 to estimate consumer's valuation, price sensitivity coefficient, time sensitivity coefficient and retail service sensitivity coefficient. The coefficients that we estimated using MLE are shown in table 2. Then, we use the coefficients in table 2 to compute the equilibrium flow and price for each case.

Table 2. $\vec{\theta}$ estimated by MLE with MNL.

Coefficient	Demand market	C	α	β	γ
Case 1	1	0.9892	0.0516	0.0494	0.0494
	2	0.9876	0.0513	0.0482	0.0482
Case 2	1	0.9779	0.0534	0.0479	0.0679
	2	0.9859	0.0524	0.0492	0.0692

For the network and cost structure of the shoes dual-channel SCN given above, assume that the utility function at markets, $U = C - \alpha P - \beta(t_{mk} \text{ and } t_{nk}) + \gamma S + \varepsilon$. Then, the equilibrium is computed in table 3.

Table 3. Computed equilibrium values on links ($\rho = 0.001$, $\varepsilon = 0.0001$).

Results	q_{nm}		q_{mk}				q_{nk}				p				s			
	q_{11}	q_{22}	q_{11}	q_{12}	q_{21}	q_{22}	q_{11}	q_{12}	q_{21}	q_{22}	p_{11}	p_{12}	p_{21}	p_{22}	s_{11}	s_{12}	s_{21}	s_{22}
Case 1	5522.0	6482.6	2015.8	576.5	1055.3	2356.2	1464.8	1464.8	1535.5	1535.5	24.0	24.7	24.2	24.9	39	39	39	39
Case 2	5346.1	6655.6	2188.2	224.0	878.7	2710.0	1467.0	1467.0	1533.4	1533.4	35.8	36.9	36.7	37.8	42	42	42	42

4. CONCLUSIONS

Dual-channel SCN is facing many problems such as low logistics efficiency, long transportation time and channel conflict. All these problems are required to be solved by a systematic and intelligent network^[34]. The tool and method of designing and managing effective dual-channel SCN would be beneficial to firms and consumers in dual-channel SCN. In this paper, a dual-channel SCN equilibrium model with customer-driven was developed considering heterogeneous consumer characteristic attributes, which is comprised of multiple product manufacturers, retailers and customers in competing markets. We began by considering operation time on links associated with different transaction cost and retail service. Next, a discrete choice model was designed to illustrate the customer choice from consumer-driven perspective in competing markets where the consumers make decisions among selling price, operation time and retail service according to their characteristic attributes. Furthermore, a finite-dimensional variational inequality was adopted to formulate the dual-channel SCN equilibrium conditions. Finally, a numerical example was provided to illustrate the model and the computational procedure.

5. ACKNOWLEDGEMENTS

This work was financially supported by the Humanities and Social Sciences of Ministry of Education Planning Fund (grant number: 16YJC630085) in China.

REFERENCES

- [1] S. Collett, "Channel Conflicts Push Levi to Halt Web Sales.," *Computerworld*, vol. 33, no. 45, p. 8, 1999.
- [2] W. K. Chiang, D. Chhajed, and J. D. Hess, "Direct Marketing, Indirect Profits: A Strategic Analysis of Dual-Channel Supply-Chain Design," *Manage. Sci.*, vol. 49, no. 1, pp. 1–20, 2003.
- [3] S. H. Chun and J. C. Kim, "Pricing strategies in B2C electronic commerce: Analytical and empirical approaches," *Decis. Support Syst.*, vol. 40, no. 2, pp. 375–388, 2005.
- [4] J. Chen, H. Zhang, and Y. Sun, "Implementing coordination contracts in a manufacturer Stackelberg dual-channel supply chain," *Omega*, vol. 40, no. 5, pp. 571–583, 2012.
- [5] B. Li, P. W. Hou, P. Chen, and Q. H. Li, "Pricing strategy and coordination in a dual channel supply chain with a risk-averse retailer," *Int. J. Prod. Econ.*, vol. 178, pp. 154–168, 2016.
- [6] B. Dan, G. Xu, and C. Liu, "Pricing policies in a dual-channel supply chain with retail services," *Int. J. Prod. Econ.*, vol. 139, no. 1, pp. 312–320, 2012.
- [7] G. (George) Cai, Z. G. Zhang, and M. Zhang, "Game theoretical perspectives on dual-channel supply chain competition with price discounts and pricing schemes," *Int. J. Prod. Econ.*, vol. 117, no. 1, pp. 80–96, 2009.
- [8] S. Choi and K. Fredj, "Price competition and store competition: Store brands vs. national brand," *Eur. J. Oper. Res.*, vol. 225, no. 1, pp. 166–178, 2013.
- [9] X. Yue and J. Liu, "Demand forecast sharing in a dual-channel supply chain," *Eur. J. Oper. Res.*, vol. 174, no. 1, pp. 646–667, 2006.
- [10] W. Y. K. Chiang and G. E. Monahan, "Managing inventories in a two-echelon dual-channel supply chain," *Eur. J. Oper. Res.*, vol. 162, no. 2, pp. 325–341, 2005.
- [11] G. Hua, S. Wang, and T. C. E. Cheng, "Price and lead time decisions in dual-channel supply chains," *Eur. J. Oper. Res.*, vol. 205, no. 1, pp. 113–126, 2010.
- [12] H. Xu, Z. Z. Liu, and S. H. Zhang, "A strategic analysis of dual-channel supply chain design with price and delivery lead time considerations," *Int. J. Prod. Econ.*, vol. 139, no. 2, pp. 654–663, 2012.
- [13] J. Chen, L. Liang, D. Q. Yao, and S. Sun, "Price and quality decisions in dual-channel supply chains," *Eur. J. Oper. Res.*, vol. 259, no. 3, pp. 935–948, 2017.

- [14] A. Nagurney, A. H. Masoumi, and M. Yu, *An Integrated Disaster Relief Supply Chain Network Model with Time Targets and Demand Uncertainty*. Springer, Cham, 2015.
- [15] G. Zhang, W. Habenicht, and W. E. L. Spieß, "Improving the structure of deep frozen and chilled food chain with tabu search procedure," *J. Food Eng.*, vol. 60, no. 1, pp. 67–79, 2003.
- [16] A. Rong, R. Akkerman, and M. Grunow, "An optimization approach for managing fresh food quality throughout the supply chain," *Int. J. Prod. Econ.*, vol. 131, no. 1, pp. 421–429, 2011.
- [17] X. Wang and D. Li, "A dynamic product quality evaluation based pricing model for perishable food supply chains," *Omega*, vol. 40, no. 6, pp. 906–917, 2012.
- [18] J. Blackburn and G. Scudder, "Supply chain strategies for perishable products: The case of fresh produce," *Prod. Oper. Manag.*, vol. 18, no. 2, pp. 129–137, 2009.
- [19] M. Christopher, *Logistics and supply chain management*. London: Financial Times, 1992.
- [20] G. Stalk Jr. and A. M. Webber, "Japan's dark side of time," *Harv. Bus. Rev.*, vol. 71, no. 4, pp. 93–102, 1993.
- [21] A. Nagurney, M. Yu, J. Floden, and L. S. Nagurney, "Supply chain network competition in time-sensitive markets," *Transp. Res. Part E Logist. Transp. Rev.*, vol. 70, no. 1, pp. 112–127, 2014.
- [22] D. Zhang, "A network economic model for supply chain versus supply chain competition," *Omega*, vol. 34, no. 3, pp. 283–295, 2006.
- [23] M. Fisher, S. Gallino, and J. Li, "Competition-Based Dynamic Pricing in Online Retailing: A Methodology Validated with Field Experiments," *Manage. Sci.*, 2017.
- [24] S. P. Anderson, A. de Palma, and J. F. Thisse, *Discrete choice theory of product differentiation*, vol. 12, no. 2. Cambridge, MA: MIT Press, 1992.
- [25] J. I. Yellott, "The relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the double exponential distribution," *J. Math. Psychol.*, vol. 15, no. 2, pp. 109–144, 1977.
- [26] S. P. Anderson, A. de Palma, and J. F. Thisse, *Discrete choice theory of product differentiation*, vol. 12, no. 2. 1992.
- [27] J. F. Nash, "Equilibrium points in n-person games," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 36, no. 1, pp. 48–49, 1950.
- [28] J. F. Nash, "Non-cooperative games," *Ann. Math. Second Ser. Ann. Math.*, vol. 54, no. 2, pp. 286–295, 1951.
- [29] J. Lions and G. Stampacchia, "Variational inequalities," *Commun. pure Appl. Math.*, vol. 20, no. 3, pp. 493–519, 1967.
- [30] A. Nagurney, D. Zhang, and J. Dong, "Spatial economic networks with multicriteria producers and consumers: Statics and dynamics," *Ann. Reg. Sci.*, vol. 36, no. 1, pp. 79–105, 2002.
- [31] A. Nagurney, *Network economics: A variational inequality approach*, vol. 10. Springer Science & Business Media, 2013.
- [32] G. M. Korpelevich, "The extragradient method for finding saddle points and other problems," *Matecon*, vol. 12, pp. 747–756, 1976.
- [33] R. Yan and Z. Pei, "Retail services and firm profit in a dual-channel market," *J. Retail. Consum. Serv.*, vol. 16, no. 4, pp. 306–314, 2009.
- [34] R. Yan and R. Yan, "Optimization approach for increasing revenue of perishable product supply chain with the Internet of Things," *Ind. Manag. Data Syst.*, vol. 117, no. 4, pp. 729–741, 2017.

The Impacts of Open and Proprietary IT on Vertical Firm Boundaries

Fengfeng Lou¹, Wen Guang Qu^{2}*

¹ School of Management, Zhejiang University, Hangzhou, 310058, China

² School of Management, Zhejiang University, Hangzhou, 310058, China

Abstract: Extensive researches have studied the relationship between IT and vertical firm boundaries, but few has distinguished between different types of IT. This study divides IT into two categories, open IT and proprietary IT and proposes that the impacts of open IT and proprietary IT on vertical firm boundaries are different. Moreover, we also hypothesize that industry dynamism has moderating effects on these relationships. We use the panel data from U.S. Bureau of Economic Analysis (BEA) to test these hypotheses. The results show that open IT has a negative effect on vertical scope of firms while proprietary IT has a positive effect on vertical scope. Furthermore, with the increase of industry dynamism, open IT has a further negative effect on vertical scope while proprietary IT has a further positive effect on vertical scope. This study suggests that different types of IT have different impacts on vertical firm boundaries, which provides implications for both research and practice.

Keywords: vertical boundaries, open IT, proprietary IT, industry dynamism, vertical scope

1. INTRODUCTION

As firms have increasingly invested in information technology (IT) resources over the past few decades^[1], more and more researchers are interested in studying how IT can affect the structure of firms^[2]. Previous literature has shown that IT makes significant impacts on vertical firm boundaries^[3, 4]. However, while some scholars claim that IT reduces the vertical scope of firms^[4, 5], others state that IT has the ability to extend a firm's vertical scope^[3, 6].

Although there are many studies focusing on the impacts of IT on vertical firm boundaries, they all focus on overall IT. Few studies on this issue have distinguished between different types of IT. To advance our understanding on the issue, we divide IT into two distinct types, open IT and proprietary IT. The main purpose of our research is to empirically examine the effects of different types of IT on vertical firm boundaries. We used annual industry-level data from U.S. Bureau of Economic Analysis (BEA) and conducted an empirical analysis. Overall, we find evidence showing that open IT and proprietary IT have opposite effects on vertical scope of firms. Specifically, the results suggest that open IT has a negative relationship with vertical scope while proprietary IT is positively associated with the vertical scope. Moreover, we find significant moderating effects of industry dynamism on these relationships. Compared to in stable industries, both the negative effect of open IT and positive effect of proprietary IT become stronger in dynamic industries.

This study contributes to the research by showing the different effects of two distinct types of IT, Open IT and Proprietary IT, on vertical firm boundaries and the moderating effects of industry dynamism on these relationships.

The rest of the paper is organized as follows. First, we review the literature about the impacts of IT on vertical firm boundaries. Second, based on prior research on open and proprietary IT and industry dynamism, we propose hypotheses regarding the impacts of different types of IT on vertical firm boundaries and the moderating effects of industry dynamism. Then we conduct an empirical test on these hypotheses based on the annual industry-level data from BEA. Finally, the paper ends with a discussion of findings and implications for research and practice.

* Corresponding author. Email: quwg@zju.edu.cn(Wen Guang Qu) , lffsghr@163.com(Fengfeng Lou)

2. LITERATURE REVIEW

Many studies have suggested that IT may have an important impact on vertical firm boundaries by influencing firms' make vs buy decisions on the materials or services for producing their final products^[1, 4, 7]. By analyzing the rich literature on the impacts of IT on vertical firm boundaries, we find that there are different arguments and evidences for the impacts of IT on vertical firm boundaries. The literature could be classified into two groups.

The first group thinks an enterprise will tend to depend more on markets after using IT because IT has the ability to reduce coordination cost^[1]. Thus, a firm's wider use of IT is associated with the reduction of vertical scope. For instance, Brynjolfsson et al. (1994) predicted that increasing use of IT leads firms to narrow down their vertical scope. Their empirical analyses also showed that following investment in IT, firms would seek to coordinate more with markets and IT also leads to smaller size^[4]. Chen and Kamal (2016) claimed that when production processes are easily codified electronically, IT will increase external transactions relative to internal transactions in MNEs (multinational enterprises), and reduce firm scope^[7]. Hitt (1999) provided evidence linking IT to the reduction of vertical integration of firms, because the effect of IT on external coordination costs is stronger than on internal coordination costs^[5].

Contrary to above findings, however, the second group of studies indicates that there is a positive relationship between IT and vertical scope of firms. For instance, Chen and Kamal (2016) provided evidence that the complex IT should facilitate MNEs (multinational enterprises) to produce in house, expanding vertical scope^[7]. Afuah (2003) revealed that IT may expand some vertical scope of firms by enabling some external activities to be performed in house^[3]. Similarly, Ray et al (2009) showed that IT is positively related to firms' vertical integration in concentrated and predictable environments because firms can use IT to coordinate more activities in house for higher revenue^[6].

Overall, the findings are inconsistent in the literature. We believe that there may be an issue in prior researches that they only focus on overall IT. Because different types of IT have different characteristics^[8-10], they may have different impacts on vertical firm boundaries. Therefore, to augment our understandings, this study distinguishes between open IT and proprietary IT and suggests that they affect vertical firm boundaries differently.

3. HYPOTHESES

IT can be categorized into two types, open IT and proprietary IT^[8-10].

Open IT is defined as one which uses public communication platforms and software, available to the public (e.g., RosettaNet and ebXML)^[9]. Plenty of benefits of open IT have been revealed by prior researches, which contains supply chain flexibility and operational cost savings^[11]. For instance, Malhotra et al. (2007) indicated that open IT enables firms to exchange information with many extended enterprise partners and heralds flexibility of business process connectivity across enterprises^[11].

On the contrary, when IT is developed and available only to a specific enterprise, requiring a private platform and software to communicate, it is considered as proprietary IT (e.g., the early ASAP system characterized as a dedicated system with proprietary protocols)^[9]. There are also numerous benefits of proprietary IT documented in the literature, including partnership relationship and competitive advantages^[8, 12]. For instance, Subramani (2004) provided empirical evidence that proprietary IT use can lead to closer cooperative relationships, which contributes to partnership relationship in buyer-supplier exchanges^[12].

3.1 Open IT and vertical firm boundaries

The impacts of open IT on vertical firm boundaries can be analyzed from the angles of transaction costs and firm flexibility.

Due to its standardization and openness, open IT enables firms to adapt to most enterprise information systems, which is convenient for inter-firm information exchange and communication. Well-functioning IT systems on both sides allow firms to share product information easily with external suppliers and reduce coordination costs^[6]. Lower coordination costs contribute to better cooperative relationship and maintain business cooperation between firms and external suppliers. Besides coordination costs, transaction costs contain costs of finding, contracting with suppliers, and resolving disputes in the business process^[6]. Open IT can help firms reduce such costs in the procurement process. Thus open IT can make firms lower transaction costs involved in procuring from suppliers. With lower transaction costs, firms prefer to procure from external suppliers and reduce their own vertical scope.

In the present competitive environment, an enterprise should retain the flexibility of changing strategies frequently to respond to market changes^[6]. Firms which use open IT usually have plenty of suitable suppliers to cooperate. So they are free to choose the most appropriate one from many potential suitable supplier partners to cooperate with, without worrying about being restricted by committed relationship. As a result, those firms that apply open IT prefer to procure outside.

Open IT can help firms reduce external transaction costs and improve enterprise flexibility when they procure from external suppliers. Thus, firms using open IT would prefer external procurement rather than internal production, which leads them to reduce their vertical scope.

Hypothesis 1: The use of open IT is negatively associated with the vertical scope of firms.

3.2 Proprietary IT and vertical firm boundaries

Usually, firms based on proprietary IT typically require extra costs like high relationship-specific investments to build committed partnership if they need to cooperate with external suppliers^[10]. Undoubtedly, this would increase coordination costs of both parties to the transaction and the switching costs of firms if they want to shift from one supplier to another. High switching costs result in the loss of flexibility when the focal firm changes suppliers. Moreover, cooperation between the focal firm and the external supplier often involves opportunistic risks^[13]. Once the supplier partner performing opportunistic behaviors, the focal firm needs to seek for another supplier partner to avoid risks but pays high switching costs in the meantime.

In addition, for firms that use proprietary IT, proprietary IT has significant effects on both internal and external costs. Coordination costs can be further divided into two distinct parts, internal coordination costs and external coordination costs^[4]. Since proprietary IT adapts to internal business procedure and characteristics of an enterprise, it becomes much easier and more convenient for the enterprise to coordinate internally and lower internal coordination costs. When IT reduces internal coordination costs more than external coordination costs, firms are expected to choose internal production decision^[4]. Besides, it will be hard and costly for the focal firm to find a suitable supplier because of the unique characteristic of proprietary IT. To avoid high costs, firms using proprietary IT prefer to produce internally rather than procure from external suppliers, which expands their vertical scope.

For firms using proprietary IT, it is hard to find a suitable supplier to cooperate with in the market. Furthermore, external procurement will cost a lot and lead to the inflexibility of firms while the internal production has lower costs and risks. Therefore, internal production which expands vertical scope would be the better decision for the firms.

Hypothesis 2: The use of proprietary IT is positively associated with the vertical scope of firms.

3.3 Industry dynamism

Industry dynamism represents the degree of change in industry environmental activities related to an organization's operations^[14]. The change is hard to predict, thus industry dynamism could create uncertainty and make an impact on firms' managerial decision-making^[15]. In this study, it is analyzed and inquired into the different effects of industry dynamism on the relationship between IT and firms' make vs buy decisions after using different type of IT (open IT or proprietary IT).

Accordingly, a firm must have the necessary agility to survive in the dynamic environment. Open IT gives firms the agility to seek and change suitable suppliers easily whenever necessary. Procuring from external suppliers is highly agile for those firms that apply open IT, which is very important for their survival in dynamic environment. Therefore, compared in stable environment, firms with open IT prefer to procure more from outside in dynamic industry environment. This leads to the narrowing of vertical scope of firms.

Within dynamic environment, firms which apply open IT need agility to survive and external procurement offers this agility. Thus, the firms tend to procure from external suppliers and reduce their vertical scope.

Hypothesis 3: The relationship between open IT and vertical scope of firms will become more negative when industry dynamism is higher.

In dynamic industry environment, IT investment decision-making of enterprise is a demanding task because of the importance and irreversibility of IT investment itself^[16]. Overinvestment in IT wastes resources while underinvestment in IT results in critical opportunity costs^[17]. As we know, IT investment is a kind of sunk cost. It is easier for IT investment to lose value in uncertain industry environment contrast to in stable environment. As a result, firms become very cautious in investing in IT resources in uncertain environment. When a firm with proprietary IT chooses to procure from external suppliers, both sides need to increase IT investments for better IT integration and building committed relationship during the cooperation period^[10]. But in dynamic environment, it's hard for the focal firm to find a supplier which is willing to investing in IT resources for the cooperation. Therefore, the firms using proprietary IT are harder to procuring from external suppliers in uncertain environment.

For firms using proprietary IT, it becomes harder to find a suitable supplier to cooperate with in dynamic environment. At the same time, the cost of internal coordination is still at a lower level. Thus, it is better for those firms using proprietary IT to choose internal production rather than external procurement, (i.e., expanding their vertical scope).

Hypothesis 4: The relationship between proprietary IT and vertical scope of firms will become more positive when industry dynamism is higher.

4. METHOD

4.1 Data and variables

To test the hypotheses, we gathered annual industry-level data from U.S. Bureau of Economic Analysis (BEA). The data includes 62 industries at the three-digit North American Industry Classification System (NAICS) level, covering the years from 1998 to 2015. We use the Use Tables from the BEA's Input-Output

Accounts to measure the vertical scope of firms and industry dynamism. For open IT investments and proprietary IT investments, we can use data in the BEA's Fixed Assets Tables to compute them. As previous studies^[18], we assume that the industry average can be considered as the data of a 'representative' firm in the industry. Thus, we can use industry-level data to carry out firm-level research.

4.1.1 Dependent variable

To better analyze the IT impacts on vertical firm boundaries, we decide to take the vertical scope of firms as the dependent variable in empirical test. Since our premise in this study is that firm's internal production and external procurement are substitutive for each other, we can take the ratio of external procurement value to internal production value as a measure of the vertical scope of firms. More precisely, we use the negative value of this ratio to measure the vertical scope, because the ratio represents the level of external procurement of firms compared to internal production. We consider intermediate inputs from the BEA's Use Tables as a measure of external procurement and value added in the Use Tables as a measure of internal production of firms. The denominator is the value of value added and the numerator is the negative value of intermediate inputs. Therefore, the measure of our dependent variable is:

$$VS = -\text{intermediate inputs} / \text{value added} \quad (1)$$

4.1.2 Independent variables

In this study, the use of open IT is measured by the investment on prepackaged software. In the BEA's Fixed Assets Tables, it refers to the stock of prepackaged software. BEA categorized IT software stock into three groups: prepackaged software stock, custom software stock and own account software stock. Prepackaged software is the category of public software, available to the public. A firm using prepackaged software is considered as using open IT. Therefore, we use a firm's investment on prepackaged software to measure its use of open IT.

We measure the use of proprietary IT by the sum of custom software stock and own account software stock in the BEA's Fixed Assets Tables. In the BEA's Fixed Assets Tables, both custom software and own account software are the categories of software which are developed and available only to a specific enterprise, which is proprietary. Therefore, we use investments on both custom and own account software to measure the use of proprietary IT of firms.

As to industry dynamism, we measure it by using industry gross output from the BEA's Use Tables. As Keats and Hitt (1988) mentioned^[19], we regress the natural logarithms of annual figures of industry gross output on time. Then the antilog of the standard error of the regression slope coefficient can be a measure of industry dynamism. This variable describes the instability of industry environment. Moreover, we take industry dynamism as a moderating variable, which is used to study the moderating effects of industry environment on the relationship between IT and the vertical scope of firms.

In addition, we include non-IT capital (firm's total capital excluding IT), computer hardware capital, and labor as control variables. Using data from the BEA's Fixed Assets Tables, we can compute the value of non-IT capital and computer hardware capital. To measure the labor, we follow the way of prior research^[20]. First, we get the data of the number of full-time equivalent employees from the BEA's Full-Time Equivalent Employees Table. Then we can calculate the value of labor by multiply the number by 2080 hours (average work hours). All independent variables (except for industry dynamism) are divided by the value-added of the corresponding industry to control the size effect.

4.2 Econometric model

To test the effects of open and proprietary IT on vertical scope of firms and the moderating effects of industry dynamism, we implement econometric model as follows:

$$VS_{it} = \beta_0 + \beta_1 Cst_{it} + \beta_2 Pkg_{it} + \beta_3 K_IT_{it} + \beta_4 L_{it} + \beta_5 Hd_{it} + \beta_6 Ind_dyn_{it} + \beta_7 (Ind_dyn * Cst)_{it} + \beta_8 (Ind_dyn * Pkg)_{it} + \varepsilon_{it} \quad (2)$$

VS_{it} represents the vertical scope of firms in an industry i at year t . Cst_{it} and Pkg_{it} are independent variables, representing proprietary and open IT investments of firms for an industry i at year t . Control variables includes non-IT capital, labor, and computer hardware capital of firms of an industry i at year t . The ε_{it} in the model represents the error term of observation. In addition, we add 61 industry dummies (62 industries totally) to control for industry fixed effects and 17 year dummies (1998-2015) to control for time effects in the analysis.

4.3 Results

Following Han et al. (2011)^[20], we analyze the data using OLS with panel-corrected standard errors (OLS-PCSE). The summary of the analysis results is provided by Table 1.

As shown in the table, Model 1 only includes direct effects. There is a significant negative relationship between open IT and vertical scope of firms ($\beta = -0.225$, $p < 0.001$ for the PCSE-PSAR1 model; $\beta = -0.227$, $p < 0.001$ for the PCSE-AR1 model). Hypothesis 1 is supported. As for proprietary IT, we found a significant positive relationship between proprietary IT and vertical scope of firms ($\beta = 0.149$, $p < 0.01$ for the PCSE-PSAR1 model; $\beta = 0.119$, $p < 0.01$ for the PCSE-AR1 model), which is consistent with Hypothesis 2.

Model 2 of Table 1 adds the hypothesized interaction terms between different type of IT investment and the moderator variable. As Table 1 shows, Hypothesis 3 is also supported. The interaction term involving industry dynamism and open IT has a negative coefficient ($\beta = -0.170$, $p < 0.001$ for the PCSE-PSAR1 model; $\beta = -0.237$, $p < 0.001$ for the PCSE-AR1 model), showing an increasingly negative effect of open IT on vertical scope in highly dynamic industries. On the other hand, the interaction term of industry dynamism and proprietary IT has a positive effect on vertical scope ($\beta = 0.285$, $p < 0.01$ for the PCSE-PSAR1 model; $\beta = 0.328$, $p < 0.001$ for the PCSE-AR1 model). The results suggest a stronger positive relationship between proprietary IT and vertical scope in dynamic industries than in stable industries, which supports Hypothesis 4.

Table 1. Main results of OLS-PCSE analysis

Variables	Model 1		Model 2	
	PCSE(PSAR1)	PCSE(AR1)	PCSE(PSAR1)	PCSE(AR1)
Control variables				
Non-IT capital(K_IT)	-0.546***	-0.524***	-0.541***	-0.514***
Labor(L)	-0.602***	-0.603***	-0.614***	-0.610***
Computer hardware capital(Hd)	-0.404***	-0.377***	-0.398***	-0.378***
Direct effects				
Open IT (Pkg)	-0.225***	-0.227***	-0.216***	-0.191***
Proprietary IT (Cst)	0.149**	0.119**	0.111*	0.068
Moderating effects				
Industry dynamism x open IT			-0.170***	-0.237***
Industry dynamism x proprietary IT			0.285**	0.328***
†p<0.1, *p<0.05, **p<0.01, ***p<0.001				

5. CONCLUSIONS

In this paper, we classify IT into two categories, namely open IT and proprietary IT, and conduct an

empirical analysis on the relationships between them and vertical scope of firms as well as the moderating effects of industry dynamism. The results show that the effects of open IT and proprietary IT on vertical scope are opposite to each other. The increasing use of open IT will reduce a firm's vertical scope while increase in proprietary IT can expand the vertical scope. In addition, the results also show that the effects of open and proprietary IT become even stronger with the increase of industry dynamism.

Open IT lowers external transaction costs^[6] and provide flexibility to firms when they procure outside. For firms using open IT, external procurement offers more benefits than internal production. As a result, it is a rational decision for those firms to choose external procurement and reduce their vertical scope. However, proprietary IT has exactly the opposite effect. Once firms use proprietary IT, there will be a high cost for procuring from external suppliers^[10]. But internal production is at a low cost. Thus firms will choose to produce internally and expand their vertical scope if they rely on proprietary IT. In addition, industry dynamism intensifies above relationships. Both effects are stronger for high levels of industry dynamism as compared to low levels. External procurement can provide agility to firms when they use open IT, which is of vital importance for them to survive in dynamic environment. So firms which use open IT are even more inclined to procure from outside in dynamic environment. But for those using proprietary IT, high industry dynamism will increase their external transaction costs and make it hard for them to find a willing and suitable external supplier. Therefore in dynamic environment, these firms prefer even more internal production.

Previous researches have examined the impacts of IT on vertical firm boundaries. To our knowledge, however, there is no study that has investigated the roles of open and proprietary IT on this issue. Our study contributes to the research by distinguishing between the effects of open IT and proprietary IT on vertical firm boundaries. In addition, this research also offers implications to enterprise managers when they making business decisions on IT. Finally, our study shows that industry dynamism has significant moderating effects on the relationships between open/proprietary IT and vertical firm boundaries. Future research may augment our understandings by examining the roles of other important industrial environment factors, such as industry munificence and industry concentration, in above relationships.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71572180.

REFERENCES

- [1] V Gurbaxani, S Whang. (1991).The impact of information systems on organizations and markets. *Communications of ACM*, 34(1):59-73.
- [2] S Dewan, F Ren. (2011).Information Technology and Firm Boundaries: Impact on Firm Risk and Return Performance. *Information Systems Research*, 22(2):369-388.
- [3] Allan Afuah. (2003).Redefining Firm Boundaries in the Face of the Internet: Are Firms Really Shrinking? *Academy of Management Review*, 28(1):34-53.
- [4] E Brynjolfsson, TW Malone, V Gurbaxani, A Kambil. (1994).Does Information Technology Lead to Smaller Firms? *Management Science*, 40(12):1628-1644.
- [5] Lorin M. Hitt. (1999).Information Technology and Firm Boundaries: Evidence from Panel Data. *Information Systems Research*, 10(2):134-149.
- [6] Gautam Ray, Dazhong Wu, Prabhudev Konana. (2009).Competitive Environment and the Relationship Between IT and Vertical Integration. *Information Systems Research*, 20(4):585-603.
- [7] W Chen, F Kamal. (2016).The impact of information and communication technology adoption on multinational firm boundary decisions. *Journal of International Business Studies*, 47(5):563-576.

- [8] WG Qu, A Pinsonneault, D Tomiuk, S Wang, Y Liu. (2015).The impacts of social trust on open and closed B2B e-commerce: A Europe-based study. *Information & Management*, 52(2):151-159.
- [9] K Zhu, KL Kraemer, V Gurbaxani, SX Xu. (2006).Migration to Open-Standard Interorganizational Systems: Network Effects, Switching Costs, and Path Dependency. *Mis Quart*, 30(30):515-539.
- [10] WG Qu, D Tomiuk, Z Wang. (2009).The Impact of Social Trust on Electronic Interfirm Cooperation. In: IEEE Computer Society, ed. 2009 2nd IEEE International Conference on Computer Science and Information Technology. Beijing: IEEE Computer Society, 316-319
- [11] A. Malhotra, S. Gosain, O.A. El Sawy. (2007).Leveraging standard electronic business interfaces to enable adaptive supply chain partnerships. *Information Systems Research*, 18(3):260-279.
- [12] M. R. Subramani. (2004).How Do Suppliers Benefit from Information Technology Use in Supply Chain Relationships? *Mis Quart*, 28(1):45-73.
- [13] OE Williamson. (1985).The Economic Institutions Of Capitalism: Firms, Markets, Relational Contracting. New York: Free Press, 450
- [14] WA Randolph, GG Dess. (1984).The Congruence Perspective of Organization Design: A Conceptual Model and Multivariate Research Approach. *Academy of Management Review*, 9(1):114-127.
- [15] B.W. Keats, M.A. Hitt. (1988).A causal model of linkages among environmental dimensions, macro organizational characteristics, and performance. *Academy of Management Journal*, 31(3):570-598.
- [16] P Weill, M Broadbent. (1998).Leveraging the New Infrastructure: How Market Leaders Capitalize on Information Technology. *Journal of Global Information Technology Management*, 6(4):620-627.
- [17] P Weill, M Subramani, M Broadbent. (2002).Building IT infrastructure for strategic agility. *Sloan Management Review*, 44(1):57-65.
- [18] A Sahaym, HK Steensma, JQ Barden. (2010).The influence of R&D investment on the use of corporate venture capital: An industry-level analysis. *Journal of Business Venturing*, 25(4):376-388.
- [19] BW Keats, MA Hitt. (1988).A causal model of linkages among environmental dimensions, macro organizational characteristics, and performance. *Academy of Management Journal*, 31(3):570-598.
- [20] K Han, RJ Kauffman, BR Nault. (2011).Returns to Information Technology Outsourcing. *Information Systems Research*, 22(4):824-840.

Impacts of Psychological Distance-based Sales Promotion on Online Purchasing Behaviors under Different Involvement

Liyi Zhang^{1}, Shan Huang², Qihua Liu³*

¹ School of Information Management, Wuhan University, China

² School of Information Management, Wuhan University, China

³ School of Information Technology, Jiangxi University of Finance and Economics, China

Abstract: Via consumer surveys after the “Double 11” promotion, we studied consumers’ consumption behavior and its influencing factors (temporal distance, social distance, product types and purchase decision involvement) based on the CLT and involvement theory with logistic regression modeling. The results show that the effect of temporal distance on purchasing decisions is increasing in high-involvement products and decreasing in low-involvement products, while social distance has a negative impact on purchasing decisions in both high and low-involvement products. Consumers’ purchase decision involvement is reinforced by temporal distance, while is no relevant to social distance. Specifically, when consumers are temporally distant from knowing the promotion issues, their purchase decision involvement tends to be higher and cost more in online promotion. Results provide practical marketing implications and help to enrich marketing theory.

Keywords: temporal distance, social distance, product involvement, purchase decision involvement, sales promotion

1. INTRODUCTION

Sales promotion is becoming increasingly penetrated among individuals, which is also an important means to clear inventory and improve sales. In China, “Double 11”(also called Single’s Day), created by Taobao.com in 2009 and launched in November 11th of each year, is the most typical online sales promotion activity, followed by many other e-commerce platforms. It attracted a substantial number of consumers and the sales on Taobao.com reached 14.3 billion US\$ in the last Single’s Day. Do online promotions influence consumers’ purchase decision? Furthermore, do consumers enjoy online promotion and buy more? It is imperative to understand how online promotions impact consumers’ purchase decision and which factors are effective in this process.

A large number of determinants of online shopping behavior have been identified and discussed on sales promotion, which focus on perceived characteristics of the web as a sale channel, website and product characteristics, and consumer characteristics^[1]. In addition to external stimulus, psychological representation could also affect the purchase decision-making process of consumers in online sale promotion situations, which need further research. The construal level theory revealed the influence of psychological distance on consumers' expectation, evaluation and purchase behavior^[2]. Although scholars have examined how the psychological distance affects consumers’ impulse buying and purchase decisions and salient highlight the information contents^[3], information forms^[4] and decision-making process^[5], they fail to take into consideration the role of involvement in the process of consumer decision making. In terms of the object of psychological distance, compared to products with low involvement, the decision making process of products with high involvement often took more time and careful consideration^[6]; regarding the subject of psychological distance, consumers who shown higher purchase-decision involvement entailed a greater expenditure of time and money^[7].

The objective of this paper is to empirically examine the impact of psychological distance on online purchase behaviors of high and low involvement products in online sales promotion. Comparing the different

* Corresponding author. Email: lyzhang@whu.edu.cn

strategies invoked by psychological distance on purchasing of high involvement vs. low involvement products, we explore the interactive relationship between psychological distance and product involvement in online sales promotion. Then, we examine the effects of psychological distance on consumers' purchase decision involvement and validate whether these two aspects would salient facilitate the consumer purchase behavior.

2. DEVELOPMENT OF RESEARCH HYPOTHESES

The construal level theory proposes that individuals tend to carry on more concrete representations (lower level construal) on near events, while represent the distant events more abstractly (higher level construal)^[8]. As high-involvement products generally possess higher capital values, they need more time to consider before making the purchase decision, and their decision-making process emphasis more on desirability demands, which represent a high level construal. On the contrary, low-involvement products always emphasize feasibility demands, which represent a low level construal. According to cognitive match principle, consumers are more likely to purchase high-involvement products in the distant future enabled with high level construal; but withal preferring to purchase low-involvement products in the near future enabled with low level construal^[6]. Consequently, we propose the following hypotheses:

H1a. When purchasing high involvement products, the temporal distance is positively related to consumers' purchase decision.

H1b. When purchasing low involvement products, the temporal distance is negatively related to consumers' purchase decision.

In terms of information receiver, even if a product's own quality is constant, consumers would perceive it a higher level construal from the recommendation of socially distant than from the corresponding proximal alternatives. Whereas, the cognition of high level construal manner requires more time-cost and cognitive effort for consumers to make purchase decisions relied on the objective evaluation, thus, the consumer is easier influenced by subjective cognition and other psychological stimuli. Accordingly, we propose the following hypotheses:

H2a. When purchasing high involvement products, the social distance is positively related to consumers' purchase decision.

H2b. When purchasing low involvement products, the social distance is negatively related to consumers' purchase decision.

Consumers' mental construal differed from diffusion path and time of the information even in the same sales promotion. Specifically, consumers who receive the activity information in the moment or from the dissimilar others will construe a high level of mental construal. The construal level generated from psychological distance would further influence consumers' evaluation behavior, including the main, goal-related properties or secondary, goal-irrelevant properties; feasibility or desirability demands; standing in the idealistic perspective or realistic perspective, etc.^[2]. Consumers who have confirmed their personal values are more willing to make additional efforts to obtain the information and reduce the risk of uncertainty, so as to show a high level of purchase-decision involvement^[9]. So, we can make the following hypotheses:

H3a. The temporal distance is positively associated with consumer's purchase-decision involvement.

H3b. The social distance is positively associated with consumer's purchase-decision involvement.

The total consumption refers to the amount of a consumer who finally consumed in the promotion. Chiou and Ting^[10] has examined the differences between online shopping motivation and product type on the searching and spending behavior in a website. They indicate that goal-oriented consumers are willing to spend more on hedonic goods without guilt and regret. Tangari^[11] has indicated that consumers' purchase intentions can be significantly influenced by temporal distance and social distance. Furthermore, temporally distant events and

socially distant events would accordingly lead to a more abstract information expression which represents a high level construal^[2]. So, we have:

H4a. In sales promotion, consumer's temporal distance is positively related to the total consumption.

H4b. In sales promotion, consumer's social distance is positively related to the total consumption.

O'Cass^[12] has found that, as an antecedent to purchase-decision involvement, product involvement would positively impact consumers' purchase-decision involvement. Hereby, we speculate that consumers involved with higher level of purchase-decision involvement pretend to purchase more high involvement products. Furthermore, as high involvement products tend to correspond to high capital value^[6], the total consumption is easier to accumulating by consumers purchase behaviors. Thus, we suggest the following:

H5. In sales promotion, purchase decision involvement positively impacts consumers' total consumption.

The theoretical framework is shown in Figure1.

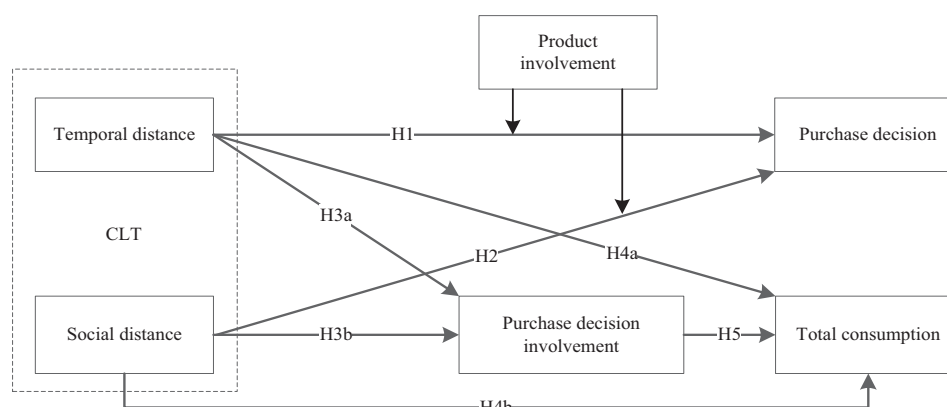


Figure 1. Theoretical framework

3. DATA AND MEASURES

3.1 Data collection

A total of 577 respondents participated in the survey during the 7 day sample collection period with the efficiency of 92.2%, of which 45 invalid questionnaires were excluded and 532 valid questionnaires were sampled. The results of descriptive statistical analysis on respondents are shown in Table 1, and manifest that the questionnaire objects basically conform to the general distribution of online shoppers.

Table 1. The descriptive statistical analysis on respondents

Item	Characteristics	N	%
Gender	Male	214	40.2
	Female	318	59.8
Age	<18	9	1.7
	18-25	359	67.5
	26-34	130	24.4
	35-45	31	5.8
	>45	3	0.6
Education background	Senior high school and less	121	22.7
	Undergraduate	336	63.2
	Postgraduate	64	12.0
	Doctor degree and more	11	2.1

3.2 Measures

Temporal distance (Temporal D) refers to the time interval between the time consumer perceiving the “Double 11” promotion and the beginning time of the promotion.

Social distance (Social D) is determined by the acquisition channel of the “Double 11” promotion.

Purchase decision (Purchase_D) measures whether consumers purchase the high or low involvement products, designated as a binary dummy variable.

Purchase decision involvement (Purchase_D_I) refers to the level of involvement for a consumer in decision making in the context of the “Double 11” promotion. We measure it with five-point scales adapted from O'Cass^[12] and Jung^[13].

Total consumption (Total_C) is utilized to measure consumer purchase amount. The questionnaire sets up five measurable indicators, which indicates the consumers total consumption increases gradually.

Control variables include *Income*, *Gender*, *Education*, *Shopping frequency* (Shopping F), *Shopping time* (Shopping T) and *Shopping equipment* (Shopping E).

The correlation coefficient matrix of the key variables (independent variables, dependent variables and control variables) and their summary statistics are shown in Table 2.

Table 2. The correlation coefficient matrix of the key variables

	Mean	S.D.	1	2	3	4	5	6	7	8	9	10
1.Total_C	2.36	0.97	1									
2.Temporal D	0.53	0.50	0.11	1								
3.Social D	0.76	0.43	-0.01	-0.01	1							
4.Purchase_D_I	3.09	1.11	0.18	0.10	-0.07	1						
5.Income	2.10	1.72	0.47	-0.04	-0.03	0.07	1					
6.Gender	0.40	0.49	0.07	-0.04	-0.10	0.02	0.21	1				
7.Education	1.93	0.65	0.12	0.12	-0.02	-0.02	0.07	-0.02	1			
8.Shopping E	1.60	0.50	-0.07	0.01	0.01	-0.04	0.01	0.09	0.11	1		
9.Shopping F	2.03	0.81	0.31	0.01	-0.02	0.19	0.39	-0.05	-0.04	-0.08	1	
10.Shopping T	2.87	1.58	-0.17	-0.23	0.03	-0.15	0.11	0.05	-0.07	0.06	-0.05	1

4. EMPIRICAL METHODOLOGY AND RESULTS

4.1 Empirical model

Our estimation strategy relies on the application of ordinary least square regression and logistic regression, processed by SPSS16.0. Following the procedure proposed by Hailpern and Visintainer^[14], we employ the following econometric specifications:

$$\text{logit}(p|Purchase_D_{ijt}) = \alpha * TD_i + \beta * SD_i + \sum_j \sigma_j * P_j + \sum_T \tau_t * C_t + \varepsilon_{ijt} \quad (1)$$

$$Purchase_D_I_{ijt} = \alpha * TD_i + \beta * SD_i + \sum_j \sigma_j * P_j + \sum_T \tau_t * C_t + \delta + \varepsilon_{ijt} \quad (2)$$

$$\text{logit}(p|Purchase_A_{ijt}) = \alpha * TD_i + \beta * SD_i + \gamma * Purchase_{D_{it}} + \sum_j \sigma_j * P_j + \sum_T \tau_t * C_t + \varepsilon_{ijt} \quad (3)$$

In the literature, we employ Equation (1) to test hypotheses H1a, H1b, H2a and H2b, Equation (2) to test hypotheses H3a and H3b, while Equation (3) to test hypotheses H4a, H4b, and H5. In addition, TD and SD represent temporal distance and social distance respectively, P_j a vector of user situational variables (including Shopping time and Shopping equipment), and C_t a vector of user characteristics variables (including Income,

Gender, Education and Shopping frequency). Lastly, α is a constant, subscript i indexes consumers, j indexes situational variables and t indexes user characteristics variables.

4.2 Results

Since the purchase decision is a binary dummy variable, which subjects to the binomial distribution and the residual sum to zero, so we apply binary logistic regression model. The regression results are shown in Table 3 and Table 4.

Table 3. Binary Logistic Regression Results of High Level Involvement Products

Parameter	B	S.E.	Wald	P value	Exp(B)
Temporal D	0.57	0.23	5.82	0.02**	1.76
Social D	-0.46	0.25	3.32	0.07*	0.63
Income	0.03	0.10	0.08	0.77	1.03
Gender	1.42	0.24	34.33	0.00***	4.15
Education	-0.02	0.18	0.01	0.92	0.98
Shopping F	0.43	0.15	8.54	0.00***	1.54

Notes: *, $p < 0.1$; **, $p < 0.05$; ***, $p < 0.01$; ns, not significant.

Table 4. Binary Logistic Regression Results of Low Level Involvement Product

Parameter	B	S.E.	Wald	P value	Exp(B)
Temporal D	-0.58	0.22	7.24	0.01***	0.56
Social D	-0.74	0.23	10.35	0.00***	0.48
Income	-0.28	0.11	6.60	0.01**	0.76
Gender	0.14	0.23	0.37	0.54	1.15
Education	0.19	0.17	1.25	0.26	1.21
Shopping F	0.33	0.15	5.16	0.02**	1.39

Notes: *, $p < 0.1$; **, $p < 0.05$; ***, $p < 0.01$; ns, non-significant.

From the regression results, we find that H1a, H1b and H2b were supported while H2a was not supported. That is, the perceived temporal distance is positively related to the purchase decision for high involvement products ($\alpha = 0.562$, $p < 0.05$), the perceived temporal distance ($\alpha = 0.555$, $p < 0.05$) and social distance ($\beta = 0.749$, $p < 0.05$) are positively associated with the purchase decision for low involvement products respectively. However, we find a negative interaction effect ($\beta = 0.459$, $p < 0.05$).

We employ ordinary least square regression to account for the situational and user characteristic effects. Table 5 presents our regression results.

Table 5. OLS Regression Results of Purchase-Decision Involvement

Parameter	B	Standard parameter estimate	S.E.	t value	P value
Temporal D	0.23	0.12	0.08	2.76	0.01***
Social D	-0.14	-0.06	0.10	-1.48	0.14
Income	0.13	0.15	0.04	3.22	0.00***
Gender	0.01	0.00	0.09	0.07	0.95
ShoppingT	-0.08	-0.12	0.03	-2.91	0.00***
Shopping F	0.26	0.21	0.06	4.57	0.00***
ShoppingE	-0.06	-0.03	0.08	-0.73	0.47
R ²	0.13				
Adjusted R ²	0.12				

Notes: *, $p < 0.1$; **, $p < 0.05$; ***, $p < 0.01$; ns, not significant.

In support of H3a, the regression results show positive relationships between temporal distance and purchase-decision involvement of consumers ($\alpha=0.23$, $p<0.05$). That is, the one who receives the promotion information earlier is more probably to show the high level of involvement in decision making. However, social distance is not significantly associated with purchase-decision involvement of consumers ($\beta=-0.14$, $p>0.05$), thus H3b is not support.

Since the total consumption is a continuous variable, we estimate the equation with ordinary least square regression. The regression results are shown in Table 6.

Table 6. OLS Regression Results of Total Consumption

Parameter	B	Standard parameter estimate	S.E.	t value	P value
Temporal D	0.20	0.08	0.09	2.15	0.03**
Social D	0.04	0.02	0.11	0.40	0.69
Purchase_D_I	0.10	0.09	0.04	2.33	0.02**
Income	0.47	0.45	0.04	10.79	0.00***
Gender	-0.04	-0.02	0.10	-0.45	0.65
ShoppingT	-0.15	-0.19	0.03	-4.85	0.00***
Shopping F	0.15	0.10	0.06	2.42	0.02**
ShoppingE	-0.14	-0.06	0.09	-1.54	0.13
R ²	0.30				
Adjusted R ²	0.29				

Notes: *, $p<0.1$; **, $p<0.05$; ***, $p<0.01$; ns, not significant.

As shown in Table 6, consumers with higher income or shopping frequently are prone to expend more (the *Total consumption is higher*) in the “Double 11” promotion. Moreover, the total consumption of consumers who participated in the promotion at the beginning is prominently higher than those participating at the end of the promotion. Additionally, the temporal distance ($\alpha=0.20$, $p<0.05$) positively impacts the total consumption, which supports H4a. And purchase-decision involvement ($\gamma=0.10$, $p<0.05$) is positively related to the consumer behavior in the “Double 11” promotion, supporting H5. However, the social distance is not significantly associated with total consumption ($\beta=0.04$, $p>0.05$), thus H4b is not supported.

5. DISCUSSION

The empirical results have rationally verified the theoretical model. On the one hand, based on the construal level theory we compared the different strategies invoked by psychological distance on purchasing of high or low involvement products, and found that when temporally or socially near from knowing the promotion issues, consumers prefer to purchase low involvement products (e.g., groceries), but high involvement products (e.g., digital products) in the distant future. However, hypothesis about the interactive relationship between social distance and purchase decision of high involvement product is not supported. The probable reason lies on that we ignore the decisive role of WOM (Word-of-Mouth) playing on consumers’ purchase behaviors. That is, compared with diffusing by advertising media, information exchange among acquaintances strengthens the effects of word of mouth specifically^{[15][16]}. Thus there emerges the unexpected phenomenon that consumers perceived socially near are more likely to make purchase decision. On the other hand, we introduce the intermediate variable—purchase-decision involvement to examine the effects of psychological distance on consumers’ purchase decision involvement and validate whether these two aspects would salient facilitate the consumer purchase behavior. As a result, the temporal distance is positively associated with consumers’ purchase-decision involvement, and then consumers’ total consumption is prominently impacted by these two

variables. Meanwhile, we find no significant relationship between social distances and purchase-decision involvement or total consumption, consistent with the sentiment of Clarke and Belk^[17]. The facilitation of social distances on consumers' purchase-decision involvement is finite, especially for high involvement products like digital products. Our study enriches the construal level theory and elucidates the underlying mechanisms behind time/social distance and consumer behaviors.

Our findings provide practical marketing implications to sales promotion. Firstly, While firm stakeholders plan to carry out online promoting activities, they should also take the consumer psychological factors into consideration to obtain better returns. Good matches of product class involvement and construal levels can efficiently promote consumers' purchase intentions. For the promotion of high involvement products like cars or digital products, marketers can recommend them to consumers well in advance, which will enhance the transmitting effects of online WOM. As for low involvement products like groceries or commodities, the strategy of online promotion can be focused on advertising campaigns. To allocate enterprise resources reasonably will be conducive to maximizing sales achievement. Secondly, since temporal distance is positively related to purchase-decision involvement, to broadcast promotion activities earlier will have a beneficial effect on the ultimate consumption. Purchase-decision involvement also provides new ideas for merchants to develop the online marketing activities. Customers with high levels of involvement in purchase decision making maybe have higher values than those with low levels of involvement.

This study still has several limitations. First, the prominent limitation is the selection of typical representations; we only select digital products as the high-involvement product type and groceries as the low-involvement product type. Although, the two types of products are typically objects, they are still just a small part of various merchandise categories. Thus, future research can take into account other categories of goods to improve the results. Second, we don't explore a salient dimension of psychological distance, spatial distance, because our data does not provide a clear deputy to spatial distance. Consider spatial distance plays an important role in affecting individuals' mental state and consumer behaviors, future research should explore possible measures of spatial distance and examine the mechanism of spatial distance worked in consumer purchase behavior.

ACKNOWLEDGEMENT

This paper was supported by the National Natural Science Foundation of China under Grant 71764006 and the MOE Project of Key Research Institute of Humanities and Social Science in Chinese Universities under Grant 14JJD870002.

REFERENCES

- [1] M. K. Chang, W. Cheung, and V. S. Lai. (2005). Literature derived reference models for the adoption of online shopping. *Information & Management*, 42(4):543-559.
- [2] Y. Trope, N. Liberman, and C. Wakslak. (2007). Construal levels and psychological distance: effects on representation, prediction, evaluation, and behavior. *Journal of Consumer Psychology*, 17(2):83-95.
- [3] P. Zou, L. He, and Y. Li. (2014). Impacts of Perceived Temporal Distance- Based Sales Promotion on Purchasing Behaviors under Information Asymmetry. *Management Science*, 1:65-74.
- [4] K. Kim, M. Zhang, X. Li. (2008). Effects of Temporal and Social Distance on Consumer Evaluations. *Journal of Consumer Research*, 35(4):706-713.
- [5] S. J. Hoch, G. F. Loewenstein. (1991). Time-Inconsistent Preferences and Consumer Self-Control. *Journal of Consumer Research*, 17(4):492-507.

- [6] C. Jiang, R. Duan, H. K. Jain, et al. (2015). Hybrid collaborative filtering for high-involvement products: A solution to opinion sparsity and dynamics. *Decision Support Systems*, 79:195-208.
- [7] R. W. Belk. (1982). Effects of gift-giving involvement on gift selection strategies. *Advances in Consumer Research*, 9(1):408-412.
- [8] Y. Trope. (2010). Construal-Level theory of psychological distance. *Psychological Review*, 117(2):440-463.
- [9] J. A. Howard, J. N. Sheth. (1969). The Theory of Buyer Behavior. *Journal of the American Statistical Association*, (1):467-487.
- [10] J. S. Chiou, C. C. Ting. (2011). Will you spend more money and time on internet shopping when the product and situation are right? *Computers in Human Behavior*, 27(1):203-208.
- [11] A. H. Tangari. (2010). The framing of psychological distance: how multiple psychological distance dimensions influence consumer responses to sustainability messages. *Dissertations & Theses-Gradworks*.
- [12] A. O'Cass. (2000). An assessment of consumers product, purchase decision, advertising and consumption involvement in fashion clothing. *Journal of Economic Psychology*, 21(5):545-576.
- [13] H. S. Jung, H. H. Yoon. (2012). Why do satisfied customers switch? Focus on the restaurant patron variety-seeking orientation and purchase decision involvement. *International Journal of Hospitality Management*, 31(3):875-884.
- [14] S. M. Hailpern, P. F. Visintainer. (2003). Odds ratios and logistic regression: further examples of their use and interpretation. *Stata Journal*, 3(3):213-225.
- [15] J. J. Brown, P. H. Reingen. (1987). Social Ties and Word-Of-Mouth Referral Behavior. *Journal of Consumer Research*, 14(3):350-62.
- [16] J. H. Park, B. Gu, and P. Konana. (2012). The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, 23(1):182-196.
- [17] K. Clarke, R. W. Belk. (1979). The effects of product involvement and task definition on anticipated consumer effort. *Advances in Consumer Research*, 6(1):313-318.

Privacy as a Commodity Is Not the Case: Privacy Calculus

Model for Connected Cars

*Minho Ju¹, Jian Mou^{*2}*

¹ Department of Interaction Science, Sungkyunkwan University

² School of Economics and Management, Xidian University, Xi'an, Shaanxi, China 710126

Abstract: With the development of information and telecommunication technology, more and more products can integrate such technologies to provide more convenient service to consumers. The connected car is the presence of devices in an automobile that connect the devices to other devices within the vehicles and or devices, networks and services outside the car including other cars, home, office or infrastructure. The data generated by telematics and vehicle infotainment systems is highly revealing of personal lifestyles, habits and preferences include customer account. The purpose of the paper is to figure out effect of social behaviors of connected car on social adjustment as partner, and how this relationship affects determining willingness to provide personal information of connected car consumers. To do so, a research model and the hypothesis have been developed accordingly. The model will be tested by using the sample from South Korea's care driver. The potential contributions have been addressed consequently.

Keywords: connected car, privacy, social behavior, social adjustment, privacy calculus model

1. INTRODUCTION

The connected car is the presence of devices in an automobile that connect the devices to other devices within the vehicles and or devices, networks and services outside the car including other cars, home, office or infrastructure. Connected car is currently defined as mutual and simultaneous relationship among the portal at the automotive company, the vehicle, the communication link between the vehicle and the portal^[20]. According to NHTSA, "level 2" automation for driving is now available with automatic sensing technologies such as LIDAR, distance sensors, position estimator in the aspect of usage of connected car, which means semi-auto drive based on driver's characteristic.

The data generated by telematics and vehicle infotainment systems is highly revealing of personal lifestyles, habits and preferences include customer account. Vehicle performance, driver behavior, biometrics and health, location, personal communications, use of features and applications^[13]. These private data have been considered as not an absolute right but is subject to economic principles of cost-benefit analysis and trade-offs. If so privacy could be treated as tradable commodity. When benefits exceed costs, users are willing to give their information to the company or service provider both individual and societal levels.

Previous studies of e-commerce on privacy calculus model could shed light on further study of privacy and connected car in the aspect of network connectivity, and range of benefits/concerns. However, social factor should be additively covered in the privacy-calculus model of the connected car due to CAS (Computers as Source).

The purpose of the paper is to figure out effect of social behaviors of connected car on social adjustment as partner, and how this relationship affects determining willingness to provide personal information of connected car consumers. The study will be based on pseudo experiment for driver perception of connected car social behavior experience, and after survey will be conducted based on structure equation modelling (SEM).

This study is organized as follows: in next section, the literature review of connected car and the term of

* Corresponding author. Email: jian.mou@xidian.edu.cn (Jian Mou)

computers as sources have been proceeded. In addition, the hypothesis has been developed. Thereafter, we introduced the research methodology that will be employed in this study. Finally, the discussion and potential contributions have been outlined.

2. LITERATURE REVIEW

2.1 Connected car

The connected car can be described as a vehicle with one or more external wireless communication possibilities, which connects the vehicle to an external network (see Figure 1). The requirement of external wireless communication distinguishes the connected car from other vehicles where internal connections already exists, e.g., the On-Board Diagnostics II (OBD-II) interface used for wired vehicle diagnostics or the USB ports that are becoming more and more common ^[12]. The connected car consists of three domains of vehicle, consisting of the in-vehicle network and ECUs, the portal at the automotive company, delivering services to the vehicle, and The communication link between the vehicle and the portal ^[20].

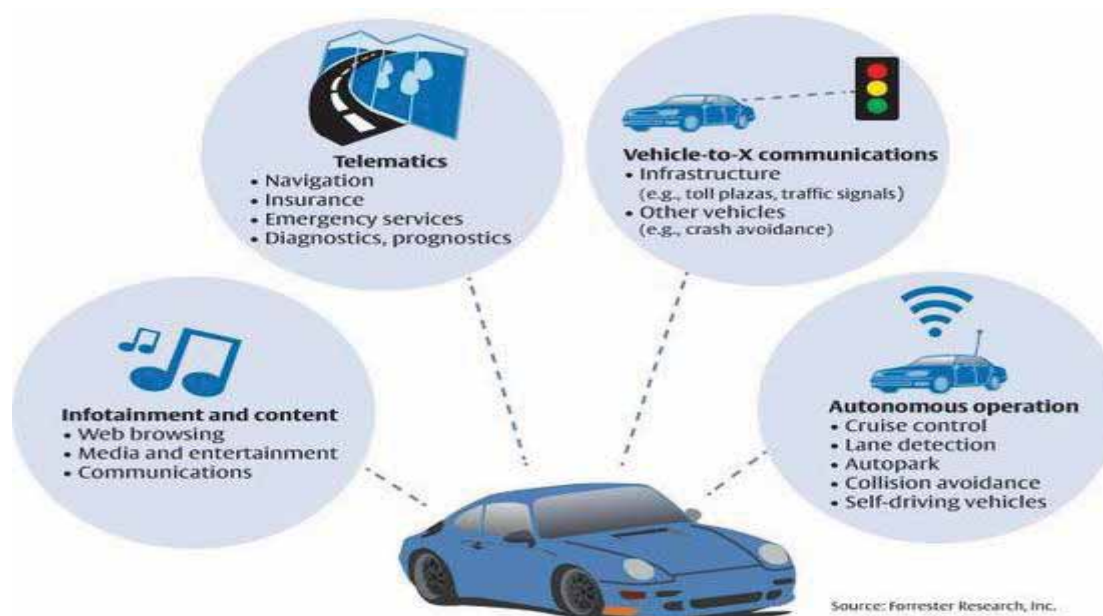


Figure 1: The structure of connected car (source: Forrester Research, Inc.)

The leading IT companies such as Apple, Alibaba, and Google struggle to take the first positioning on the connected car market. It is expected that Apple's connected car interface "iDash" will resemble touch screen dashboard of iPad with cameras for head and eye tracking, lasers for light, and voice actuation for hands free adjustments. Under the new system a driver will be able to control not only the radio and the temperature, but the headlights, wipers, mirrors, and much more [10]. Senior managing consultant at NTT DATA described car company side of connected car, automated driving calls for access to the extensive sensor data of the vehicle. He asserted that connected car would allow drivers benefits in security, efficiency, and comfort. For instance, long range radar, laser scanner on front and rear view camera would allow driver prediction of traffic and road conditions on voice or image on highly automated driving. Moreover, dynamic route planning significantly saves time by avoiding traffic jams. When these vehicles share data about their destinations with other cars, city traffic management roads would be much more efficiently used which help us better using resources. ESP and lane keeping give tactile feedback, and gesture based control would be available on the modern cockpits of

connected car [7].

However, as consumers demand more personalized services and smarter, safer, and more seamless in-car features that means more data and more questions about privacy. The data generated by telematics and vehicle infotainment systems is highly revealing of personal lifestyles, habits and preferences. In addition to customer account data and vehicle performance data, it includes driver behavior data, biometrics and health data, location data, personal communications (voice, text, email, and social networking), web browsing data, personal contacts and schedules, use of features and applications, and choice of music, radio and other streamed audio or video content ^[13]. Some of these data is personal data, deserving privacy protections therefore data and questions about privacy should be covered in using connected car. Drivers give their private information to get better perceived ease of use like personalization of service, financial rewards (insurance). In this purpose aspect connected car using experience has similarity with e-commerce using experience because people share their private information to gain personalized services. Furthermore, possible privacy concerns such as large amounts of data include private data is collected by service provider or telecommunication company, and unauthorized secondary use and possibility of hacking is another similarity. In these aspects, e-commerce private information sharing case could be adapted toward connected car private information sharing.

2.2 Computers as Sources

Connected car can be considered as computers as source (CAS) due to the fact that humans apply the same social heuristics toward inanimate machines as they do in human-human interaction. Computers simply generate output as programmed, they seem to become oblivious to the asocial nature of interaction and rather automatically apply the same social heuristics toward inanimate machines as they do in human-human interaction ^[19].

When computers emulate human-like attributes, such as interactivity and speech, people tend to focus on those cues, and fail to take into account the asocial nature of the interaction. Interactive cues of connected car like head-body tracking and voice actuation triggers driver to adapt human-human interaction rules. These interactive cues seem promising in their ability to cue cognitive heuristics pertaining to credibility assessments because they are all structural features that underlie the design aspects or surface-level characteristics associated with powerful first impressions of connected car credibility. Moreover, when the computer first provided some information about its technical capacity (self-disclosure), participants' responses were more intimate than when the computer simply asked the same questions without revealing about itself. This triggers spontaneously social behaviors grounded in interpersonal relationships. After receiving useful information from the computer to accomplish an experimental task, participants spent more time on the second task to "help" the computer than when the computer had offered them irrelevant information ^[9]. Depending on how particular affordances and information manifest itself to drivers, those can lead to positive or negative outcomes. In short social behavior of computer i.e., self-disclosure, giving useful information, positive feedback affect participants consider computer as a partner, which fits with definition of social adjustment of the privacy calculus model.

2.3 Privacy cost

Privacy is defined as the ability of individuals to decide when, what, and how information about them is disclosed to others ^[8]. There have been two opinions about privacy, one is people have the right to be let alone, and the other is individuals should not have a right to conceal facts about themselves ^[28]. The latter asserts that privacy is not an absolute but can be assigned an economic value and traded for goods and services. Value-based definitions of privacy argue that a call for greater privacy is, fundamentally, antagonistic to the political economy of the information markets ^{[5] [14] [21]}. In this view, privacy is not an absolute right but is subject to the

economic principles of cost-benefit analysis and trade-offs ^[11]. Bennett ^[3] further formalized the notion of privacy as a tradable commodity. It can be assigned an economic value and can be considered in a cost-benefit calculation at both individual and societal levels.

Privacy concerns is receiving increased attention due to the huge amount of personal information being collected, stored, transmitted and published on the internet. Smith et al. ^[23] identified four dimensions of an individual's concern about privacy, namely: (1) collection, (2) errors, (3) unauthorized secondary use and (4) improper access. The four factors provide a framework to explain the concerns for information privacy ^[24]. Due to the fact that most data from connected car is based on actual driving and map, inaccuracy possibility could be reduced. However, still large amounts of personally identifiable data are being collected by company, companies could use personal information for undisclosed purposes, and companies are able to fail to protect consumer's privacy. This brings us Knightian uncertainty from unknown purpose of large data collection, privacy risk belief that when private information is revealed, and privacy protection belief. Consumer privacy concerns vary dramatically by information type. Consumers are likely to avoid revealing personal information that may identify themselves to companies in exchanges for values or services that companies would provide ^[23].

Privacy risk is the degree to which an individual believes that a potential for loss is associated with the release of personal information to an entity ^[17]. The construct is related to the potential loss of control over personal information ^[6]. In the privacy literature, privacy risk has been treated as multidimensional construct. Risky events are those for which individuals can assign a known or knowable probability to each possible outcome. Ambiguous events are those for which such assignment is impossible, either because the probabilities are unknown/unknowable or because one or more possible outcomes are not known/knowable, and therefore are characterized by a qualitatively different type of uncertainty. Hence, the term uncertainty is reserved to denote situations characterized by quantifiable risk and unquantifiable risk like Knightian uncertainty. Knight stated the word 'risk' to describe the "measurable uncertainty", where the possible outcomes are known and they can be categorized in groups with assigned probabilities "either through calculation a priori or from statistics of past experience". The 'true' uncertainty, on the other hand, applies to situations where no probability can be computed, as agents do not have the information necessary to assign a probability measure "because the situation dealt with is in a high degree unique" ^[29]. Through past literature we believe that privacy concern, privacy risk and Knightian uncertainty are the dimension of privacy cost that may negatively influence the willingness to provide personal information in connected car context. Thus the following hypothesis have been proposed:

H1. Privacy concern will have a negative effect on willingness to provide personal information of connected car users.

H2. Privacy risk will have a negative effect on willingness to provide personal information of connected car users.

H3. Knightian Uncertainty from collecting large amounts of data will have a negative effect on willingness to provide personal information of connected car users.

2.4 Privacy rewards

Three main types of benefits of information disclosure have been assumed according to the prevailing literature review on the privacy calculus. Financial rewards, personalization, and social adjustment benefits ^[18]. Though perceived usefulness still has an important value, drivers currently use navigation systems which enable them the fastest way to the destination with estimated time. Furthermore, perceived usefulness can be divided into quality values such as conciseness, comparability, timeliness and these values strongly connected affordances such as interactivity ^[25]. Secondary services from relationship could create the additional perceived

ease of use or convenience, but this paper would focus on the effect of direct connected relationship among domains of the connected car.

The effect of financial benefits to attract customers to reveal their personal information has been proven, and the effect on connected car is predictable with UBI-Usage-Based Insurance or discount on purchasing car when using connected car service on subscription. Also, time or cost savings from semi/full-auto drive could be considered as possible financial rewards from connected car. Liu et al. ^[15] found that personalized services played a significant moderating effect on the relationship between users' disclosed information and their perceived benefits. The value consumers place on personalization was almost twice as influential as their concerns for privacy when determining usage of personalization services ^[4]. Social adjustment benefits (defined as the establishment of social identity by integrating into desired social groups) can also have an effect on willingness to disclose personal information ^[16]. Social adjustment benefits could be considered in the aspect of computers as sources (CAS) theory on connected car cases. Trust, partnership with the machine from receiving useful information on time with positive feedback could be different from e-commerce cases. We therefore hypothesises:

H4. Financial rewards will have a positive effect on willingness to provide personal information of connected car users

H5. Personalization will have a positive effect on willingness to provide personal information of connected car users.

H6. Social adjustment will have a positive effect on willingness to provide personal information of connected car users.

Based on above discussion, we considered the dimension of disclosure privacy rewards as financial rewards, personalization and social adjustment benefits; while the disclosure privacy costs include privacy concern, privacy risk and Knightian uncertainty. In addition, we also considered gender, age, driving experience, past invasion of privacy and media exposure as control variables. The research model is shown in Figure 2.

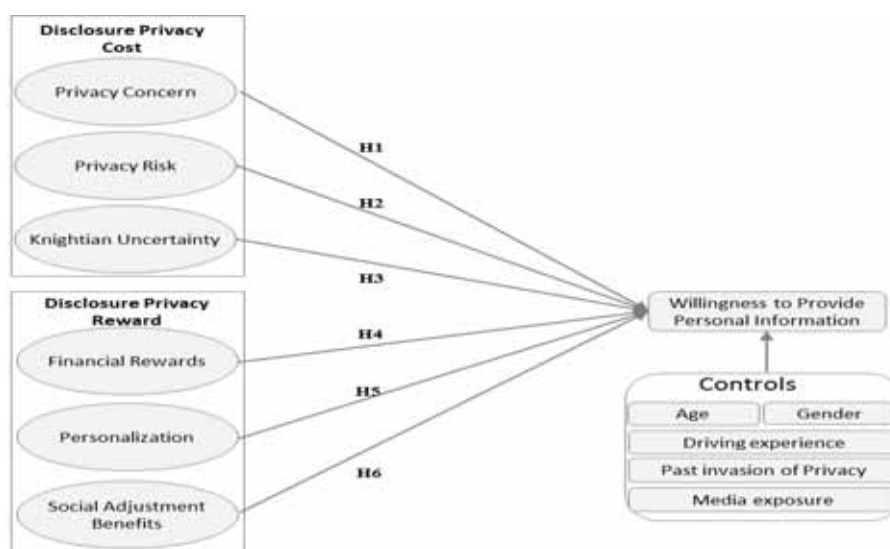


Figure 2. Research Model

3. RESEARCH METHODOLOGY

Human computer communication with human-like attribute and social behavior is the significant difference of connected car from an e-commerce experience. The difference will be tested as treatment condition to figure out its effect toward social adjustment and willingness to give private information. Survey study will be

employed after the treatment condition. Survey study that included items for the constructs specified in the model will be employed in this research. Data ($N = \text{more than } 300$) will be collected via an online survey in South Korea. The target population is adult drivers in South Korea. This is admittedly only a small subset of a possibly much larger population of interest. For example, it would be interesting to collect samples drawn from an international population, because international socio-cultural differences might influence user behavior. Due to resources availability and time constraints, however, we leave these developments to future research. Structural equation modelling such as PLS-SEM will be employed as a data analysis approach because of its ability to account for measurement errors for unobserved constructs and to simultaneously examine the predictive relationships among them^[22]. The hypotheses will be tested by examining the structural models. The explanatory power of a structural model could be evaluated by looking at the R^2 value in the final dependent construct. Specifically, we follow the two-step approach suggested by Anderson and Gerbing^[2]. As such, we first analyze a measurement model to assess the measurement quality of constructs by using a confirmatory factor analysis (CFA) approach. Subsequently, we estimate a structural model to test the research hypotheses. PLS-based SEM to evaluate both the measurement and structural models will be utilized. Specifically, we will use the statistical software packages SmartPLS 3.2.1. The bootstrap technique is well-known and commonly used in PLS analysis to estimate the significance of weights, loadings and path coefficients.

Table 1: Survey Instrument Details

Variable	Sample Statement	Source
Financial rewards	An offer of a financial reward will decrease concerns about self-disclosure.	Andrade et al. (2002)
Personalization	I value the connected car that are personalized for my usage experience preferences	Chellappa and Sin (2005)
Social adjustment	Friend's suggestions and recommendations will affect my decision to use the connected car.	Tan et al. (2012); Taylor and Todd (1995),
Knightian Uncertainty	I have often only vague and limited knowledge of the actions to protect (or give away) my personal information.	Acquisti et al. (2005)
Privacy risk	What do you believe is the risk for connected car drivers due to the possibility that: Records of driving information could be sold to third parties	Dinev and Hart (2006)
Privacy concern	I am concerned that the information I submit on the connected car could be misused.	Dinev and Hart (2006)
Willingness to provide personal information	To what extent are you willing to provide personal information? Driving a connected car requires me to provide highly personal and password-protected financial information (e.g., driving history and GPS data)	Dinev and Hart (2006)

4. DISCUSSION AND POTENTIAL CONTRIBUTIONS

The purpose of the paper is to figure out the effect of social behaviors of connected car on social adjustment as partner, and how this relationship affects determining willingness to provide personal information of connected car consumers. To do so, a research model and the hypothesis have been developed accordingly. The model will be tested by using the sample from South Korea.

This study will contribute both in theoretical and practical aspects in several ways. Theoretically, we classified the privacy cost as the dimension of privacy concern, privacy risk and Knightian uncertainty. Given the connected car systems need collect driver's personal information, privacy risk and the potential risk may rise that can negatively influence driver's behavior. Further, we defined three dimensions as privacy reward. Rewards may always promote an individual to take an action such as usage. In connected car context, we

specify financial rewards, personalization and social adjustment benefits as the potential rewards that may positively influence connected car user's willingness. This study can also help connected car product related company to better understanding how to increase an individual provide his/her information. Only the connected care systems collect enough information from individuals, it can provide better services to car drivers. In addition, firm also can consider provide privacy rewards to achieve their purposes such as collect more individual data.

ACKNOWLEDGEMENT

This study supported by the Fundamental Research Funds for the Central Universities of No. JBX180604; and the Chinese Ministry of Education Humanities and Social Sciences project of No. 16YJC630038; and the National Natural Science Foundation of China of No. 71573199.

REFERENCES

- [1] Acquisti A, Heinz, H J, Grossklags J. (2005). Uncertainty, ambiguity and privacy. The Workshop on the Economics of Information Security (pp.2--3).
- [2] Anderson J C., Gerbing D W. (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411-423.
- [3] Bennett C. (1992). *Regulating privacy*. Ithaca, NY: Cornell University Press.
- [4] Chellappa R K, Sin R G. (2005). Personalization versus privacy: an empirical examination of the online consumer's dilemma. *Information Technology & Management*, 6(2-3), 181-202.
- [5] Cohen J E. (2001). Privacy, ideology, and technology: a response to jeffrey rosen. *Georgetown Law Journal*, 89(6), 2029-2045.
- [6] Dinev T, Hart P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research* 17(1), pp. 61-80.
- [7] Dmoch, T. (2015). Connected Car could be the ultimate Human-Machine-Interface. Retrieved September 27, 2017, from <https://www.linkedin.com/pulse/connected-car-ultimate-human-machine-interface-thomas-dmoch>.
- [8] Duri S, Gruteser M, Liu X, Moskowitz P, Perez R, Singh M., Tang J M. (2002, September). Framework for security and privacy in automotive telematics. In *Proceedings of the 2nd international workshop on Mobile commerce* (pp. 25-32). ACM.
- [9] Fogg B J, Nass C. (1997, March). How users reciprocate to computers: an experiment that demonstrates behavior change. In *CHI'97 extended abstracts on Human factors in computing systems* (pp. 331-332). ACM.
- [10] Hughley J. (2015). Apple's Connected Car Interface. <https://www.linkedin.com/pulse/apples-connected-car-interface-john-hughley>.
- [11] Johnston C. (2012, February 8). Google paying users to track 100% of their Web usage via little black box. <http://arstechnica.com/gadgets/2012/02/googlepaying-users-to-track-100-of-their-web-usage-via-little-black-box/> [1 May 2012].
- [12] Kleberger P, Olovsson T, Jonsson E. (2011, June). Security aspects of the in-vehicle network in the connected car. In *Intelligent Vehicles Symposium (IV)*, 2011 IEEE (pp. 528-533). IEEE.
- [13] Lawson P, McPhail B, Lawton E. (2015) *The Connected Car: Who is in the Driver's Seat? A Study on Privacy and On board Vehicle Telematics Technology*. British Columbia Freedom of Information and Privacy Association, Vancouver, 2015, p. 32.
- [14] Lessig L. (2000). Code is law: On liberty in cyberspace. *Harvard Magazine*, (January-February), 1-2.
- [15] Liu Z, Bonazzi R, Fritscher B, Pigneur Y. (2011). Privacy-friendly business models for location-based mobile services. *Journal of theoretical and applied electronic commerce research*, 6(2), 90-107.

- [16] Lu Y, Tan B C Y, Hui K L. (2004). Inducing Customers to Disclose Personal Information to Internet Businesses with Social Adjustment Benefits. *International Conference on Information Systems, Icis 2004*, December 12-15, 2004, Washington, Dc, USA (pp.571-582). DBLP.
- [17] Malhotra N K., Kim S S, Agarwal J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research*, 15(4), 336-355.
- [18] Moloney M, Poti V. (2013). A behavioral perspective on the privacy calculus model. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2310535.
- [19] Nass, C., & Moon, Y. (2000). Machines and mindlessness: social responses to computers. *Journal of Social Issues*, 56(1), 81–103.
- [20] Nilsson D K, Larson U E, Jonsson E. (2008). Creating a Secure Infrastructure for Wireless Diagnostics and Software Updates in Vehicles. *International Conference on Computer Safety, Reliability, and Security (Vol.5219, pp.207-220)*. Springer-Verlag.
- [21] Posner R A. (1978). Economic theory of privacy. *Regulation*, 2, 19.
- [22] Rigdon E E, Schumacker R E, Wothke W. (1998). A comparative review of interaction and nonlinear modeling. *Interaction and nonlinear effects in structural equation modeling*, 1-16.
- [23] Smith H J., Milberg S J, Burke S J. (1996). Information privacy: measuring individuals' concerns about organizational practices. *MIS Quarterly*, 20(2), 167-196.
- [24] Stewart K A, Segars A H. (2002). An empirical examination of the concern for information privacy instrument. *Information Systems Research*, 13(1), 36-49.
- [25] Sundar S S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility.
- [26] Tan G W H, Ooi K B, Sim J J, Phusavat K. (2012). Determinants of mobile learning adoption: An empirical analysis. *Journal of Computer Information Systems*, 52(3), 82-91.
- [27] Taylor S, Todd P A. (1995). Understanding information technology usage: A test of competing models. *Information systems research*, 6(2), 144-176.
- [28] Warren S D, & Brandeis L D. (1985). The right to privacy. *Harvard Law Review*, 4(5), 193-220.
- [29] Knight, F. (1921). *Risk, uncertainty and profit*. Boston, MA: Hart, Schaffner & Marx.

Considering correlation retarded growth for personalized recommendation in social tagging

Xuwei Pan^{1}, Ling Ding², Leimin Ye²*

¹Department of Qixin Honors & Entrepreneurship, Zhejiang Sci-Tech University, Hangzhou, China

²Department of Management Science and Engineering, Zhejiang Sci-Tech University, Hangzhou, China

Abstract: Due to the massive amounts of data, finding social media suited to their need is a challenging issue. To help such users retrieve useful social media content, we propose a new model of personalized recommendation system by using annotating information from relationship among users, tags, and items. However, the frequency of users' tagging has strong or weak correlation, which affects the dynamic interest mining of users. In this paper, CRGI is proposed to describe the correlation between users and tags or tags and items. Our approach has two phases, in the first phase, we describe the correlation between users, items and tags by CRGI and in the second phase, we build a tag-item weight model and a user-tag preference model on the basis of the first phase. Then we utilize the two models to find the suitable items with the highest scores. The experimental results demonstrate that the item recommendation performance is improved in both the accuracy and the diversity, and validate that the proposed personalized approach is effective for improving the social media recommendation.

Keywords: social tagging, collaborative tagging, personalized recommendation, correlation retarded growth

1. INTRODUCTION

Nowadays, there are huge amount of information that are available for users. The expansion of information has become a double-edged sword. Users can not only get more abundant contents, but also spend a long time to find the contents needed. Researchers started to utilize social tagging known as Folksonomy to solve the problem of finding the suitable resource for each user query according to his own taste. In social tagging system, users are allowed to add one or more tags to items freely. Due to that flexibility in Folksonomy, it can be used as a good tool to organize and share items on the web.

As a core component of Web 2.0 ^[1], social tagging is consists of users, tags and items. These three elements are combined together to form a complex dynamic network. The statistical properties and evolution mechanism of the network are helpful to reveal the structural characteristics and behavior characteristics ^[2]. Some studies, Yang et al. ^[3] analyzed a large number of data in Folksonomy experimentally and found a pile of tags contain much information to identify the core ideas of Multimedia content. In a word, the tag has two functions in social tagging system: 1) the tag can help users organize and manage content, and 2) users use tags to discover the similar content that others share ^[4]. Therefore it is very crucial to studying collaborative tagging as a tool to improve personalization in social media recommendation.

User's tagging activities show the relationship among users, tags and items. And the closeness of the relationship is affected by many factors, such as annotation frequency, tagging time. The acquisition of users' preference is directly affected a lot without an accurate description of tight correlation. Tagging behavior is the concentrated embodiment of user's participation in annotation and interactive behavior. However, there is a correlation upper limit of user's annotation ability ^[5]. So it is not true to recognize the correlation between user-resource-tag is infinitely increasing proportionally with the raise in the number of annotations. This paper introduces the logistic population growth model ^[6] to construct an index of correlation retarded growth (CRGI)

* Corresponding author. Xuwei Pan, Email: panxw@zstu.edu.cn

to convert the labeling frequency of the user, tag and resource, which provides a favorable basis for next work. We compute first the tag-item weight model with respect to similar tags and the user-item weight model with respect to similar items. Then we combine both weights to give a personalized recommendation result according to anyone's own preference.

2. LITERATURES REVIEWING

2.1 social tagging system

At present time, there exists so many popular social tagging systems, such as Delicious, Flickr, You tube, etc. Users can not only organize and manage items, but also acquire items of interest according to the sharing of other users. In the Internet, the a large number of annotation links not only users, tags and items, but also forge links between one and the other user, so as items and tags^[7]. Among the multimedia content, there are two ways to mark them^[8]: manual annotations, which are edited entirely by the user with appropriate words; semi-automatic tagging, social tagging systems that recommend the appropriate list of tags for the user to select. Social tagging has become an indispensable tagging behavior in the modern network. More and more researches have been conducted, mainly focusing on the patterns and characteristics of social tagging, usage characteristics of social tags and the recommendation model based on social tagging^{[9]-[15]}.

2.2 tag based recommendation

The widely used approach is forming weighting vector to build user interest model with tags. Centering on the information such as the tagging frequency, number of times, characteristic and so on from users, it is easy to calculate the laws of these information by using probability computing model, etc. Researchers regard folksonomy as a graph-based perspective to provide personalized service. Some of these studies use a tripartite hyper-graph in which each hyper edge connects a user, tag and items^{[16]-[17]}. Hotho et al.^[16] proposed graph based algorithms. It is basic thought is making use of potential information between tags and resource, tags and users. Nonetheless, the algorithm doesn't consider the personalized message of tags well so that can't recommend personalized tags to different users. Firan et al.^[18] put forward a music recommendation system basing on tags which expresses user's preference by frequency of tags. The tags not only reflect types and characteristic of music, but also reflect the user's preference for music. Symeonidis et al.^[19] built a new model to recommend basing tensor decomposition which represented social tagging data-set as three-order tensor and then use higher-order singular value decomposition (HOSVD) to mine the potential semantic among users, tags and resource. Yeung et al.^[20] built user model with frequent tags set that used just tag co-occurrence. The model reflects the varying degree of users among different interests.

The above researches all use annotation frequency to express the tightness between users, tags and items. However, the ability of users to annotate items changes at a variable rate over time. In this paper, CRGI (correlation retarded growth index) is proposed to describe the correlation between them. Integration of this indicator into a tag-based recommendation model would get better recommendation results.

3. CORRELATION RETARED GROWTH INDEX

The unit nodes of the users, tags and items in the social tagging system are the indicators that reflect the correlation characteristics among these elements. The association feature index establishes the connection between the user, label and item according to the associated activity. The degree and intensity of this connection are affected by many factors, such as tagging frequency, etc. We find that the rate of change of user tagging ability is similar to that of Logistic population growth model, so we construct the indicator to describe the correlation between users and tags and between tags and items. The establishment of correlation indicators can measure the close relationship between them more accurately.

The correlation index is used to describe the degree of correlation between user U_u and tag T_t , item I_i

and tag T_t . Take the correlation index $a'_{u,t}$ between user U_u and tag T_t as an example. The correlation index has the following characteristics:

- (1). The number of related activities $a_{u,t}$ between user U_u and tag T_t is independent variable, the correlation index $a'_{u,t}$ is dependent variable. The data range of $a_{u,t}$ is natural numbers N , that means $a_{u,t}=0,1,2,\dots$. In order to simplify the calculation, we normalize the result of $a'_{u,t}$, that is the data range of $a'_{u,t}$ is $[0,1]$.
- (2). $a'_{u,t}$ is the increasing function of $a_{u,t}$ that is the more number of related activities between user U_u and tag T_t , the higher the degree of correlation.
- (3). When $a_{u,t}=0$, $a'_{u,t}=0$. In the initial stages of $a_{u,t}$ changes from 1, $a'_{u,t}$ grows faster. Then $a'_{u,t}$ grows slowly. Finally, after a certain stage, $a'_{u,t}$ gradually stops increasing. Obviously, $a'_{u,t}$ is not the linear function of $a_{u,t}$.

Logistic retarded growth model was discovered by Pierre Franois Verhulst in 1844-1845 while he was studying the law of population growth. The model is near exponential growth in the initial stages, and then its growth slows with the saturated increasing. At last, the growth stops after maturity. Logistic function has a typical statistical significance, its standard form is:

$$f(x) = \frac{L}{1+e^{-r(x-x_0)}} \quad 3-(1)$$

The growth rules of correlation index $a'_{u,t}$ are very similar to Logistic function. The related activities between user U_u and tag T_t are similar to the population growth in bio-system. Its correlation is not only influenced by the number of related activities, but also has upper limit that is the correlation could not increase infinitely. So we use the variant of Logistic function to define the correlation index $a'_{u,t}$.

$$a'_{u,t} = \begin{cases} 0 & , a_{u,t} = 0 \\ \frac{1}{1+e^{-r(a_{u,t}-a_0)}} & , a_{u,t} > 0 \text{ and } a_{u,t} \in N \end{cases} \quad 3-(2)$$

The r in the function is growth rate, the larger the r is, the faster the growth of $a'_{u,t}$ is. The a_0 is defined as the median of dependent variable, when $a_{u,t} = a_0$, $a'_{u,t} = 0.5$. r and a_0 can be defined as different values according to different application scenarios. The definitional domain of standard Logistic function is $(-\infty, +\infty)$, the range is $[0,L]$. The function 3-(2) limits the definitional domain to natural numbers N , and limits range to $[0,1]$ by setting up $L=1$. In the standard Logistic function, when x tends to $-\infty$, $f(x)$ tends to 0.

4. PROPOSED MODEL

The recommendation system recommends items which users may interested according to recommendation algorithm and user model. We propose a reference algorithm by analyzing the relationship among users, items and tags. As shown in the Figure 1:

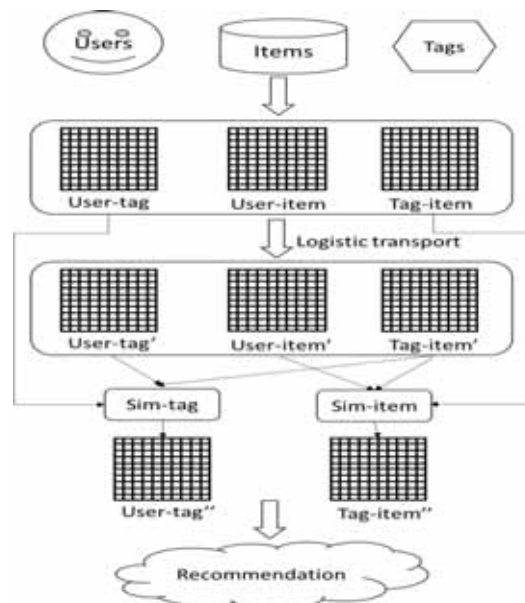


Figure 1. The algorithm process of the users, tags and items

4.1 Tagging system decomposition

A tag system is a three-dimensional system consisted of user, tag and item. Users can label an item with multiple tags while the same tag can be used to label multiple items. Obviously, tag is a link between the user and the item. All users in the network application system form the user set $U = \{u_1, u_2, u_3, \dots, u_{|U|}\}$. All items in the network application system form the item set $I = \{i_1, i_2, i_3, \dots, i_{|I|}\}$. All tags in the network application system form the tag set $T = \{t_1, t_2, t_3, \dots, t_{|T|}\}$. $|U|, |I|, |T|$ represent the number of users, items and tags. The U represents the users who can select and use the items in I under the tag in T . Because of the relationship among user, item and tag, U, I, T form into a three-dimensional matrix which is complex to make recommendation. We reduce the three-dimensional matrix to three two-dimensional matrices.

The User-tag matrix $A = [a_{u,t}]_{|U| \times |T|}$ represents the number of items chosen by U_u with T_t . The User-item matrix $B = [b_{u,i}]_{|U| \times |I|}$ represents that if the users U_u tags item I_i , marking “1” and otherwise marking “0”. The Tag-item matrix $C = [c_{t,i}]_{|T| \times |I|}$ represents the number of users who choose item I_i with T_t .

To compute the degree of correlation between item I_i and tag T_t , we employ CRGI in section 3. Figure 2. illustrates the calculation process of building the correlation index.

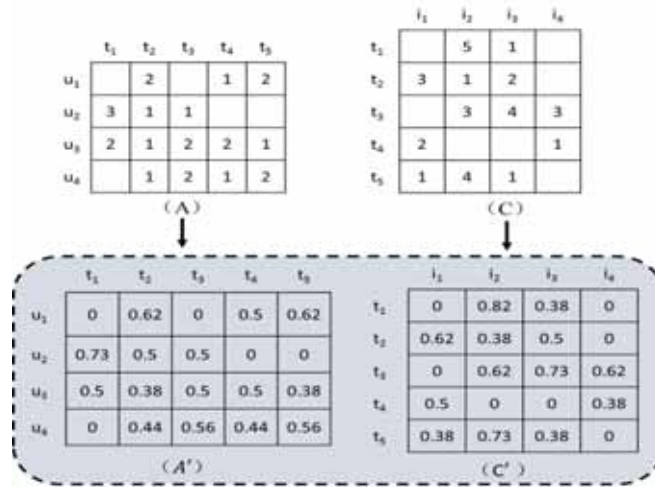


Figure 2. Concrete numerical demonstration of Logistic transformation

4.2 Similarities analysis

Computing the similarity between tags or items is an important part. In the tagging system, items and items are linked together by users and tags. By projecting into the tag space, we can get user-based item-item relational network, and by projecting into the user space, we can get tag-based item-item relational network. Therefore, the similarity between items depends on the user-based item-item similarity N_{user} and the tag-based item-item similarity N_{tag} . There are many algorithms to compute similarities including cosine similarity, adjusted cosine similarities and Pearson correlation coefficient. In the paper, we use cosine similarity which has obvious advantage process huge data with high speed.

In order to compute the Item-Item similarity matrix $N_{|I| \times |I|}$, we divide the process up into three steps. Frist, we utilize the User- Item correlation matrix B' by computing the weights of users tag items. Then we use the cosine similarity according to 3-(4) to compute the user-based item-item similarity N_{user} and the tag-based item-item similarity N_{tag} respectively. Finally, we define the item-item similarity according to 3-(3):

$$N_{(i_x, i_y)} = \lambda N_{\text{user}}(i_x, i_y) + (1 - \lambda) N_{\text{tag}}(i_x, i_y) \quad 3-(3)$$

where i_x and i_y represent the row and column of item i_x and item i_y , $\lambda \in (0, 1)$. In this article, the value of λ is set of 0.5 due to the fact that the two types of cosine similarities follow a similar distribution.

$$n_{i_x, i_y} = \cos(i_x, i_y) = \frac{i_{x, i_y}}{\|i_x\| \cdot \|i_y\|} = \frac{\sum_{s=1}^{|I|} (n_{i_x, s} \cdot n_{i_y, s})}{\sqrt{\sum_{s=1}^{|I|} (n_{i_x, s})^2} \cdot \sqrt{\sum_{s=1}^{|I|} (n_{i_y, s})^2}} \quad 3-(4)$$

s represents a particular tag, and $|T|$ represents the number of tags. If the similarity value jumps to the top of k among the similarities in a given column, we keep the similarity value. Otherwise the similarity value is set to 0.

As a result, get the N^k similar items, where k is the number of the top Item-Item similarities.

we follow the same steps as in computing the Item-Item similarity matrix N to compute Tag-Tag similarity matrix M . In the end, we can get the M^k similar tags, where k is the number of the top Tag-Tag similarities.

4.3 Building the recommendation model

We build two models to reflect the user-item relationship in a given tag. The first model called user-tag preference model reproduces the answer to the following question:

(1)How to judge the level of interest of users in similar tags to a specific tag?

The model computes the users' potential preference for tags denoted as $E_{|U| \times |T|}$, which is deduced by the result of User-tag correlation matrix A' and Tag-tag similarity matrix M as shown:

$$E_{|U| \times |T|} = A' \times M^k \quad 3-(5)$$

The model shows that the potential impact from the similar tags to a tag. The result reflects that the degree of users' favor to tags as shown in Figure 3.

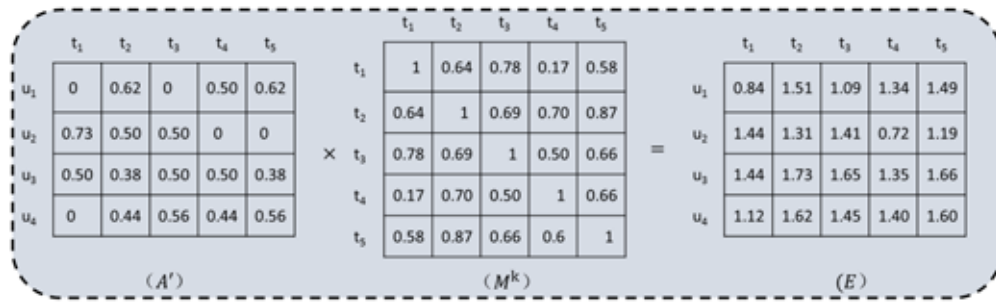


Figure 3. An illustration of the process of computing the user-item matrix E

The second model reproduces the answer to the following question:

(2)How to judge the connection degree between tags and items?

The model called tag-item weight model and computes the connection degree between tags and items denoted as $F_{|T| \times |I|}$, which is deduced by the product of Tag-item correlation matrix C' and Item-item similarity matrix N . The model shows that the potential impact from the similar items to an item. The result reflects that the degree of connection between tags and items as shown in Figure 4.

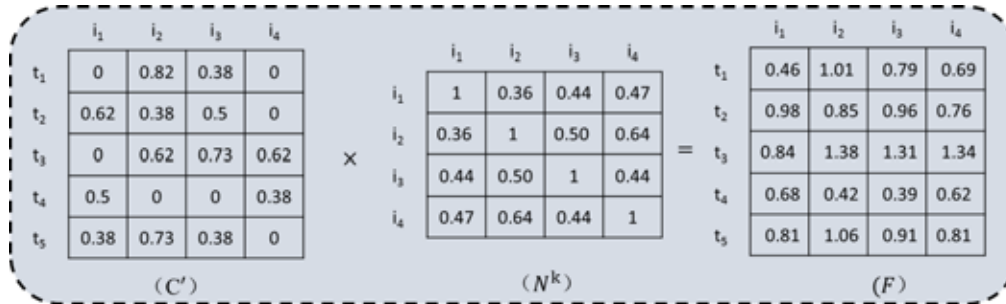


Figure 4. An illustration of the process of computing the tag-item matrix F

Our goal is getting the preference degree of users to items in a specific tag. We build the final model shown as user-item matrix which is computed by combining the two models E and F . We call our final model as Tag-Based-Recommend (TBR). For a given set of tags in a query q , $q = \{t_1, t_2, \dots, t_n\}$ $n \leq |T|$, the relevance score of item I_y for user U_x can be computed as:

$$TBR_U(i, t) = \sum_{t \in q} E_{u,t} \times F_{t,i} \quad 3-(6)$$

By utilizing both models, E and F , items that fit a user's needs rank higher in the recommendation list.

5. EVALUATED MEASUREMENTS

In this section we describe the experiments conducted. In the first phase, we present how the user data are used to train a model and corresponding results. In the second phase, we evaluate and compare the performance

of the different recommendation methods. Traditional item based collaborative filtering method is compared to explore that weather annotating information from relationship among users, tags, and items makes contribution to recommendation. Besides, we compare our approach with ICM approach .The ICM approach ^[21] connects users, tags and items through the number of tagging directly, which ignores the tagging ability of the user changes over time by varied rate.

In this experiment, a MovieLens dataset is used to evaluate our algorithms. It contains 35163 tags applications across 19545 movies. And it is created by 7801 users between December 24, 2005 and March 31, 2015. During the testing phase, it is divided into two groups containing training set and testing set. We withhold 20% records for testing while the remaining 80% is used for training. In order to make the experimental results more accurately, we use the 10% cross-validation method.

We adopt precision and recall, which can be used to measure the relevance between a set of ranked results and the user. The collection of k items recommended to user u is $R(u)$, user u favorite collection of items on the testing set is $T(u)$. Accuracy and recall show the effect of recommendation from different point of view. In evaluating the model, the higher the accuracy and recall are, the better. But in some extreme cases, there are some contradictions between the two indicators. So, we need to consider them together. F1-measure is a weighted average of precision and recall. F1 is higher, test method is more effective. Three indicators can be seen in 4-(1),4-(2),4-(3).

$$\text{Precision} = \frac{\sum_{u=1}^{|U|} |R(u) \cap T(u)|}{\sum_{u=1}^{|U|} |R(u)|} \quad 4-(1)$$

$$\text{Recall} = \frac{\sum_{u=1}^{|U|} |R(u) \cap T(u)|}{\sum_{u=1}^{|U|} |T(u)|} \quad 4-(2)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad 4-(3)$$

6. CONCLUSION AND RECOMMENDATION

In total, we have tested 3000 different queries. After that we compute for the whole queries, the average precision related to our approach, and the average precision related to the other two approaches.

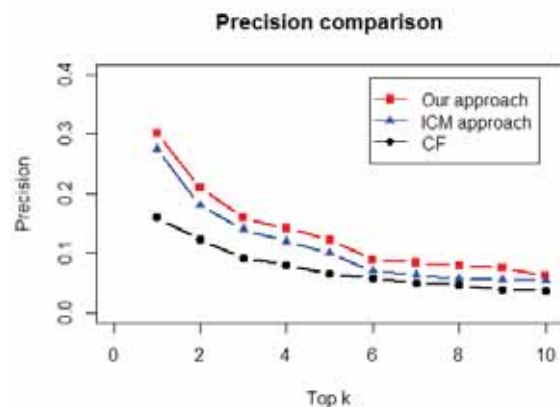


Figure 6. Precision comparison at top k

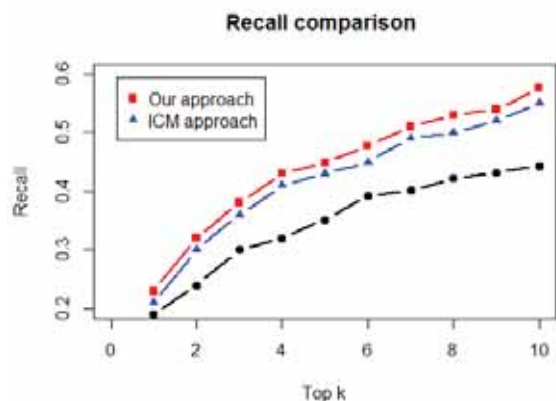


Figure 7. Recall comparison at top k

First, Figure 6 shows the results of the precision performance with respect to different values of k (top item recommended) according to 4-(1). The results show that the item based CF approach has worst performance compared to the other two approaches. Then examine recall of each algorithm as shown in Figure 7 using 4-(2). Our method obtained approximately 2.3 % on recall (at top 10). Finally, we examine F1-measure of each algorithm as shown in Figure 8 using 4-(3).

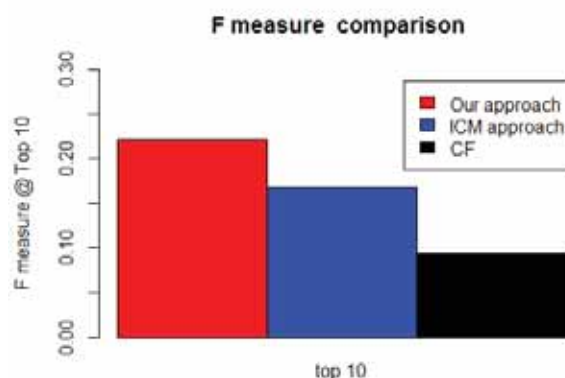


Figure 8. The F-measure at top10 of each approach

Considering that it is difficult to obtain scoring information when we recommend items to users in a context-aware environment, this article attempts to use tag information to make recommendations. Through tagging information, this paper excavates the potential relationship between users and items. Moreover, considering the correlation intensity, the numbers of tagging are not directly proportional in the actual applications. Therefore, we convert the relationship of users, tags and items into a more reasonable correlation index through the logistic population growth formula. Experiment shows that the quality of proposed model is higher than other models.

The work presented here shows prospects for further research. In fact, the quality of recommendation also relates to time and number of repeated tagging, so we are now developing a more perfect algorithm which considers tagging time and number of repeated tagging. We hold that the relational strength of items that are marked recently is higher than items marked previous. Similarly, we also recognize that the stability strength item which is marked many times is higher than the item marked only once. Although our investigation has provided promising results, we believe that our contribution is an initial step in the study of tag-aware RS. Additional research in this field is still to be explored.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (GrantNo.71471165, No.71501172), Zhejiang Provincial Social Science Funds of China (Grant No.17NDJC256YB)

REFERENCES

- [1] Zhai Shuang, Song Wen. (2010). Overview of the Research on Social label .Library and Information work, 54(20): 41-44. (in Chinese)
- [2] Zhang Zike, Zhou Tao, Zhang Yicheng. (2011). Tag-Aware Recommender Systems:A State-of-the-Art Survey .Journal of computer Science and Technology, 26(5):767. (in Chinese)
- [3] Yang X, Zhang T, Xu C, et al. (2015). Automatic Visual Concept Learning for Social Event Understanding. IEEE Transactions on Multimedia, 17(3):346-358. (in Chinese)
- [4] Krestel R, Fankhauser P, Nejdl W. (2009) Latent dirichlet allocation for tag recommendation. ACM Conference on Recommender Systems, Recsys 2009, New York, Ny, Usa, October. DBLP, 61-68.
- [5] Yang Zhaoxiang. (2016). Situation acquisition based on social tagging in computing environment. M Thesis. Zhejiang Sci-Tech University. (in Chinese)
- [6] Liu H L, Liu W F, Wang Yanjiang, et al. (2016). Regularization Logistic regression model of Hessian.computer

- Engineering and Application, 52(5): 236-240.
- [7] Wei Jian-liang, Zhu Qinghua.(2009). Summary of socialized tagging theory .Journal of the Chinese Library, 35(6)88-96. (in Chinese)
- [8] Doerfel S, Singer P, Singer P, et al. (2016). What Users Actually Do in a Social Tagging System: A Study of User Behavior in BibSonomy. *Acm Transactions on the Web*, 10(2):14-17.
- [9] Noll M G, Meinel C. (2008) Exploring social annotations for web document classification. *ACM Symposium on Applied Computing*. ACM, 2315-2320.
- [10] Yeung C M A, Gibbins N, Shadbolt N.(2008) A Study of User Profile Generation from Folksonomies. *World Wide Web Conference Series*.
- [11] Pan X, He S, Zhu X, et al.(2016). How users employ various popular tags to annotate items in social tagging: An empirical study. *Journal of the Association for Information Science & Technology*, 67(5):1121-1137. (in Chinese)
- [12] Wang L.(2017).Personalized Movie Recommendation Based on Social Tagging Systems.International Conference on Advanced Design and Manufacturing Engineering. (in Chinese)
- [13] Fu Qingmiao.(2017). Research on the characteristics of user tags in the social label system. M Thesis. Zhejiang Sci-Tech university. (in Chinese)
- [14] Gemmell J, Schimoler T, Mobasher B, et al. Tag-based resource recommendation in social annotation applications.(2011) *International Conference on User Modeling, Adaption, and Personalization*. Springer-Verlag,6787 111-122.
- [15] Lebib F Z, Mellah H, Drias H.(2017). Enhancing information source selection using a genetic algorithm and social tagging. *International Journal of Information Management*.
- [16] Hotho A, Jäschke R, Schmitz C, et al.(2006). Information retrieval in folksonomies: search and ranking. *Semantic Web Research & Applications*,4011, 411-426.
- [17] Ramezani M.(2011). Improving Graph-based Approaches for Personalized Tag Recommendation. *Journal of Emerging Technologies in Web Intelligence*, 3(2).
- [18] Campos L M D, Fernández-Luna J M, Huete J F, et al.(2010). Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. *International Journal of Approximate Reasoning*, 51(7): 785-799.
- [19] Bobadilla J, Serradilla F, Bernal J. (2010). A new collaborative filtering metric that improves the behavior of recommender systems. *Knowl-Based Syst. Knowledge-Based Systems*, 23(6): 520-528.
- [20] Reyn N, Shinsuke N, Jun M, et al. (2007). Tag-Based Contextual Collaborative Filtering. *Iaeng International Journal of Computer Science*, 34(2): 214-219.
- [21] Alhamid M F, Rawashdeh M, Hossain M A, et al. (2016) Towards context-aware media recommendation based on social tagging[J]. *Journal of Intelligent Information Systems*, 46(3): 499-516.

Will the Recommendation Information be Sure to Improve the Brand Evaluation: The Impact of Distrust

Xinyan Liu¹, Cheng Hu^{2}*

¹ School of Business Administration, Zhongnan University of Economics and Law, China

² School of Business Administration, Zhongnan University of Economics and Law, China

Abstract: In the process of consumption, consumers will come into contact with all kinds of information related to the goods they want to buy. At this time, the consumer's trust of the information that they contact with will greatly affect the consumer's evaluation and attitude towards the alternative products. In the psychology field, the psychological state of distrust will activate the alternative explanation of the given information. This article expands the research field about distrust from psychology to the marketing field, and finds that when consumers see the recommended information, if they do not trust it, it will lead to lower evaluation of the brand in the recommended information. At the same time, consumers' evaluation of the competitive brand in the information will be reduced, too.

Keywords: distrust, alternative explanation, brand evaluation

1. INTRODUCTION

At present, people are increasingly keen on online shopping. Without seeing the physical goods, consumers can only understand the product through the information given by others. Imagine such a scene: you want to buy a mobile phone on the Internet, and hesitate in brand A and brand B. At this time, someone strongly recommends the mobile phone of brand A to you. In general, people's purchase intention at this time is inclined to brand A. But if this person has deceived you or has hurt your interests, many people may not accept this person's recommendation without hesitation, but will have doubts on the recommendation of brand A, even in the alternative situation will tend to buy brand B.

Therefore, the initial point of this article is that in the decision-making process of consumption, if the information is distrustful to a consumer, the recommended information will reduce his evaluation of the original brand in the recommended information, and will improve his evaluation of the competitive brand. In fact, this is a problem that has been studied in the field of psychology.

In the psychological field, distrust has been defined as a “lack of confidence in the other, a concern that the other may act so as to harm one, that he does not care about one’s welfare or intends to act harmfully, or is hostile”^[1](Kramer RM, 1999). ” Psychologists believe that the psychological state of distrust is a warning signal to people that others' motives, intentions and future behaviors may be ambiguous, and the information they convey may be fictive and misleading. There is a basic conclusion about distrust that, when an individual believes in a source, the receiver tends to focus on congruent messages; however, if the source is suspected to be unreliable, the receiver spontaneously activates the incongruent messages, because they are thinking about what will happen if the message is invalid. Psychologists have concluded that when individuals think that specific information sources are not credible, their natural response is to consider alternative explanations of information provided by them. Alternative interpretation is information that is incongruent with the given information. That is to say, when individuals distrust sources of information, they will consider information that is different, or even opposite, from the one given.

* Corresponding author. Email: yanzido@163.com

However, in the marketing field, few scholars take "distrust" as an independent variable, but more often use it as a mediate variable. According to Margaret's research in 2000, if persuasion scenario is considered to have ulterior motives, it will cause consumers' distrust of salespeople, so consumers' persuasive knowledge will affect their evaluation of salespeople^[2]. The study of GuangXin Xie and David M in 2015 also concluded that when consumers realize that a product is implicitly sold, they will have poorer brand attitude and lower purchase intention^[3]. In this process, the psychological state of distrust is also a mediating mechanism.

In this paper, distrust is taken as an independent variable, and the hypotheses are tested by the experimental method. The final conclusion is that when a consumer is in a distrust state, recommended information will reduce his evaluation of the original brand, and at the same time, it will reduce his evaluation of the competitive brand, too.

2. THEORETICAL BACKGROUND

2.1 Distrust in psychology

Distrust, as a relative concept of trust, was rarely ever studied by 1990 alone, but more often regarded as the antithesis of trust. In the traditional view of psychology, trust is an attribute of individual differences, and distrust is the same concept as trust. They are the two ends of a continuum. Therefore, the low trust represents a distrust of high. Someone distrust a person means that he is not thought to does not think that person will act for his own best interests, and may even engage in potentially harmful actions against him, and also means that this person will produce a variety of non-cooperative behavior^[4].

Later, some scholars questioned the opposing views of trust and distrust. Research on positive and negative emotions shows that these two emotions are not the same two ends of the continuum, but two different concepts. Therefore, the view that "positive and negative attitude is the two ends of a continuum" needs to be reexamined. In 1979, Luhmann proposed that the effect of trust and distrust is achieved in different ways^[5]. As the opposite of trust, distrust should be a negative expectation of a person's ability and goodwill. But Luhmann's research found that distrust is the act of unacceptable behavior, or even the determination of positive expectation for harmful behavior. That is to say, trust is the expectation that an individual will take a favorable course of action, while distrust is the expectation that an individual will harm others. In 1999, a literature review of Kramer RM on trust and distrust gave a proper definition of distrust. He defined distrust as "lack of confidence in the other, a concern that the other may act so as to harm one, that he does not care about one's welfare or intends to act harmfully, or is hostile." That is to say, distrust is not only a lack of trust, but also an early warning that the individual will harm someone's interests.

In the early days, scholars mostly focused on the behavioral aspects of distrust, but little on the intention, motivation and other aspects. Later, with the development of distrust research, some scholars suggest that distrust reflects the recipients' perception of the misleading intentions of information sources^[6] (Schul, Mayo, 2007). Distrust is related to concealing the truth and lack of transparency. It is an indefinite state in which this state of mind conveys a warning sign that the motives, intentions and future actions of others may be ambiguous, and the message they convey may be fictitious and misleading.

Therefore, the development of the concept of distrust has gone through a series of evolution process. Scholars finally put forward that distrust and trust are two different concepts. And distrust arouses great concern in academic fields.

Psychological studies have found that, when people are in a state of distrust, the information processing mechanism of them is more complex than in the state of trust. When in a state of trust, people directly accept the content of the message; but when in a distrust state, people will consider that the information itself may be

correct, and also consider that the opposite of information may be right^[7] (Kruglanski, 1989). Because distrust is related to concealing the truth, when distrust is arisen, people tend to look for informational content that is incongruent with the given information. Such incongruent information is called alternative explanation.

In 2000, Galinsky and Moskowitz proposed that when counterfactual thinking patterns are activated, people are more likely to propose new ways of using objects or test the opposite results of a given hypothesis^[8]. Counterfactual thinking is a kind of thinking process that individuals replace the unreal conditions or possibilities. When the mental state of distrust is activated, it has the effect similar to counterfactual thinking^[9] (Schul, Mayo, 2004). The recipient will avoid being misled by testing potential alternative explanations. When a source of information is not trusted, information is questioned, so the "unconventional" concept that is incongruent with the information will be conveyed to the recipient spontaneously. For example, when a person is in a state of distrust, if you give him a stimulus word "dark", he may think of the word "light". Or when someone is trying to sell something, he would say that this is the last inventory, or that price will be doubled tomorrow, in the distrust condition, people may think that such things are likely to have a large amount of inventory, or tomorrow the price is likely to remain unchanged, or even falling.

Based on the mechanism that distrust will activate in congruent alternative explanations, many scholars conduct further studies. In the underlying mechanisms of distrust, one considers the opposite of the given message. Due to the vigilance of the reverse relationship, distrust people will be more likely to find rules when they are more suitable for the less common condition^[10] (Schul, Mayo, 2008). And the sensitivity to unconventional events means an increase in the accessibility of uncommon and remote events, which is consistent with the flexibility of thinking and represents a creative enhancement. So distrust inspires creativity^[11] (Thomas Mussweiler, 2011). Thomas Mussweiler also studied the relationship between distrust and the reduction of stereotypes in 2013. Distrust stimulates the unconventional process of information processing, then the unconventional process of information processing turns to reduce stereotypes, so distrust can reduce stereotypes^[12].

2.2 Distrust in marketing

In the field of marketing, there is little studies use distrust as a single concept and studies it as an independent variable. However, in the process of persuasion, the use of persuasive knowledge is actually related to the mental state of the distrust, and the distrust acts as a mediator. In the research of persuasion knowledge, the basic mechanism is: when consumers have persuasion knowledge, they will doubt the intention of marketers, resulting in a lack of positive evaluation for marketers and even products. In 1980, Mona and Robert found that subjects will try to get rid of the advice of disliked or unbelievable communicators^[13]. If the consumer considers that the salesperson have motivate to persuade him, then the salesperson will be thought insincere. If persuasive scenes are considered to have ulterior motives or consumers have an unlimited cognitive ability, persuasion will affect consumers' evaluation of salespeople (Margaret, 2000). When consumers realize that some kinds of products are sold recessively, their attitude to the brand will be poorer and the purchase intention will be lower (GuangXin Xie and David M, 2015). In such marketing scenarios, like salesmen are believed to have persuasive motivation, persuasive scenes are considered to have ulterior motives, and some products are sold recessively, all of these conditions stimulate consumers' distrust.

When the marketing scene took place online, the mental state of distrust may appear more obviously. Low trust clues will lead to high vigilance to the motive behind the persuasive incident. When shopping online, the consumers who are in the state of distrust will doubt the motivation, intention and behavior of network sellers, even think they are malicious, thereby terminate the purchase behavior^[14] (Carol and Choon, 2009). When people browse online reviews, many comments may have some boastful content. When people are in the state of

trust, boast is acceptable for the reviewers with strong professional knowledge, leading to more convincing. However, when people are in the state of distrust, boast will stimulate the reader's high vigilance about the motivation of the review, resulting in a decline in the persuasiveness of the commentary^[15] (Grant, 2016).

Through reviewing the literature in marketing field, we can find that when studying distrust in marketing, most studies regard distrust as an intermediary mechanism, and concluded that some behaviors can lead to results of non-cooperation, no-dependence or cancellation of transactions because of distrust. But according to the basic mechanism in psychological research, distrust can be studied as an independent variable. When the consumer receives the recommendation information, if it is in a state of distrust, the content of information that is incongruent with the recommended information will be activated spontaneously. Distrust activates the alternative interpretation of the proposed concept, which makes the accessibility of the substitution concept and the original concept to the same degree, and reduces the influence of the original concept^[16] (Tali Kleiman, Noa Sher, 2015). That is to say, in the state of distrust, when the consumer sees a recommended product information, the alternative explanation, which is incongruent to the recommendation, is immediately activated. In other words, consumers will consider this product is unworthy to be recommended. At this point, the effect of the recommended content in the information is weakened. According from these, this paper puts forward following hypothesis:

H1: When consumers distrust the sources, the recommendation information will reduce the consumer's evaluation of the recommended brand.

In fact, alternative interpretations can come from many aspects. For example, apart from “this product is not recommended”, other alternative explanation to the information above can be the alternative brand of the recommended product. That is to say, in the state of distrust, mentioning a brand name in an advertisement may remind people of the competitive brand, weakening the influence of the brand in advertisement, and enhance the influence of the competitive brand. The same is true for the other recommended cases. Therefore, this paper makes other hypothesis:

H2: When consumers distrust the sources, the recommendation information will improve the consumer's evaluation of the competitive brand.

3. EXPERIMENTAL DESIGN

The purpose of this experiment is to test the two hypothesis mentioned above, that is, when people distrust the source, recommendation information will reduce people's evaluation of the brand and improve the evaluation of consumers' competitive brand. The main contents of the experiment include the manipulation of distrust and the measurement of brand evaluation.

3.1 Experimental process

118 college students participated in the experiment, and the subjects were randomly divided into four groups. In the end, 12 participants were excluded because their manipulation tests were not passed. The experimental conditions were the inter group experiment of 2 (the vs control group of the group of distrust) × 2 (the original brand vs competitive brand).

All participants were asked to imagine the following scenario:

You want to buy a camera, and after comparison, you finally think that the camera of brand A and the brand B are in line with your expectations. And in the process of collecting information before, you find there is no significant difference in the evaluation of the two cameras. At this point, you see an organization assessing the camera of the brand A.

The organization has made a comprehensive assessment of the brand A camera, and the final conclusions

are as follows:

1. The camera of brand A has top quality.
2. The camera of brand A has unbeatable performance.
3. The camera of brand A has more features than other brands.
4. No other camera gives you more value than the camera of brand A.

In the distrust condition, the participants would be told that there was news exposure that the evaluation organization had written false advertisements before. Several products' quality was not consistent with the evaluations from this organization. And there is no such hint in the control group.

Under the condition of original brand, participants were asked about their attitudes towards the brand A, while under the competitive brand condition, they were asked about their attitude towards the brand B camera.

3.2 Experimental measurement

3.2.1 Brand attitude

In order to understand the participants' attitudes toward the two brands under different scenarios, participants were asked to fill in a questionnaire (Table 1). One to five represents very disagree to very agree.

Table 1. Brand attitude test

	1	2	3	4	5
The camera of brand A/B is good					
The camera of brand A/B is satisfactory					
The camera of brand A/B is useful					
The camera of brand A/B is favorable					
The camera of brand A/B is appealing to me					

If the experimental condition is original brand, the participants were asked about their attitude towards the brand A's camera. If the experimental condition is competitive brand, then they are asked about their attitude towards the brand B's camera.

3.2.2 Manipulation test

In order to test whether the manipulation of distrust in the experiment was successful, participants were asked to fill in another questionnaire (Table 2). One to five represents very disagree to very agree.

Table 2. Manipulation test

	1	2	3	4	5
This evaluation is truthful					
This evaluation is honest					
This evaluation is misleading					
This evaluation is deceptive					

3.3 Result and discussion

3.3.1 Manipulation test

To test whether the manipulation of distrust was successful, the data are divided into two groups: the distrust group and the control group. The results of data analysis showed that when distrust was manipulated, the distrust level of participants ($M=4.11$, $SD=0.46$) was significantly higher than that in the control group ($M=2.18$, $SD=0.65$; $F=287.18$, $p<0.01$). That was to say, the manipulation test of this experiment was successful, and the

distrust level of the participants under the distrust condition is higher than that of the participants under the control condition.

3.3.2 Attitude towards brand A

The results of data analysis showed that the participants' evaluation of brand A in distrust condition ($M=1.88$, $SD=0.55$) was lower than those who were in control condition ($M=3.88$, $SD=0.62$; $F=143.64$, $p<0.01$). This result validates the first hypothesis of this paper.

3.3.3 Attitude towards brand B

The results of data analysis showed that the participants' evaluation of brand B in distrust condition ($M=2.29$, $SD=0.74$) was lower than those who were in control condition ($M=3.37$, $SD=0.73$; $F=28.37$, $p<0.01$). This data result is contrary to the second hypothesis of this paper.

In summary, the following statistical analysis can be obtained (Figure 1):

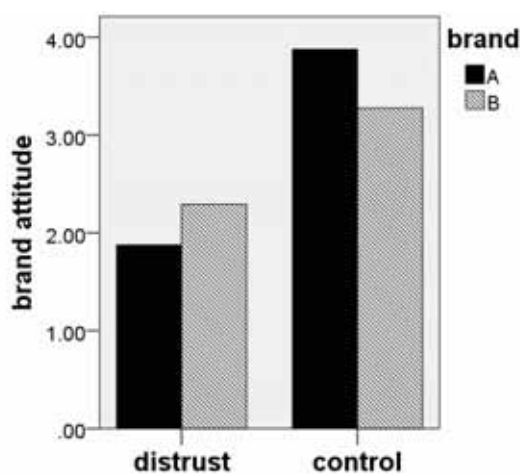


Figure 1. Experiment result

When individuals are in a distrust mental state, they will consider the alternative explanation of the given information, so when consumers distrust the source, recommendation information will reduce the consumer's evaluation of the recommended brand. The reason why the evaluation of competitive brands is also reduced may be that consumers' awareness of self-protection will lead to defensive bias. Because the original and competing brands are belong to the same category, this defensive bias will make negative evaluation on original brand spill over, leading to the reduced evaluation of competitive brands.

4. GENERAL DISCUSSION

In the experiment above, this paper verifies that when consumers distrust the source, compared to the undisturbed condition, they will reduce the evaluation of the original brand in the recommendation information. It is proposed in the psychology field that when the individual is in the state of distrust, an alternative explanation which is incongruent with the given message is activated spontaneously. However, the experimental results also indicate that when consumers are in the distrust condition, they will reduce the evaluation of the competitive brand in the recommendation information, which is contrary to the original hypothesis. The possible reason for such a result is that consumers have defensive bias and the spillover effect of the negative evaluation.

The main theoretical contribution of this paper is to expand the research and conclusions in psychology field to marketing field, and find that the mental state of distrust really activates people's alternative explanations. In the marketing field, this paper no longer regarded distrust as an intermediary mechanism, but studied it as an

independent variable. The dependent variable also extends from non-cooperative and other distrust behaviors to attitude towards competitive brands. The study found that when consumers read recommendation information, distrust could activate their alternative explanation of recommendation information, so that consumers can reduce the evaluation of original brand in recommendation information.

At the same time, the conclusions of this paper also have some reference for enterprises and businesses. Because consumers will have a negative evaluation of the original brand and the competitive brand in the recommendation information when they are in the state of distrust, the enterprises and businesses need to be careful when they choose the advertising media. At the same time, when consumers are generally distrust a type of advertisement or an advertising medium, the enterprises need to stop the advertisement in time. And the competitive brands of the advertiser also need to take measures to weaken the effect of spillover of distrust and reduce the negative attitude of consumers to the competitive brand.

Of course, there are some limitations in this paper. Firstly, the control of the product category is relatively simple. This paper only select the camera as a representation, which may not be universal. Secondly, after discovering negative effect of the competitive brand, no further experiments are designed to explore the reasons and influence mechanism. Therefore, the future research can explore that for different categories of goods, if there is the same conclusion. At the same time, distrust has negative impact on competitive brands, so for different categories of alternative commodity brand, will distrust have a positive effect on it or not? These problems can be used as the future research direction to do further exploration.

REFERENCES

- [1] Roderick M. Kramer. (1999). Trust and Distrust in Organization: Emerging Perspectives, Enduring Questions. *Annual Reviews*, 50:569-598
- [2] Margaret C. Campbell, Amna Kirmani. (2000). Consumers' Use of Persuasion Knowledge: The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent. *Journal of Consumer Research*, 27: 69-83
- [3] Guang-Xin Xiea, David M. Boushb, Raymond R. Liua (2015). Tactical deception in covert selling: A persuasion knowledge perspective. *Journal of Marketing Communications*, 21(3): 224-240
- [4] Lewicki, R. J., Mcallister, D. J., Bies, R. J. (1998). Trust and distrust: new relationships and realities. *Academy of Management Review*, 23(3):438-458
- [5] N. Luhmann. (1979). Trust: A mechanism for the reduction of social complexity. *Trust and Power*. Wiley, New York, 4-103.
- [6] Schul Y, Mayo R, Burnstein E. (2007). How people cope with uncertainty due to chance or deception. *Journal of Experimental Social Psychology*, 43(1): 91-103
- [7] Kruglanski, A. W. (1989). The psychology of being "right": the problem of accuracy in social perception and cognition. *Psychological Bulletin*, 106(3): 395-409
- [8] Galinsky, A. D., Moskowitz, G. B. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality & Social Psychology*, 78(4): 708-724
- [9] Mayo, R., Schul, Y., Burnstein, E. (2004). "I am not guilty" vs "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, 40: 433-449
- [10] Schul, Y., Mayo, R., Burnstein, E. (2008). The value of distrust. *Journal of Experimental Social Psychology*, 44(5): 1293-1302
- [11] Mayer, J., Mussweiler, T. (2011). Suspicious spirits, flexible minds: when distrust enhances creativity. *J Pers Soc Psychol*, 101(6): 1262-1277
- [12] Posten, A. C., Mussweiler, T. (2013). When distrust frees your mind: the stereotype-reducing effects of distrust. *Journal of Personality & Social Psychology*, 105(4): 567-84

- [13] Clee, M. A., Wicklund, R. A. (1980). Consumer behavior and psychological reactance. *Journal of Consumer Research*, 6(4): 389-405
- [14] Ou, C. X., Sia, C. L. (2009). To trust or to distrust, that is the question: investigating the trust-distrust paradox. *Communications of the Acm*, 52(5): 135-139
- [15] Packard, G., Gershoff, A. D., Wooten, D. B. (2016). When boastful word of mouth helps versus hurts social perceptions and persuasion. *Journal of Consumer Research*, 43(1): 26-43
- [16] Kleiman, T., Sher, N., Elster, A., Mayo, R. (2015). Accessibility is a matter of trust: dispositional and contextual distrust blocks accessibility effects. *Cognition*, 142: 333-344.

A Personalization-Privacy Paradox in Usage of Mobile Health

Services: A Game Theoretic Perspective

Fanbo Meng¹², Xitong Guo^{1}, Kee-hung Lai², and Xinli Zhao¹*

¹School of Management, Harbin Institute of Technology, Harbin, China.

²Faculty of Business, The Hong Kong Polytechnic University, Kowloon, Hong Kong.

Abstract: As health information privacy concern of the public raises, people are hesitant on disclosure of their private health information for personalized health services from using mobile health. The tension between personalization and privacy hinders users' adoption of mobile health services. In this study, we draw on game theory to explain the personalization-privacy paradox in the usage of mobile health services. The results show that: (1) In a one-shot game, the strategy set of mobile health marketers and users will be contrary to their original motivations. (2) In a repeated game, collecting users' private health information in a friendly way and disclosing private health information will be dominant strategies for both players. Managers need to pay attention to these scenarios in promoting usage of mobile health services.

Keywords: game theory, mHealth, personalization, privacy concern, technology acceptance

1. INTRODUCTION

As the mobile information communication technology evolves, the usage of mobile devices for accessing personalized healthcare services has dramatically increased ^[1]. Mobile health (mHealth) can capture users' health information and preferences in digital format and suggest corresponding individual-tailored health services to users. Using mHealth services will bring benefits to users including personalized health information, lower healthcare costs, improved medical outcomes, a more effective health services process, and a more efficient personal health data management ^[2]. Due to the capable features of mobile information communication technology, such as Global Position System (GPS) and sensors, the mHealth service providers or marketers can easily collect individuals' real-time health data including walking steps, heart rate, and blood pressure through their mobile device. To seek personalized health services and consult health professionals through mobile platforms, users are required to upload or report their health records, medical histories, demographic characteristics, and contact numbers, and so forth onto the platform. Based on real-time and self-reported health information, the service providers command a better position to understand users' health needs and hence provide customized health services. Despite the popularity of personalized health services, they are reluctant to report and disclose their private health-related information and data to mHealth marketers in view of the growing privacy concern in recent years. Such dilemma can be considered as a personalization-privacy paradox ^[3], suggesting "the tension between how the developers and marketers of IT applications exploit users' information to offer them personalized services, and those users' growing concerns about the privacy of that information, which can restrain their use of such applications" ^[4]. In the management information systems (MIS) literature, individuals' intentions to adopt a new technology is promoted by personalization but discouraged by privacy concern. Recent empirical studies highlight such personalization-privacy paradox in the use of online personalized services ^[5], electronic health records (EHRs) ^[6], smartphone ^[4], and mHealth services ^[7].

One novel aspect of this study is to use the game theory to explain the role of personalization-privacy

paradox in the adoption of mHealth services. From the game-theoretical perspective, mobile health users and mobile health marketers can be regarded two game players. The strategy set of mobile health marketers includes a friendly way and a hostile way to collect private health information, respectively. The former requires mobile health marketers to collect and analyze users' private health information in a friendly way, offering them personalized health information and services based on users' private health information. The latter performs such activities in a hostile way with the intention to abuse or disclose users' private health information for illegal profit-making. The strategy set of mobile health users includes 1) disclosing or 2) protecting personal private health information. In the first one, mobile health users allow mobile health marketers to collect their private health information and enjoy personalized health information and services in return for health care. Alternatively, they refuse to share their private health information with mobile health marketers, keeping their private health information safe under the second strategy.

Payoff functions of this game present the profits and losses of two players. The losses and payoffs of two players' strategic actions are significant determinants of their decision-making. Based on the game-theoretic perspective, this study aims to investigate how to reduce users' health privacy concern and leverage the power of personalization in promoting the usage of mobile health services.

In a two player game, mobile health users will make a response according to the strategies of mobile health marketers. To maximize payoffs, mobile health marketers' dominant strategy is to collect users' private health information in a hostile way when users choose to disclose their private health information. Therefore, based on the marketers' strategy, users are reluctant to share their private health information because the latter worried about unintentional disclosure of their private health information or even abuse for illegal profit-making. While this strategy set could partially explain the personalization-privacy paradox, yet this result may be only applied in the context of a one-shot game. Indeed, many mobile health marketers prefer a long-term relationship with their users for maximum profit rather than a one-shot game. Although mobile health users have a concern about their health privacy, they are keen on personalized health services. Strategies of two players in a long-term relationship may differ from those of a one-shot game. To better explain the personalization-privacy paradox in mHealth context, we discuss both players' strategies and payoffs from perspectives of both a one-shot game and a repeated game.

In this study, we follow two rules: 1) different users have different tolerance levels in private health information disclosure; 2) the possibility of users' private health information disclosure will increase if mobile health marketers frequently come to collect users' private health information. Therefore, we assume that the way marketers collecting users' private health information under users' tolerance level is seen as a friendly way, however, the way of collecting users' private health information beyond users' tolerance level is seen as a hostile way. Obviously, different strategies of two players will bring different payoffs and losses to each other. Both players will take into consideration their prior beliefs in evaluating payoffs and losses in the game. Therefore, a repeated game theory may be appropriate to explain this process, and Nash equilibrium may exist in this game.

The remainder of this paper proceeds as follows. In the next section, we discuss strategies and payoffs of mobile health marketers and users in a one-shot game. Next, we describe strategies and payoffs of two players respectively in a repeated game. We conclude with a description of our results and implications for research and practice.

2. STRATEGIES AND PAYOFFS OF TWO PLAYERS IN A ONE-SHOT GAME

In this game, two game players include private health information marketers and mobile health users. We discuss their different strategies and payoffs when using different strategies in this game respectively.

$U_profit_{tri}^{dis}$ refers to the profit that mobile health users gained when mobile health marketers take the

strategy of collecting users' private health information in a friendly way, and mobile health users take the strategy of disclosing their private health information. This also refers to the profit mobile health users lost when mobile health marketers take the strategy of collecting users' private health information in a friendly way, however, mobile health users are reluctant to disclose their private health information. This profit indicates that mobile health users could gain personalized health information and services to better serve their health needs.

$M_profit_{fi}^{dis}$ refers to the profit mobile health marketers gained when mobile health marketers choose to collect users' private health information in a friendly way, while users allow marketers to collect their private health information from them. This profit means that mobile health marketers could better understand users' specific health need and design mobile health services through collecting and analyzing users' private health information.

$U_loss_{hos}^{dis}$ refers to the loss mobile health users suffered when mobile health marketers choose to collect users' private health information in a hostile way, while users allow marketers to collect their health information. The loss is that mobile health users' private health information is abused and leaked out for illegal profit-making by mobile health marketers.

$M_income_{hos}^{dis}$ refers to the profit mobile health marketers gained when mobile health marketers take the strategy of collecting users' private health information in a hostile way, while mobile health users choose to disclose their private health information. This profit is different from the profit gained through collecting private health information in a friendly way. Although mobile health marketers may undertake higher risk, they can gain a higher profit from this strategy of collecting users' private health information in a hostile way.

$U_income_{hos}^{pro}$ refers to the profit mobile health users gained when mobile health marketers collect users' private health information in a hostile way, while mobile health users refuse to share their private health information with marketers. These profits indicate that mobile health users protect their health privacy and avoid the risk that they may suffer from private health information disclosure and abuse.

Based on the above analysis, both game players should choose following strategies: collecting private health information in a friendly way and disclosing private health information respectively, in order to gain profits and nurture a trusting relationship. When two players choose strategy set: collecting private health information a hostile way and disclosing private health information respectively, mobile health marketers could gain higher payoffs while mobile health users suffer losses in this game. When both game players choose strategy set: collecting private health information in a friendly way and protecting private health information respectively, mobile health marketers could not gain any profit and mobile health users miss the opportunity to enjoy personalized health information and services. When both game players choose strategy set: collecting health information in a hostile way and protecting private health information respectively, mobile health marketers could not gain any profit in this process, however, mobile health users prevent their private health information from private health information abuse. This game matrix is presented in Table 1.

Table 1 Game Matrix in a One-shot Game

Mobile Health Users	Mobile Health Marketers	
	Friendly	Hostile
Disclose	$U_profit_{fi}^{dis}, M_profit_{fi}^{dis}$	$-U_loss_{hos}^{dis}, M_profit_{hos}^{dis}$
Protect	$-U_profit_{fi}^{dis}, 0$	$U_profit_{hos}^{pro}, 0$

From the perspective of mobile health marketers, when mobile health users choose the strategy of disclosing private health information, the dominant strategy of mobile health marketers is to collect private health information in a hostile way. That is $M_profit_{hos}^{dis} > M_profit_{fi}^{dis}$. When mobile health users choose the strategy of protecting their private health information, the profit for mobile health marketers will be 0. From the

perspective of mobile health users, when mobile health marketers choose the strategy of collecting private health information in a friendly way, the dominant strategy of mobile health users is disclosing private health information. That is $U_profit_{fri}^{dis} > -U_profit_{fri}^{dis}$. When mobile health marketers choose the strategy of collecting private health information in a hostile way, the dominant strategy of mobile health users is protecting their private health information. That is $U_profit_{hos}^{pro} > -U_loss_{hos}^{dis}$. Based on the above strategies, there is a pure strategy Nash equilibrium $(U_profit_{hos}^{pro}, 0)$ in this game matrix. However, in fact, both mobile health marketers and users aim to communicate and share information through this emerging platform in a long term. Therefore, both marketers and users may not actually have the outcome reaching this Nash equilibrium. Although users may face the risk of privacy disclosure or abuse, they still choose to share their private health information in order to get better-personalized health information and services. To get continued profit, mobile marketers may change their strategies for longer-term customer relationship development. The dominant strategy of this game may be contrary to both players' original motivations. Therefore, this echoes a prisoner's dilemma between mobile health marketers and users which is discussed in a repeated game context below.

3. STRATEGIES AND PAYOFFS OF TWO PLAYERS IN A REPEATED GAME

Further to a one-shot game, we draw on a repeated game to discuss the prisoners' dilemma between mobile health marketers and mobile health users. Strategy sets of both players are a finite set. Then, in this repeated game process, the payoff of each strategy of mobile health marketers and mobile health users will be discussed as follows. δ refers to a discount factor of mobile health users' future payoff and ζ refers to a discount factor of mobile health marketers' future payoff.

From the perspective of mobile health users, when mobile health marketers collect private health information from users in a friendly way, users can choose to disclose their private health information or not. In this context, the payoff of users will be presented as $u_profit_{fri}^{dis}$. Therefore, when mobile health marketers collect users' private health information at the n^{th} time, mobile health users' payoff is $U^*_profit_{fri}^{dis}$.

$$U^*_profit_{fri}^{dis} = u_profit_{fri}^{dis} + u_profit_{fri}^{dis} * \delta + u_profit_{fri}^{dis} * \delta^2 + \dots + u_profit_{fri}^{dis} * \delta^n = u_profit_{fri}^{dis} * \frac{1 - \delta^n}{1 - \delta}$$

When mobile health marketers collect users' private health information in a hostile way, mobile health users choose to disclose their private health information. In this context, the payoff of users will be presented as $u_loss_{hos}^{dis}$. Therefore, when mobile health marketers collect users' private health information at the n^{th} time, mobile health users' payoff is $U^*_loss_{hos}^{dis}$.

$$U^*_loss_{hos}^{dis} = u_loss_{hos}^{dis} + u_loss_{hos}^{dis} * \delta + u_loss_{hos}^{dis} * \delta^2 + \dots + u_loss_{hos}^{dis} * \delta^n = u_loss_{hos}^{dis} * \frac{1 - \delta^n}{1 - \delta}$$

When mobile health marketers collect users' private health information in a hostile way, mobile health users choose to protect their private health information. In this context, the payoff of users will be presented as $u_profit_{hos}^{pro}$. Therefore, when mobile health marketers collect users' private health information at the n^{th} time, mobile health users' payoff is $U^*_profit_{hos}^{pro}$.

$$U^*_profit_{hos}^{pro} = u_profit_{hos}^{pro} + u_profit_{hos}^{pro} * \delta + u_profit_{hos}^{pro} * \delta^2 + \dots + u_profit_{hos}^{pro} * \delta^n = u_profit_{hos}^{pro} * \frac{1 - \delta^n}{1 - \delta}$$

From the perspective of mobile health marketers, when marketers choose to collected users' private health information in a friendly way, mobile health users disclose their private health information. In this context, the payoff of marketers will be presented as $m_profit_{fri}^{dis}$. Therefore, when mobile health marketers collect users' private health information at the n^{th} time, mobile health marketers' payoff is $M^*_profit_{fri}^{dis}$.

$$M^*_profit_{fri}^{dis} = m_profit_{fri}^{dis} + m_profit_{fri}^{dis} * \zeta + m_profit_{fri}^{dis} * \zeta^2 + \dots + m_profit_{fri}^{dis} * \zeta^n = m_profit_{fri}^{dis} * \frac{1 - \zeta^n}{1 - \zeta}$$

When mobile health marketers collect users' private health information in a hostile way, mobile health users choose to disclose their private health information. In this context, the payoff of marketers will be presented as $m_profit_{hos}^{dis}$. Therefore, when mobile health marketers collect users' private health information at the n^{th} time, mobile health marketers' payoff is $M_profit_{hos}^{dis}$.

$$M_profit_{hos}^{dis} = m_profit_{hos}^{dis} + m_profit_{hos}^{dis} * \zeta + m_profit_{hos}^{dis} * \zeta^2 + \dots + m_profit_{hos}^{dis} * \zeta^n = m_profit_{hos}^{dis} * \frac{1 - \zeta^n}{1 - \zeta}$$

When mobile health marketers collect users' private health information in a hostile way, mobile health users choose to protect their private health information. In this context, the payoff of marketers will be presented as $m_loss_{hos}^{pro}$. Therefore, when mobile health marketers collect users' private health information at the n^{th} time, mobile health marketers' payoff is $M_loss_{hos}^{pro}$.

$$M_loss_{hos}^{pro} = m_loss_{hos}^{pro} + m_loss_{hos}^{pro} * \zeta + m_loss_{hos}^{pro} * \zeta^2 + \dots + m_loss_{hos}^{pro} * \zeta^n = m_loss_{hos}^{pro} * \frac{1 - \zeta^n}{1 - \zeta}$$

Based on the above payoffs and losses of mobile health users and mobile health marketers, the game matrix of two players in this context is listed in Table 2.

Table 2 Game Matrix in a Repeated Game

Mobile Health Users	Mobile Health Marketers	
	Friendly	Hostile
Disclose	$U_profit_{fri}^{dis}, M_profit_{fri}^{dis}$	$-U_loss_{hos}^{dis}, M_profit_{hos}^{dis}$
Protect	$-U_profit_{fri}^{dis}, 0$	$U_profit_{hos}^{pro}, -M_loss_{hos}^{pro}$

Under this repeated game, when mobile health users' strategy is to disclose their private health information, the dominant strategy for mobile health marketers is to collect users' private health information in a hostile way rather than collecting data in a friendly way. Therefore, we can have that $M_profit_{hos}^{dis} > M_profit_{fri}^{dis}$. When mobile health users' strategy is to protect their private health information, the dominant strategy for mobile health marketers is to collect users' private health information in a friendly way. Accordingly, we can have $-M_loss_{hos}^{pro} < 0$. From the perspective of mobile health users, when marketers' strategy is to collect private health information in a friendly way, disclosing private health information is profitable for users. Therefore, we can have that $U_profit_{fri}^{dis} > -U_profit_{fri}^{dis}$. When marketers' strategy is to collect private health information in a hostile way, protecting private health information is profitable for users. Accordingly, we can have that $U_profit_{hos}^{pro} > -U_loss_{hos}^{dis}$.

Based on the above statements, in this repeated game context, we could easily find the dominant strategy for two game players. From the perspective of mobile health marketers, we can see that

$M_profit_{hos}^{dis} > M_profit_{fri}^{dis} > 0 > -M_loss_{hos}^{pro}$. In this case, collecting users' private health information in a hostile way will lead to loss rather than 0, when mobile health users' strategy is to protect their private health information. In a long-term, therefore, collecting users' private health information in a friendly way is continuously profitable for marketers. From the perspective of mobile health users, we can see that $U_profit_{fri}^{dis} > U_profit_{hos}^{pro} > -U_loss_{hos}^{dis}$. Therefore, when marketers' strategy is to collect users' private health information in a friendly way, the best response strategy of users will be disclosing private health information. When marketers' strategy is to collect users' private health information in a hostile way, the best response strategy of users will be protecting their private health information. In this case, marketers will choose to collect users' private health information in a friendly way for continued profit, and then users will disclose their private health information. Therefore, the best response set for marketers and users will be $(U_profit_{fri}^{dis}, M_profit_{fri}^{dis})$. This strategy set will lead to mutual benefits for both players in the long run.

4. DISCUSSION AND CONCLUSIONS

In this study, we draw on the game theory to explain the personalization-privacy paradox in the context of mobile health services. Health services, different from other services, is closely related to peoples' lives. Individuals may take cautious attitudes toward this emerging health services from a mobile platform. Users could enjoy personalized health services from mobile platforms, such as smartphones if they agree to upload and report their health record, medical history, and demographic characteristics. However, users are hesitant that mobile health marketers or servers may disclose or abuse their private information for illegal profit-making. With respect to mobile health marketers, in a one-shot game, the dominant strategy of marketers is to collect users' private health information in a hostile way for maximum profit. However, in a repeated game, collecting users' private health information in a friendly way is profitable for the marketer in a long term.

This study contributes to the knowledge stock of mobile health marketers by presenting an insight of the personalization-privacy paradox from game-theoretical perspective. Marketers aim to build a long-term relationship with users in order to continuously collect users' private health information to better understand users' health needs and provide personalized health services to users for profit. To make their payoffs maximum, marketers should choose the strategy of collecting users' private health information in a hostile way in a one-shot game. Instead, marketers choose to collect users' private health information in a friendly way in the repeated game in a long term. In practice, compared to a hostile collection of users' private health information, collecting private health information in a friendly way could not only reduce users' health privacy concern and offer users personalized health services but also avoid network security inspection. For marketers, our recommendations are that they should incorporate privacy-safe features and user-friendly designs in the mobile health application development.

For mobile health users, they hope to gain benefits from personalized health services but are highly concerned about health privacy. In this game, users will choose the best way to respond the strategy of mobile health marketers in order to maximize their own payoff. If mobile health marketers' strategy is to collect users' private health information in a hostile way, users' best response is to protect their private health information. If mobile health marketers' strategy is to collect users' private health information in a friendly way, users' best response is to disclose their private health information. In practice, comparing to protecting private health information, disclosing private health information is beneficial to users who can gain personalized health services, and to marketers who can utilize private health information to generate individually tailored health services for maximum payoff. Our study recommends that users should alleviate their health privacy concerns, and feel more secure that their private health information never actually be abused in a long run.

REFERENCES

- [1] Akter S, D'Ambra J, Ray P. User perceived service quality of m-Health services in developing countries [J]. 18th European Conference on Information Systems, 2010, 1-12.
- [2] Rai A, Chen L, Pye J, et al. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences [J]. *Journal of Medical Internet Research*, 2013, 15(8): e149.
- [3] Awad N F, Krishnan M S. The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization [J]. *MIS Quarterly*, 2006, 30(1): 13-28.
- [4] Sutanto J, Palme E, Tan C-H, et al. Addressing the Personalization-Privacy Paradox: An Empirical Assessment from a Field Experiment on Smartphone Users [J]. *MIS Quarterly*, 2013, 37(4): 1141-64.
- [5] Chellappa R K, Sin R G. Personalization versus privacy: An empirical examination of the online consumer's dilemma [J]. *Information Technology and Management*, 2005, 6(2): 181-202.

- [6] Angst C M, Agarwal R. Adoption of electronic health records in the presence of privacy concerns: the elaboration likelihood model and individual persuasion [J]. *MIS Quarterly*, 2009, 33(2): 339-70.
- [7] Guo X, Sun Y, Yan Z, et al. Privacy-Personalization Paradox in Adoption of Mobile Health Service: The Mediating Role of Trust; proceedings of the PACIS, F, 2012 [C].

Understanding continuance intention to use travel review websites

Ping Wang

Department of Entrepreneurship and Management, Turku School of Economics, University of Turku,
Finland

Hongxiu Li

Department of Industrial and Information Management, Tampere University of Technology, Finland

Yong Liu

Department of Information and Service Economy, Aalto University School of Business, Finland

Abstract: Travel review websites have become important information source for travellers. Given the widespread use of travel review websites among travellers, little research has been conducted to investigate what motivate travellers to use travel review websites. This study investigates travellers' continuance intention to use travel review websites by integrating three social media related constructs into the IS success model, and empirically test the research model among the Chinese travellers. In this research we found that the perceived enjoyment and curiosity fulfillment determine travelers' intention to continue using travel review websites together with the two determinants (information quality and system quality) suggested in the IS success model. Finally the implications for research and practices are discussed.

Keywords: Social media, travel review website, continuance intention, IS success model, eWOM

1. INTRODUCTION

Social media have rapidly penetrated into the society, and have been used by individuals as well as in business. In the past years, we have witnessed how social media have transformed the travel industry. People are increasingly using social media and travel review websites to search travel information, to make travel plans, to share travel information, and to communicate with others. In travel industry, more and more travellers heavily rely on user-generated reviews from different social media channels to support their travel-related purchasing decisions^[1].

In the literature plenty of studies have researched on travel reviews/eWOM on travel review websites, such as travel review generation behavior, travel review usefulness, travel review credibility, travel review valence, travel review use as well as the impact of travel reviews on marketing, sales and purchasing decisions. However, there is a lack of research on travel review websites, such as users' perceptions, trust, intentions, and behaviours related travel review websites. Such as little research has attempted to explore what determines travellers' continued use of travel review websites. As Bhattacharjee^[2] suggested, the continued use of an IS is important for its long-term success. Thus, understanding the factors leading to travellers' continuance use of travel review websites will provide new insights to travel review service providers on its strategies on retaining customers and to achieve its sustainable success.

In this research we investigate travellers' continuance intention to use travel review websites by integrating three social media related constructs into the IS success model. The proposed research model was empirically tested among the Chinese travellers.

In the next section, a literature review on IS success model is provided. Then the proposed research model and hypotheses are presented. We then discuss the research methods and the research results based on data analysis followed by the discussion on the research findings. Finally, we conclude the paper by highlighting both the theoretical and practical implications as well as pointing out the research limitations and future research directions.

2. LITERATURE REVIEW

The IS success model was originally developed by DeLone and McLean^[3]. The original IS success model identified six dimensions related to information systems (IS) success, including system quality, information quality, user satisfaction, system use, individual impact, and organizational impact. Later DeLone and McLean^[4] proposed an updated IS success model based on a literature review work on the different factors leading to IS success. In the updated IS success model, service quality was added as a construct to reflect the importance of service in supporting IS success and to provide a comprehensive evaluation of the overall quality of an IS together with information quality and system quality. In addition, intention to use an IS was included as an alternative to system use in the updated IS success model, and net benefits was added as a construct to evaluate both individual and organizational impact. The updated IS success is shown in Figure 1.

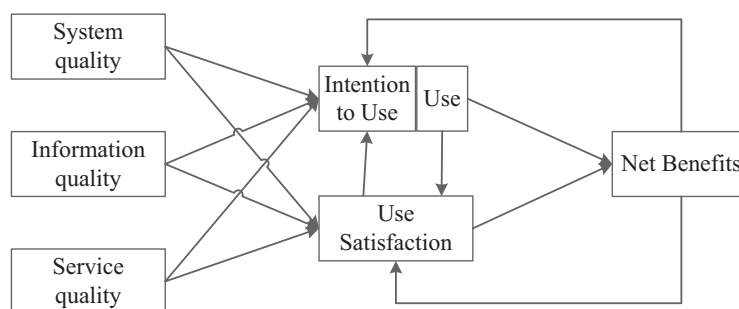


Figure 1. Updated IS success model^[4]

Both the original and the updated IS success model have been widely used to examine individuals' IS adoption behaviour in different research contexts, such as online shopping websites^[5], e-government services^[5], knowledge management systems^[6], and social media^[7]. As Wang^[9] and Wang et al.^[5] argued that the original and the updated IS success models can serve as a basis for the selection of appropriate IS measures, and researchers still need to choose some other appropriate measures of IS success based on the research objectives as well as the research phenomena under investigation when constructing the research model. Such as Wang et al.^[5] added perceived value into the updated IS success model to explore the stickiness intention to use online group purchasing websites together with the commitment-trust theory, and found that information quality, system quality and service quality determines user satisfaction together with perceived value, and customer satisfaction affects stickiness intention together with trust and relationship commitment. Dong et al.^[8] proposed a research model to examine continuance intention to use Facebook by integrating privacy protection service and benefits of social interaction into the IS success model. The research results indicate that information quality, system quality, and privacy protection service affect continued usage of social network sites indirectly via user satisfaction and benefits of social interaction. Chen et al.^[6] investigated e-government website usage in the context of online tax filing based on the trust theory and the updated IS success theory. They found that trust in e-government websites affect the three quality dimensions (information quality, system quality and service quality). Information quality is the most important factor influencing user perceptions of the usefulness of e-government websites and their satisfaction, followed by service quality and system quality, and user satisfaction, whereas perceived usefulness determine the perceived net benefits of using e-government websites.

3. RESEARCH MODEL AND HYPOTHESES

The current research employs the IS success model as its basic research framework, and incorporates social

interaction, curiosity fulfilment and perceived enjoyment in the research model in order to provide a comprehensive understanding of travel review website usage.

Travellers turn to travel review websites to search travel-related information to support their travel decisions, and eWOM information has become the main information they search for in travel decision process. Information quality is a dominant factor determining user satisfaction with an IS ^[5]. As travellers will use eWOM information from travel review websites to make travel plans and support their travel-related purchasing decisions, they must ensure the quality of the eWOM from the travel review websites. If the eWOM information provided at the travel review websites is relevant, complete and timely, travellers will feel satisfied with the travel review websites as the eWOM information helps them to make decisions on travel. Based on the above ground, we propose that:

H1: eWOM information quality positively affects user satisfaction with travel review websites.

System quality has been found to be an important antecedent of user satisfaction with a website. Good system quality means that the website is reliable, available and timely. As Lin ^[9] stated that system quality evaluates the technical feature of a website, and has positive impact on user satisfaction with a website. If a travel review website can help travellers navigate the site and find the right relevant information to support their travel decisions effectively and conveniently, the travellers will consider their use of the travel review website to be satisfactory, and vice versa. Based on the above ground, we propose the following hypothesis:

H2: System quality positively affects user satisfaction with travel review websites.

Social interaction is an important part in using travel review websites. Travel review websites provide an important channel for travellers to communicate with other travellers and the sellers. Interactive travel review websites make the information on the website to be more relevant and credible as most of the eWOM will be generated by other travellers. Travellers can also send request to other travellers and receive support for their travel decisions from others via social interaction. Yoo et al. ^[10] found that the interactivity of eWOM systems improves user satisfaction. Thus, it is assumed that if travellers achieve more social interaction in using a travel review website, they should be more satisfied with using the travel review website, and the following hypothesis is suggested:

H3: Social interaction positively affects user satisfaction with travel review websites.

The recent literature on social media research shows that perceived enjoyment is an important motivator driving an individual's use of social media as social media appear to be a multi-purpose-oriented system, such as both utilitarian and hedonic oriented systems ^[11,12]. It has been empirically validated in different IS contexts that individual users are more likely to be satisfied with an IS if they perceive more happiness and fun in using an IS. Based on the above discussion, we assume that perceived enjoyment is one of the antecedents determining a traveller's satisfaction with using a travel review website, and suggest the following hypothesis:

H4: Perceived enjoyment positively affects user satisfaction with travel review websites.

According to Agarwawl and Karahanna ¹³, curiosity fulfilment reflects the cognitive exploration in seeking out experiences in IS use. Curiosity plays an important role in the context of decision-making with uncertainty ^[14]. A website which can fulfil the curiosity of users seems to be more useful and more effective ^[14,15]. Travel is experience-based service, and there are a lot of uncertainties in travel decision-making process. Rich travel information from travel review websites can arouse and fulfil travellers' curiosity when they are using the travel review websites and address their uncertainties in making travel decisions, which will lead to travellers' satisfaction with the travel review websites. Thus, we propose that the perceived curiosity fulfilment in using travel review website will increase user satisfaction with travel review websites, and it is proposed that:

H5: Curiosity fulfilment positively affects user satisfaction with travel review websites.

Satisfaction have been suggested to be a dominant determinant of continuance intention to use an IS ^[1,2]. As

Bhattacharjee^[2] suggested that user satisfaction with using an IS plays critical roles in shaping continuance intention to use an IS. The prior IS literature has empirically validated the positive impact of user satisfaction on continuance intention in different research contexts^[1,16]. It is expected that user satisfaction with travel review websites is also associated with continuance intention to use travel review websites, and the following hypothesis is posited:

H6. User satisfaction positively affects continuance intention to use travel review websites.

The research model and the proposed research hypotheses are depicted in Figure 2.

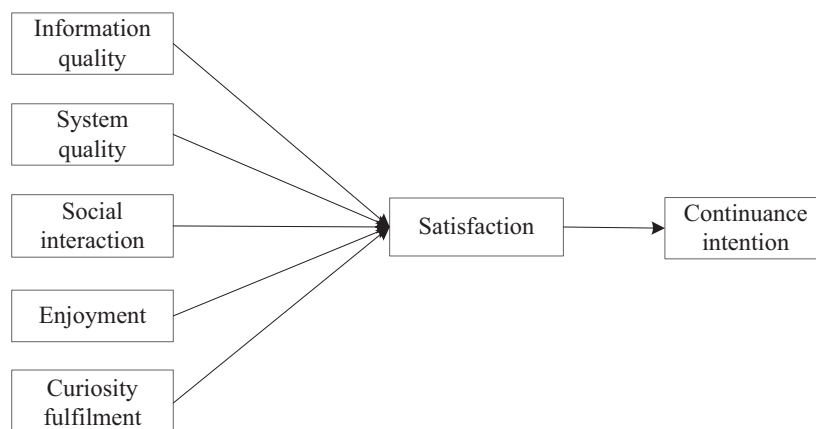


Figure 2 Proposed research model and research hypotheses

4. RESEARCH METHODOLOGY

Measurement

In this research, seven constructs are included, including information quality, system quality, social interaction, enjoyment, curiosity fulfilment, satisfaction, and continuance intention. The constructs included in the research model were measured with multiple items adopted from the existing IS literature, and were modified in order to make the items fit to the research context of travel review websites. This study employed a five-point Likert scale anchored from strongly disagree (1) to strongly agree (5) to measure all the measurement items included in the research instrument.

The constructs of satisfaction and continuance intention were measured using the items adapted from Bhattacharjee^[20] and Bhattacharjee^[2] respectively. The four items for the construct of eWOM information quality were modified from Wixom and Todd^[16] and Cheung et al.^[17]. The system quality items were taken from Wixom and Todd^[16]. The construct items of curiosity fulfilment were adapted from Agarwal and Karahanna^[14]. The items measuring perceived enjoyment were taken from Heijden^[15]. Social interaction was taken from Ko et al.^[18], and modified to fit to the research context of travel review websites.

Table 1. Constructs and items

Constructs	Items
eWOM information quality [16,17]	The eWOM on the travel review website provided the correct information for my travel plan. The eWOM on the travel review website provided me with a complete set of information for my travel. The eWOM on the travel review website is relevant to my travel plan. The eWOM on the travel review website provided me with the most up-to-date information for my travel-related decision.
System quality [16]	This travel review website is versatile in addressing needs as they arise. This travel review website pulls together information that used to come from different websites and information sources. The travel review website operates reliably. It takes short time for the travel review website system to respond to my requests.

Curiosity fulfilment [13]	Using this travel review website aroused my imagination about travel. Using this travel review website made me curious about new travel. Using this travel review website and reading other tourists' stories excited my curiosity for a new travel experience.
Enjoyment [15]	It was fun to use this travel review website. It was very enjoyable to use this travel review website.
Social interaction [18]	Using the travel review websites enabled me to see what other travellers said. Using the travel review websites enabled me to keep up with what's going on with regard to my travel. Using the travel review websites enabled me to express myself freely regarding my own travel.
Satisfaction [20]	Overall, I am satisfied with my website usage. My choice to use this travel review website was a wise one. My experience with using this travel review website was very satisfactory.
Continuance intention [2]	I intent to continue using the travel review website rather than discontinue its use. I will continue using the travel review websites rather than using alternative websites. I will frequently return to this website in the future.

Data collection

In this study a survey method was adopted to examine continuance intention to use travel review websites, and a questionnaire in Chinese was developed to collect data among Chinese travellers. The questionnaire was originally developed in English and was translated into Chinese. A back translation from Chinese to English was conducted in order to ensure the consistency of the survey instrument. A pilot study was conducted among 20 Chinese users of travel-related eWOM websites using the initial Chinese version to ensure the clarity of the texts before data collection.

An online travel service provider in China helped with data collection in this research. This online travel service provider offers online travel services to both domestic and international individual travellers in China, such as customized trip plan, local guided tours, ticket booking and so on. The online travel service provider distributed the link of the questionnaire to the customers and the followers at its official social media channel – WeChat - in 2016. Those who completed the online survey got a small gift provided by the research project. The online survey lasted for 2 weeks. The target respondents are the travellers who had the experience of using travel review websites, such as TripAdvisor. We received 351 responses. 31 responses were excluded from this study because the answers were not complete or not consistent. In addition, 8 respondents indicated in the survey that they had never used a travel review website, thus also excluded from this study. The 312 responses were used as valid database for this study. The demographic information and some travel-related characteristics are presented in Table 2.

Table 2. Demographic and travel-related characteristics of the respondents

Dimension	Items	Frequency	Percentage
Gender	Female	187	40.1
	Male	125	59.9
Age (years)	18-25	97	31.1
	26-30	109	34.9
	31-40	80	25.6
	41-50	23	7.4
	>51	3	1.0
Income (RMB/month)	<=5000	111	35.6
	5001-10,000	124	39.7
	10,001-15,000	42	13.5
	15,001-20,000	13	4.2
	>=20,001	22	7.1

Leisure travel frequency last year	0	9	2.9
	1	47	15.1
	2	113	36.2
	3	60	19.2
	4	25	8.0
	>=5	58	18.6
Preferred travel style	Self-organized	212	67.9
	Group travel	16	5.1
	Customized travel	81	26.0
	Others	3	1.0

Data validity and reliability

In this study Partial Least Squares (PLS) was employed to test the measurement and the structure models.

Convergent validity and discriminant validity were tested for the measurement model. Convergent validity was tested following the three criteria: (i) the factor loadings of the each measurement item on the respective constructs must be over 0.7; (ii) the composite reliability (CR) of each construct must be above the cut-off value of 0.8; and (iii) the average variance extracted (AVE) by each construct must exceed 0.5^[20,21,22]. As shown in Table 3, the factor loadings of all items are above 0.70; and the values of the composite reliability (CR), and average variance extracted (AVE) of each construct all are above the cut-off value of 0.8, 0.7 and 0.5 respectively (See Table 3). The test results indicate that there is a good internal consistency and reliability of this research and support the convergent validity of the research data.

Table 3. The measurement model

Constructs	Items	Factor Loadings	Cronbach's Alpha	CR	AVE
eWOM information quality (IQ)	IQ1	0.841	0.853	0.901	0.695
	IQ2	0.877			
	IQ3	0.861			
	IQ4	0.749			
System quality (SQ)	SQ1	0.783	0.803	0.871	0.629
	SQ2	0.814			
	SQ3	0.759			
	SQ4	0.815			
Curiosity fulfilment (CF)	CF1	0.934	0.937	0.959	0.887
	CF2	0.951			
	CF3	0.942			
Enjoyment (ENJ)	ENJ1	0.918	0.885	0.929	0.813
	ENJ2	0.935			
	ENJ3	0.849			
Social interaction (INT)	INT1	0.852	0.819	0.892	0.734
	INT2	0.837			
	INT3	0.881			
Satisfaction (SAT)	SAT1	0.919	0.907	0.942	0.843
	SAT2	0.917			
	SAT3	0.918			
Continuous intention (CI)	CON1	0.866	0.808	0.884	0.719
	CON2	0.775			
	CON3	0.897			

Discriminant validity indicates whether measurements reflect the construct in question or whether they reflect another related construct included in the research. Discriminant validity can be measured by testing whether the variance of the square root of the AVE for each construct is larger than any correlation between the tested construct and any other construct^[23]. The test results in this study show that the square root of the AVE for each construct is greater than the correlation estimates with other constructs (see Table 4), suggesting adequate discriminant validity for all the measurements.

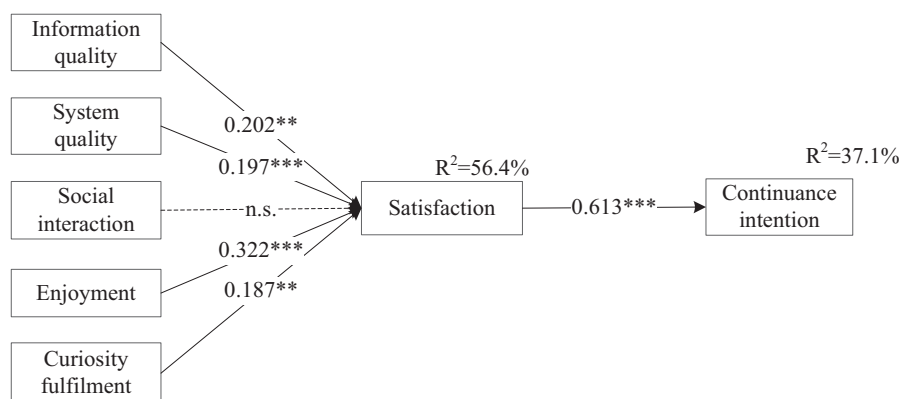
Table 4 Correlation matrix and discriminant assessment

	IQ	SQ	CF	ENJ	INT	SAT	CI
IQ	0.834						
SQ	0.639	0.793					
CF	0.432	0.320	0.942				
ENJ	0.546	0.429	0.739	0.902			
INT	0.551	0.498	0.518	0.552	0.857		
SAT	0.598	0.536	0.588	0.668	0.509	0.918	
CI	0.481	0.479	0.392	0.501	0.374	0.613	0.848

Note: Diagonals represent the squared root of average variance extracted for each construct. Off-diagonals are the correlations between latent variables.

Test of the structural model

The structural model was tested using a bootstrapping procedure in PLS. As shown in Figure 3, almost all the proposed hypotheses are supported, except H3 (Social interaction to satisfaction). User satisfaction was found to be positively related to eWOM information quality ($\beta=0.202$, $p<0.01$), system quality ($\beta=0.197$, $p<0.001$), enjoyment ($\beta=0.322$, $p<0.001$) and curiosity fulfilment ($\beta=0.187$, $p<0.01$), and continuance intention was found to be related with user satisfaction ($\beta=0.613$, $p<0.001$). The proposed research model explains 56.4 percent of the variation of user satisfaction with travel review websites and 37.1 percent of continuance intention to use travel review websites respectively.



* $p < 0.05$, $t > 1.96$; ** $p < 0.01$, $t > 2.58$; *** $p < 0.001$, $t > 3.29$

Figure 3 Structural analysis of the proposed research model

5. DISCUSSION

In this research, we found that eWOM information quality and system quality affect user satisfaction with travel review websites significantly, and user satisfaction exerts a positive effect on continuance intention to user travel review websites. The findings are consistent with the findings in prior research^[6,7], which support the use of IS success model as a basic research framework in understanding travel review website usage. Enjoyment and curiosity fulfilment were also found to have significant positive influence on user satisfaction with travel review website together with eWOM information quality and system quality. Interestingly, enjoyment was found to be the most important factor determining user satisfaction with travel review websites, followed by eWOM information quality, system quality and curiosity fulfilment.

The strong impact of perceived enjoyment on user satisfaction with travel review websites might be due to the following reason. When travellers use a travel review website, they not only achieve the goal of searching travel information to support their travel decisions, but also obtain happiness and fun during their use of travel review websites, such as getting a perception of happiness and fun when they are reading some interesting travel stories from travellers and some beautiful pictures and videos shared by other travelers. Travellers pay more

attention to the fun and happiness in using travel review websites, and take the perceived enjoyment in travel review website usage as the most important factor leading to their satisfaction with travel review websites.

Curiosity fulfilment also affects user satisfaction with travel review websites. Curiosity is always accompanied with information searching behaviour. Travellers search for travel information from travel review websites to address their knowledge gap or uncertainties in their travel decision-making process, and fulfil the curiosity raised in the process. When travellers feel that their curiosity are fulfilled when they use a travel review website to support their travel decisions, they will feel satisfied with the travel review website use experience.

Surprisingly, social interaction was not found to be a significant antecedent of user satisfaction with travel review websites. This finding is different from prior research findings on the role of social interaction in the context of social networking websites. This might be due to the difference of travel review websites and other social media. On travel review websites, users are not always online to do social interaction with other travellers, which makes social interaction to be much more difficult compared to Facebook and Twitter, etc. Thus, travellers have less concern on social interaction in their use of travel review websites. Thus, social interaction exerts no impact on their satisfaction with travel review websites.

6. CONCLUSION

This research attempted to advance the understanding of individual travellers' intention to continue using travel review websites from the IS success and the social media function perspectives. The study findings in this research contribute to the current understanding of travel review websites by examining the hedonic, the social and the utilitarian function of travel review websites. It also contributes to the literature on social media research by examining the role of social interaction in the travel review websites, and offers some new insights to social media research in understanding social interaction in different social media contexts. Meanwhile this research provides some practical guideline to travel review website service providers. Travel review website should be designed as a multi-purpose oriented system as enjoyment and curiosity are both found to be the determinants of user satisfaction with travel review websites. In addition, travel review websites should take different strategies to motivate travellers to provide quality eWOM for others to use as information quality affects user satisfaction. Furthermore, travel review system should keep good quality on system accessibility, reliability and timely response in order to keep travellers satisfied.

Like other researches, this study has several limitations. First, the sample only includes the data collected among Chinese travellers. Second, the research was empirically tested in the scenario of travel industry. Thus, the research findings might not be possible to be generalized to other industries. Third, other factors, such as trust, privacy, have not been considered in the current research as determinants of user satisfaction. Thus, future research should test the research model in a different context, or incorporate other factors into the IS success model to test different social media contexts.

REFERENCE

- [1] Li H, Liu Y. Understanding post-adoption behaviors of e-service users in the context of online travel services. *Inf Manag.* 2014;51(8). doi:10.1016/j.im.2014.07.004.
- [2] Bhattacharjee A. Understanding Information Systems Continuance: An Expectation-Confirmation Model. *MIS Q.* 2001;25(3):351. doi:10.2307/3250921.
- [3] DeLone WH, Mclean ER. Information Systems Success: The Quest for the Dependent Variable. *Source Inf Syst Res.* 1992;3(1):60-95. doi:10.1287/isre.3.1.60.
- [4] DeLone WH, Mclean ER. The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *J Manag Inf Syst / Spring.* 2003;19(4):9-30. doi:10.1073/pnas.0914199107.

- [5] Chen JV, Jubilado RJM, Capistrano EPS, Yen DC. Factors affecting online tax filing - An application of the IS Success Model and trust theory. *Comput Human Behav.* 2015;43:251-262. doi:10.1016/j.chb.2014.11.017.
- [6] Un Jan A, Contreras V. Success model for knowledge management systems used by doctoral researchers. *Comput Human Behav.* 2016;59:258-264. doi:10.1016/j.chb.2016.02.011.
- [7] Dong J, Li H, Zhang X. *Classification of Customer Satisfaction Attributes: An Application of Online Hotel Review Analysis*. Vol 445.; 2014.
- [8] Wang YS. Assessing e-commerce systems success: A respecification and validation of the DeLone and McLean model of IS success. *Inf Syst J.* 2008;18(5):529-557. doi:10.1111/j.1365-2575.2007.00268.x.
- [9] Lin HF. Determinants of successful virtual communities: Contributions from system characteristics and social factors. *Inf Manag.* 2008;45(8):522-527. doi:10.1016/j.im.2008.08.002.
- [10] Yoo CW, Kim YJ, Sanders GL. The impact of interactivity of electronic word of mouth systems and E-Quality on decision support in the context of the e-marketplace. *Inf Manag.* 2015;52(4):496-505. doi:10.1016/j.im.2015.03.001.
- [11] Li H, Liu Y, Xu X, Heikkilä J, Van Der Heijden H. Modeling hedonic is continuance through the uses and gratifications theory: An empirical study in online games. *Comput Human Behav.* 2015;48. doi:10.1016/j.chb.2015.01.053.
- [12] Gan C, Li H. Understanding the effects of gratifications on the continuance intention to use WeChat in China: A perspective on uses and gratifications. *Comput Human Behav.* 2018;78. doi:10.1016/j.chb.2017.10.003.
- [13] Agarwal R, Karahanna E. Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Q.* 2000;24(4):665. doi:10.2307/3250951.
- [14] van Dijk E, Zeelenberg M. When curiosity killed regret: Avoiding or seeking the unknown in decision-making under uncertainty. *J Exp Soc Psychol.* 2007;43(4):656-662. doi:10.1016/j.jesp.2006.06.004.
- [15] Heijden H Van Der. User acceptance of hedonic information systems. *MIS Q.* 2004;28(4):695-704. doi:10.2307/25148660.
- [16] Wixom BH, Todd PA. A theoretical integration of user satisfaction and technology acceptance. *Inf Syst Res.* 2005;16(1):85-102. doi:10.1287/isre.1050.0042.
- [17] Cheung CMK, Lee MKO, Rabjohn LN. *The Impact of Electronic Word-of-Mouth: The Adoption of Online Opinions in Online Customer Communities*. Vol 18.; 2008. doi:http://dx.doi.org/10.1108/10662240810883290.
- [18] Ko H, Cho C-H, Roberts MS. INTERNET USES AND GRATIFICATIONS: A Structural Equation Model of Interactive Advertising. *J Advert.* 2005;34(2):57-70. doi:10.1080/00913367.2005.10639191.
- [19] Bhattacharjee A. An empirical analysis of the antecedents of electronic commerce service continuance. *Decis Support Syst.* 2001;32(2):201-214. doi:10.1016/S0167-9236(01)00111-7.
- [20] Chin W. The partial least squares approach to structural equation modeling. *Mod methods Bus Res.* 1998;295(2):295-336. doi:10.1016/j.aap.2008.12.010.
- [21] Hulland J. Use of partial least squares (PLS) in strategic management research: a review of four recent studies. *Strateg Manag J.* 1999;20(2):195-204. doi:10.1002/(SICI)1097-0266(199902)20:2<195::AID-SMJ13>3.3.CO;2-Z.
- [22] Tenenhaus M, Vinzi VE, Chatelin Y-M, Lauro C. PLS path modeling. *Comput Stat Data Anal.* 2005;48(1):159-205. doi:10.1016/j.csda.2004.03.005.
- [23] Fornell C, Larcker DF. Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *J Mark Res.* 1981;18(3):382. doi:10.2307/3150980.

Gender Differences in Selective Attention and Shopping Intention

in the case of Taobao Live-show: An Eye-Tracking Study

Qing Xu¹, Mengqi Fei¹, Huizhong Tan¹¹

¹School of Management, Zhejiang University, China

Abstract: Live-show video was introduced into electronic marketing place two years ago. However, it still remains unknown how such new communication ways between buyers and sellers influence individual online shopping intention. How the limited attention resource would be assigned to the main information of endogenous and exogenous cues in the video, and then lead to shopping intention become interesting and important research questions. This study examines these questions from the view of gender difference and social influence with eye-tracking tool. An experiment is designed to capture the visual pattern of selective attention to endogenous and exogenous cues when participants watching Taobao live-show video. Data from Eye-tracking index and questionnaire will be collected to validate the proposed model. Potential contributions and implications for future research are discussed.

Keywords: gender differences; selective attention; live-show; eye-tracking; NeuroIS

1. INTRODUCTION

Individual got used to shopping online everyday with the rapid development of e-commerce. With the available of WIFI everywhere and the popularity of smart phone, video became an important way of online communication. Some platforms in China, such as Taobao and JD, adopted a new business model of live-show to provide better service for mobile customers in 2016. A new role of anchor was introduced between seller and buyer. The anchor will try on clothes, show the detail of design, and explain the possible match of collection before video. At the same time, customers can ask any questions about the product by type text in an interaction window. Such live-show effectively increase telepresence of consumer with vivid product presentation and instant communication. Consumers showed great passion for this new e-business mode since 2016.

The main interface features of the Taobao live-shows video screen (Figure 1) include four parts. No.1 area give basic information the seller name and location, the number of viewers, the live-show room ID, and updates information about the customer ID who enter the live-show room. No.2 area outlines the appearance of anchor and product. No.3 area presents social influence information including herding cues of “XXX is on the way to buy”, as well as text that consumers type in to interaction with anchor and other consumers. No.4 area is in the bottom screen. It includes a shopping bag with all the hyperlink of products showed in the video, an box window for consumer input text. Three buttons in the left are set to share with others, send gifts to the anchor, and show preference to this live-show.



Figure 1. Information in Taobao live-show

When people watch a movie or a video, they pay attention on where their eyes fix on. Since attention is a selective process, and the selective attention has two types: ‘endogenous’ and ‘exogenous’^[1]. In the traditional

¹ Corresponding author. Email: xuqing@zju.edu.cn (Qing Xu) , 11620004@zju.edu.cn (Mengqi Fei).

theory, endogenous attention competes with exogenous attention on the control over attention (Godijn and Theeuwes, 2002; Yantis, 1998, 2000; Yantis and Jonides, 1990). In the case of Taobao live-show, the interaction window displays the information of consumer who is on the way to purchase, which act as a herding cue on shopping intention of potential consumer. Therefore, both the endogenous attention on the main parts of video (including product and anchor) and the exogenous attention on the interaction window should have the effect on consumer's understanding of product and intention to click on shopping bag. In addition, previous research reported the different influence of social influence on individual between male and female. The social influences, such as herding cues, showed greater effect of exogenous attention on females than males. It's interesting to know the exact effect of endogenous and exogenous attention on shopping intention in such case of live-show.

Therefore, in this study, we try to figure out the gender differences in selective attention and shopping intention in Taobao live-show and propose the following research questions:

RQ1: What's the difference of selective attention between gender when individual watch Taobao live-show?

RQ2: How does the selective attention on endogenous and exogenous cues influence following shopping intention?

RQ3: What's the differences between gender in the influence of selective attention on shopping intention?

2. LITERATURE REVIEW

2.1 Endogenous attention and exogenous attention

Attention is limited resource especially in information explosion age. Selective attention enables us to tune out unimportant details and focus on what really matters to guides our behavior. Posner cueing paradigm was a basic paradigm to examine selective attention. Posner (1980) divided selective attention into 2 types of attention: 'endogenous' and 'exogenous'^[1]. The former is a voluntary system that corresponds to our ability to willfully monitor information at a given location; the latter is an involuntary system that corresponds to an automatic orienting response to a location where sudden stimulation has occurred^[2]. Lots of neurophysiological studies focused on whether common neurophysiological substrates underlie endogenous and exogenous attention and the relationship between them (Gandhi et al., 1999; Corbetta et al., 2000; Giesbrecht et al., 2003; Peelen et al., 2004; Hopfinger & West, 2006; Serences et al., 2005, 2007; Busse et al., 2008; Lovejoy & Krauzlis, 2010; Chica et al., 2013; Mysore and Knudsen, 2013). For example, Busse et al. (2008) suggested that endogenous and exogenous attention existed in two distinct attention systems but in the same capacity-limited system, where they compete with each other for the control over attention.

The performance mechanisms of endogenous and exogenous cues have been studied in a variety of tasks, e.g., texture segmentation (Yeshurun et al., 2008), letter identification (Talgar, Pelli, & Carrasco, 2004), and temporal resolution (Yeshurun, 2004). However, few studies applied real-life tasks instead of simplified experiment tasks with variants of Posner's paradigm. Recently, Tang et al. (2015) proposed a framework in which attention modulates multisensory processing in both endogenous (goal-driven) and exogenous (stimulus-driven) ways^[4]. They argued that an audiovisual cue may elicit a larger spatial cueing effect than corresponding visual cue, and endogenous and exogenous attention differentially but mutually modulate multisensory processing. In the context of Taobao live-show, buyers got to know more about clothes by using all available cues in the video. The endogenous cues (goal-driven) are anchor and product, which appear in the main middle area of the video window; the exogenous cues (stimulus-driven) are interactive communication, which appear in the left corner. It is worthwhile to note that exogenous cues include the flickering barrages of herding and textual interaction which may strengthen the endogenous response due to herding influence.

2.2 Gender differences in information processing

Traditionally, males showed a superior performance at visual-spatial tasks (Collins & Kimura, 1997), while females demonstrated an advantage in verbal and episodic memory tasks (Herlitz et al., 1997). Gender schema theory states that males and females differ in the extent to which they take advantage of schema to conduct cognitive processing (Martin et al., 2002). According to the theory, the male schema is associated with success and achievement to a greater degree than is the female schema (Noble et al., 2006), while females' schemas are oriented to a greater extent toward communal activities and goals (Putrevu, 2001). Merritt et al. (2007) found that females show larger validity effects in endogenously cued tasks, but not with a peripheral cue or exogenous cue in a basic Posner cueing paradigm ^[5]. In this way, online consumers' gender schema influences the information sources used for purchase decision (Ilie et al., 2005).

Some neurophysiological studies argued that males and females used different brain hemispheres to conduct their information-processing strategies and differed in information-processing thresholds. Males mainly rely on selective right hemisphere processing that involving a subset of highly available and salient cues. And females often depend on comprehensive left hemisphere processing to perform sequential and detailed analyses (Tsichla et al., 2014; Goodrich, 2014; Meyers-Levy & Loken, 2015). Hewig et al. (2008) explored the gender differences in gaze patterns when participants looking at the body of men and women with eye-tracking equipment ^[6]. The own-gender bias appears to be present in women regardless of the difficulty of the experimental task or how much of a face is presented (Lewin & Herlitz 2002; Lykins et al., 2008; Lovén et al., 2011; Man & Hills, 2016). In real-life situations, Shen and Itti (2012) found the eye-tracking evidence of orient attention difference between male and female during conversational listening ^[7]. Baptista et al. (2015) reported gender difference on eye-tracking patterns in an ironic statement and literal situations with focus on facial expression, written commentary and pictorial cues ^[8]. McIntyre and Graziano (2016) measured gender differences in selective visual attention toward person- and thing-related image content and examined how selective attention to people and things manifests in language use ^[9]. Coutrot et al. (2016) shown that female gazers follow a much more exploratory scanning strategy than males in eye-tracking data from participants watching videos of another person ^[10]. However, few empirical studies have used eye-tracking experiments to address gender differences in selective attention in the video watching task and online shopping context.

2.3 Gender differences in social influence

Humans beings are social animals. Individual mind and behavior may easily be influenced by others in the society. Social influence had been reported in conformity, socialization, peer pressure, obedience, leadership, persuasion, sales, and marketing. Herd behavior describes how individuals in a group can act collectively without centralized direction, such as the behavior of humans in demonstrations, riots and general strikes (Braha, 2012), decision-making, judgment and opinion-forming.

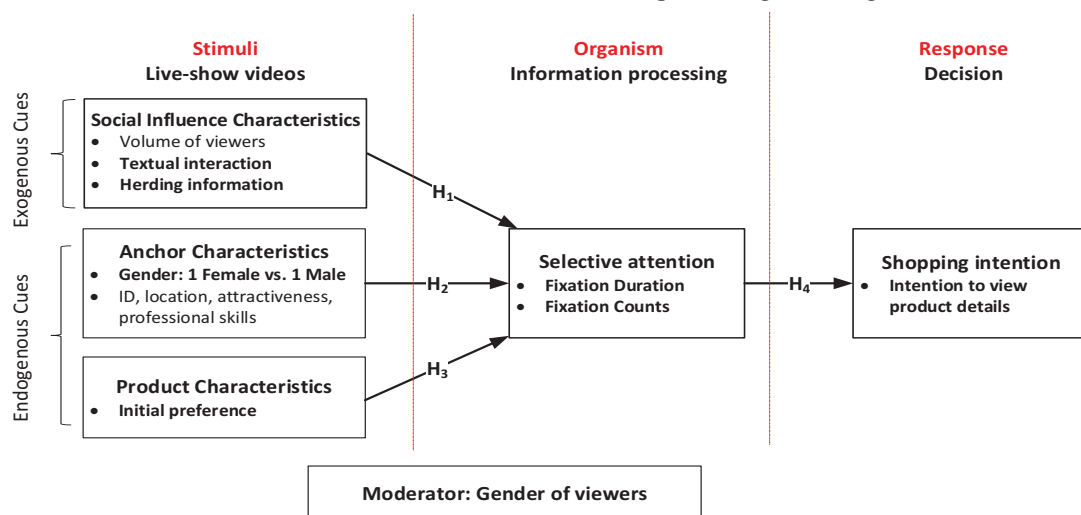
Existing studies showed that different gender had different response to the social influence. For example, Men showed more confidence when interacting via computer-mediated communication (CMC) with authority cues ^[11]. Besides the authority cues, there are other types of social influence such as herding cues, information cues and persuasions. Women often score higher values on factors related to communication (or information flow), security, and functionality than did their male counterpart ^[12]. Another study on online engagement through websites and social media showed that peer recommendations had a significantly stronger influence on attitudes of females than on attitudes of males ^[13]. Women will be more appealed when there appears a herding cue. This study aims to explore the herding cues on shopping intention in the context of live-show by using the eye-tracking tool. With the advantage of eye-tracking tool, such study will extend our understanding of social influence on fundamental level with physiology data.

2.4 Gender differences in shopping intention

Gender differences may lead to different shopping intention. There are strong evidences that females have indeed significantly less positive associations with the homo economics concept than men (Awad, 2008; Garbarino, 2004; Dittmar, 2004; Van Slyke, 2002). Females were more anxious about privacy in online shopping than males do ^[14]. An eye-tracking study showed that participants with high product involvement paid more attention to product and information cues ^[15]. Consumers with high product involvement show longer fixation duration to a positive advertisement cue of product ^[16]. Males showed higher enjoyment when they shopping utilitarian products online, while females preferred to shopping hedonic products online ^[17]. A fMRI experiments found that females activated more brain areas than males did in the context of online shopping ^[18]. Different genders process information differently in addressing cognitive problems, and lead to different purchase decisions ^[19]. Feminine activities are more focused in the pursuit of attachment, empathy, and playing a nursing and caring role ^[20]. So, existing evidences showed that women were more sensitivity to “sensory stimuli” than men. However, there are few studies examine the gender differences on shopping intention from the difference of visual pattern between male and female. How such visual attention difference influence shopping intention remain unknown.

3. RESEARCH MODEL

Based on the research questions and literature review, this study proposed a 2-phase model (Figure 2) to explain the gender differences on information processing and shopping intention in the context of live-show. This model is proposed on the stimulus-organism-response (S-O-R) framework. It was first introduced into marketing science to examine retail atmosphere from environmental psychology by Donovan and Rossiter (1982). Recently, it was often used to examine online store atmosphere (Eroglu et al., 2003; Dailey, 2004; McKinney, 2004; Wu, Cheng & Yen, 2008; Dong-Mo & Seon-Hee, 2010; Chen & Wu, 2016). In Taobao live-show, the video provides rich information for consumers. Among them, exogenous cues and endogenous cues were set as the “stimuli”. Selective attention of information processing was “organism”. And decision of



shopping intention was “response”.

Figure 2. The research model

3.1 Gender of viewers

To figure out the gender differences in selective attention and shopping intention in the live-show, gender of viewers was set as a moderator.

3.2 Exogenous cues and endogenous cues

In this study, endogenous cues are assumed to elicit voluntary shifts in attention while exogenous cues are thought to involve automatic processes^[1]. In the context of Taobao live-show, endogenous cues occupying the largest part of screen are the direct related shopping information which viewers seek for actively, whereas exogenous cues occurring unexpectedly in the corner are the indirect information which may distract viewers' attention. Because we selected two types of Taobao live-show video as experiment stimuli: One was female anchor perform women's coat and the other was male anchor perform men's clothing. Both anchors are manipulated to have similar attractive and professional level. The two types of video differ in product characteristics, anchor characteristics and social characteristics. Therefore, in this experiment, the endogenous cues refer to the area of anchor and product and exogenous cues mainly refer to the flickering areas of social influence cues.

3.3 Selective attention

When viewers receive the video information as the "stimuli", their brain as the "organism", are working to process this message for further decision. In this phase, different information cues will induce different attention of viewers and we focus on the process of attracting attention by eye-tracking method. According to the theory of selective attention^[1-4], there are 2 types of attention mechanism - endogenous and exogenous attention, which are attracted by different cues and may influence the final decision in different ways. To measure the selective attention of viewers, we focus on the fixation duration and count as the primary eye tracking metrics and also take other metrics like pupil size into consideration for correct judgement.

With the previous empirical evidences of gender differences in selective attention (Bayliss et al., 2005; Merritt et al., 2005, 2007; Spence & Pratt, 2007; Wang Liyan et al., 2010), it can be inferred that male and female viewers pay different attention to different cue types. And because fixation duration and count are related to consumers' cognitive processing and visual attention, longer fixation duration on area of interest (AOI) may indicate that the object on the AOI is more attractive (Ehmke & Wilson, 2007). Due to the "distraction effect" of social influence cues on female and male viewers, we hypothesize that:

H_{1a}: When herding information (or textual interaction) occurs more frequently, fixation duration on the AOI of exogenous cues will be significantly longer.

H_{1b}: Fixation counts on the AOI of exogenous cues with similar frequency will exist significant difference between male and female.

According to the previous evidence^[6-10], gender of both the participant (viewer) and the anchor showed influence on gaze patterns during watching live-show video. Viewers will primarily gaze at anchor's face. However, they will use different face-scanning strategy when processing different gender faces. Female viewer may follow an exploratory scanning strategy than males with longer and more fixations to the eyes, and both gender of viewers may look earlier and longer at opposite-sex anchors. Therefore, we hypothesize that:

H_{2a}: Fixation durations on the AOI of female and male anchor's face will exist significant difference between female and male viewer.

H_{2b}: Fixation counts on the AOI of female and male anchor's face will exist significant difference between female and male viewer.

Fixation reflected as attention can indicate interests in products (Jacob & Karn, 2003; Chae & Lee, 2013; Jing Luan et al., 2016). Product (referred as coat in this study) has itself characteristics like design style, quality and price, which leads to different first impression among viewers. When we focus on the perceived interests of products showed by viewers' initial preference of each product, products are divided into different attracting level. As for different gender of viewers and product characteristics, different attracted products also attracted

different attention with moderating effect of female and male viewers. And hence we hypothesize that:

H_{3a}: Fixation duration on the AOI of product will be significantly longer When viewers have higher initial preference of the product.

H_{3b}: Fixation counts on the AOI of product with similar initial preference will exist significant difference between female and male viewer.

3.4 Shopping intention

After processing video information, viewers are asked to decide whether they are willing to view the details of coat by clicking the product in shopping bag, which is the “response” of the “organism”. This study defines viewers’ willingness to view product details as shopping intention, as well as the dependent variable. It’s known that observed data of viewers’ attention with eye-tracking tool can be used to predict their intention (Milosavljevic & Cerf, 2008). On the basis of previous studies on gender differences in online shopping attitude [17-19], we argued that different selective attention of female and male viewers will lead to different shopping intention. In the context of watching Taobao live-show to select clothing, if viewers are interested in the presented clothing, they will fixate more on product and herding cues, which can be reflected by eye-tracking metrics – fixation duration and counts. Specifically, endogenous attention will improve shopping intention, while exogenous attention with social cues will compete for limited cognitive resources but possibly strengthen the positive influence of endogenous attention for herding influence on intention. In addition, we take account of gender difference in the influence of selective attention on final decision as the moderating role. Therefore, we hypothesize that:

H_{4a}: Longer fixation duration on the AOI of endogenous area will increase shopping intention.

H_{4b}: More fixation counts on the AOI of exogenous area will increase shopping intention.

H_{4c}: Fixation on exogenous area will have positive interaction influence with endogenous attention on shopping intention.

H_{4d}: Fixation of female and male viewers will have significantly different influence on shopping intention.

4. EXPERIMENT DESIGN

4.1 Stimulus and pretest

This study captured 12 videos in the Taobao live-show on mobile. There are 2 anchors, one is male and the other is female, and 6 videos for each anchor. To control the overlap between AOIs as possible as less, medium-sized coat was set as the clothing type. Each piece of videos lasted 40s. In each video, the anchor tried on coat themselves, showed the details of coat, and suggested the possible collection package of coat. For further studying about the social influence that the herding cues and comments brought, the amounts and frequencies of the herding cues and comments in the videos are counted.

This study argues that there will be different visual pattern occur when the gender is different among the participates and the different gender of live-show anchors. As for live-show anchors, there is different styles of anchors in Taobao website, they have different levels of professional skills, mandarin standards, sounds and faces. In order to focus on the influence of anchors’ gender on attention, there is a pretest designed for testing what the participates think the different levels of live-show anchors before the experiment. Now we have done such a pretest to test which live-show anchors will attract participates most and select one male and one female anchor with similar quality on these characteristics. There are 6 pieces of videos that had different live-show anchors including 3 male anchors and 3 female anchors. Participants are asked to answer several questions about attractiveness of appearance, sound characteristics and professional skills after watching each piece of video.

4.2 Measurement and manipulation control

This study explored the distribution of attention when viewers watching Taobao live-show video with an eye-tracking experiment. Eye movements typically are analyzed with respect to fixations measured normally in counts and duration. Numbers and duration of visual fixations, respectively, indicate how many and how long customers' eyes remain focused on a particular AOI on the target screen.

Table 1. Video information included in the AOIs

AOI	Cue type	Information
1	endogenous	a. anchor's face
		b. product (e.g. coat)
2	exogenous	a. herding information
		b. textual interaction

Each screen included 2AOIs which are consisted of supplementary information according to Table 1. The endogenous AOI includes two subareas with model's face and the product. The exogenous AOI includes two subareas with herding information and textual interaction. As outcomes from the eye-tracking experiments, this study tends to use two main metrics – total fixation duration (TFD) and fixation count (FC) which reflect participants' visual attention and cognitive processing when they are interested in a particular AOI (Lee & Ahn, 2012; Resnick & Albert, 2014). TFD represents the duration of fixation and FC represents number of fixations to a specific AOI. Furthermore, we will analyze other indexes to robust the conclusions, such as the fixation dwell times within each AOI, mean fixation and saccade durations, sequence and duration of AOI visits (Joseph, 2002).

To observe the social influence cues of female and male viewers, we need to take the number and frequency of flickering cues of each video into the consideration (total viewing volume nearly fixed is also included but not paid attention to). To investigate the influence of product characteristics on fixation, the selected products are all coats with similar types and unknown price and the perceived interests of products is measured by participants' initial preference scale of each product before videos. To check whether participants are actively listen to the anchors, we ask them to answer a follow-up question about some conversation detail in the video after each trial.

4.3 Participants and procedure

The main experiment will recruit 60 individuals from Zhejiang University. The experimental lab is equipped with a 5.5-in mobile phone, an adjustable chair and a table for the participants, and another table for the researcher. The participants' visual attention patterns were tracked and recorded with the Red-5 eye-tracking system, which uses a 250-Hz sampling rate. The experimental procedures will be as follows:

When a participant arrived, s/he was asked to complete the demographic questionnaire. They will be told that this research is going to test the efficiency of Taobao live-show. Their task is to actively watch videos and select proper coats for their friend. And after watching each video, they need to answer a follow-up question about video detail and decide whether to click the product detail hyperlink or not. At last, there's a post-experiment questionnaire about attitudes to live-show and evaluation of anchor when they finished the eye-tracking experiment. The whole experiment lasts for approximately 20 min for each participant.

5. DISCUSSION AND FURTHER RESEARCH AGENDA

Although this research is currently underway and initial results are being analyzed, it's expected to have important theoretical contributions and practical implications. First, it will improve our understanding on selective attention in the context of live-show, by using eye-tracking experiment which provides units of

analysis with the identification of multiple AOIs rather than simply discrimination on overall performance between subjects. Compared to traditional empirical methodologies such as self-report survey, neuroscience methods such as eye-tracking and electroencephalography (EEG) are more accurate and efficient to reflect individuals' subconscious cognitive processing (Luan et al., 2016). We employed eye-tracking method to testify our hypotheses with self-report and behavioral data to robust our conclusions, which will deepen our understanding about complex consumers' behavior and advance NeuroIS research promoted by many IS researchers (Dimoka, Pavlou, & Davis, 2011). Second, it will offer insight into selective attention influence on shopping intention and even herding behavior in live-show with gender as interaction terms. Third, it will help Taobao live-show platform to improve user-interface design and empower anchors with professional skills to achieve attractive promotion and patient communication with customers, which will lead to better customers' satisfaction with their decisions.

However, the proposed research still has some limitations that require future work to overcome. To simplify the experiment paradigm, several influencing factors of shopping intention in live-show were not investigated despite their potential relevance, (e.g., product involvement and individual characteristics) for pragmatic reasons related to complex reality and sample size. Further, only one cognitive variable (i.e., attention by eye-tracking) is observed clearly, and other bio-data can be obtained by other neurophysiological tools (e.g., EEG or fMRI) in future studies. Therefore, after get first knowledge on how viewers allocate their attention for different cues and produce shopping intention during live-show watching, we plan to figure out deep neural changes in viewer's brain and explain more detailed cognitive process from live-show stimulus to decision response. More specifically, we will combine with EEG measurement approach for assessing viewer's cognitive and affective change at stimulus onset time to investigate the role of herding cues in a follow-up live-show research.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71202157, Science Foundation of Ministry of Education of China under Grant 12YJC630251 and Social Science Foundation of Zhejiang Province under Grant 13ZJQN038YB.

REFERENCES

- [1] Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25.
- [2] Macaluso E. Orienting of spatial attention and the interplay between the senses[J]. *Cortex*. 2010, 46(3):282-297.
- [3] Hopfinger J B, West V M. Interactions between endogenous and exogenous attention on cortical visual processing[J]. *NeuroImage*, 2006, 31(2): 774–789.
- [4] Tang X et al. The interactions of multisensory integration with endogenous and exogenous attention[J]. *Neuroscience and Biobehavioral Reviews*, Elsevier Ltd, 2016, 61: 208–224.
- [5] Merritt P et al. Evidence for gender differences in visual selective attention[J]. *Personality and Individual Differences*, 2007, 43(3): 597–609.
- [6] Hewig J et al. Gender differences for specific body regions when looking at men and women[J]. *Journal of Nonverbal Behavior*, 2008: 1–12.
- [7] Shen J, Itti L. Top-down influences on visual attention during listening are modulated by observer sex[J]. *Vision Research*, Elsevier Ltd, 2012, 65: 62–76.
- [8] Baptista N I et al. Looking more and at different things: Differential gender eye-tracking patterns on an irony comprehension task[J]. *Psychology and Neuroscience*, 2015, 8(2): 157–167.
- [9] McIntyre M M, Graziano W G. Seeing People, Seeing Things: Individual Differences in Selective Attention[J].

- Personality and Social Psychology Bulletin, 2016, 42(9): 1258–1271.
- [10] Coutrot A et al. Face exploration dynamics differentiate men and women[J]. *Journal of Vision*, 2016, 16(14): 16.
- [11] Okdie B M et al. Social Influence Online: A Tale of Gender Differences in the Effectiveness of Authority Cues Bradley[J]. *International Journal of Interactive Communication Systems and Technologies*, 2013, 3(1): 20–31.
- [12] Murphy G B, Tocher N. Gender differences in the effectiveness of online trust building information cues: An empirical examination[J]. *Journal of High Technology Management Research*, Elsevier Inc., 2011, 22(1): 26–35.
- [13] Nadeem W et al. Engaging consumers online through websites and social media: A gender study of Italian Generation Y clothing consumers[J]. *International Journal of Information Management*, Elsevier Ltd, 2015, 35(4): 432–442.
- [14] Midha V. Impact of consumer empowerment on online trust: An examination across genders[J]. *Decision Support Systems*, 2012, 54(1): 198–205.
- [15] Behe B K et al. The effect of involvement on visual attention and product choice[J]. *Journal of Retailing and Consumer Services*, Elsevier, 2015, 24(C): 10–21.
- [16] Yang S-F. An eye-tracking study of the Elaboration Likelihood Model in online shopping[J]. *Electronic Commerce Research and Applications*, Elsevier B.V., 2015, 14(4): 233–240.
- [17] Rodgers S, Harris M A. Gender and E-Commerce: An Exploratory Study[J]. *Journal of Advertising Research*, 2003, 43(43):322-329.
- [18] Riedl R et al. Are There Neural Gender Differences in Online Trust? An fMRI Study on the Perceived Trustworthiness of eBay Offers[J]. *MIS Quarterly*, 2010, 34(2): 397–428.
- [19] Chang C. The Relative Effectiveness of Comparative and Noncomparative Advertising: Evidence for Gender Differences in Information-Processing Strategies[J]. *Journal of Advertising*, 2007, 36(1): 21–35.
- [20] Meyers-Levy J, Maheswaran D. Exploring Differences in Males’ and Females’ Processing Strategies[J]. *Journal of Consumer Research*, 1991, 18(1): 63.

Signaling Peer Trust in Accommodation-sharing Services: Effects of Similarity and Reviews on Listing Sales

Jiang Wu^{1*}, Mengmeng Jin²

¹Center for Studies of Information Resources, Wuhan University, Wuhan, 430072, China

²Center for E-commerce Research and Development, Wuhan University, China

Abstract: Online P2P accommodation-sharing has been a rising sub-market of sharing economy in recent years. However, the trust issues still exist because of information asymmetry and economic risks. Based on signaling theory, we argue that the traveler-host demographic similarity and host review volume both foster traveler's trust-building. By an empirical study, we found that the traveler-host age similarity and education similarity have a significant positive effect on listing sales. Furthermore, it is found that the host's review volume moderate the relationship of age and education similarity of traveler-host on listing sales. The findings take the initiative to verify the effect of traveler-host demographic similarity in sharing economy, which contributes to the accommodation-sharing literature theoretically and provides practical guidelines for developing trust.

Keywords: accommodation-sharing, demographic similarity, online review

1. INTRODUCTION

Over the past few years, as an emerging business model, the “sharing economy” has grown rapidly. According to PwC, the sharing economy market will grow to \$335 billion by 2025 ^[1]. A rising sub-market of sharing economy is the area of peer-to-peer (P2P) accommodation-sharing services, which happens when a host rents an apartment or a room they own to travelers through a digital platform such as Airbnb. According to the iResearch report, it is estimated that the P2P accommodation-sharing trading volume of China will be \$2.6 billion in 2018 ^[2]. The sharing economy in online short-term accommodation rental will still flourish in a long run. Although the P2P accommodation-sharing platforms provide direct interactions for travelers and hosts, the trust issues still exist. Since the suppliers and customers in P2P accommodation-sharing are both strangers, trust between them is the basis for trading and has a significant impact on consumer's decision-making ^[3]. Bacharach and Gambetta(2001) point out that trust can be regarded as a signaling problem^[4], that is to say, trust can be developed on the basis of perceived cues or signals that indicate trustworthiness about individual's identity.

Previous literature has discussed the following signals about trust. On the one hand, the service provider's profiles such as demographics and other personal information exposed on sharing economy platforms can facilitate consumers' trust on them. In P2P accommodation-sharing platforms, trust can be enhanced by host's profiles, which serve as a means of identification ^[5] and a way to increase the sense of personal, sociable, human contact ^[6]. Accordingly, we consider that people who are similar in demographics (e.g. age, hometown, education) tend to trust each other more easily. On the other hand, the reputation system including online review and rating is expected to encourage trust among traders, given the potential risks in trading with strangers in P2P marketplaces ^[7]. Exposure to online reviews offers additional information about listings for travelers to measure its quality and fit their needs and preferences, which further drive their perceived trust. Based on above research, we propose that online reviews, as the other trust signal, can also facilitate trust of travelers to hosts and positively affect the listing sales. Meanwhile, online reviews represent the predecessor feedback about listings and hosts, while peer similarity among travelers and hosts are traveler's self-identification which can be

* Corresponding author. Email: jiangw@whu.edu.cn(Jiang Wu)

substituted by predecessor feedback. Thus, we assert that there's a moderation relationship between the online review and the peer similarity of travelers and hosts.

In fact, little has been known about the specific effects of demographic similarities between travelers and hosts when compared to online reviews of hosts in sharing economy. This paper aims at answering the main effect and moderating effect of traveler-host demographic similarity and host's review volume on listing sales. Drawing on the signaling theory, we made an empirical study on an accommodation-sharing platform of China to address the research question.

2. LITERATURE REVIEW AND HYPOTHESIS

2.1 Signaling theory

Signaling theory is fundamentally concerned with reducing information asymmetry between two parties^[8]. The primary elements of signaling theory consist of signaler, signal, receiver, and feedback^[9]. In e-commerce, signaling is displaying of certain website features (signals) that carry information about an individual, product or firms from sellers (signaler) to buyers (receiver) and aim at motivating the buyer's behavioral intention (feedback). Signaling plays a significant role in resolving information asymmetries about the latent and unobservable quality of a product and evaluating the credibility and validity of a seller's qualities^[10]. In the online P2P accommodation-sharing context, a traveler not only rents a listing but also literally purchasing the host's offline service. On the one hand, the review volume of a host received is a primary signaling cue indicating his/her popularity and trustworthiness. On the other hand, as both traveler's and host's characteristics are important in P2P room-sharing transactions, a match of these characteristics can also be assumed to be important. Consequently, this study intends to integrate two focuses: demographic similarities and review volume. We are curious about the effect of two signals on listing sales and how they interact.

2.2 Demographic similarity

The effects of demographic similarity on trust in P2P transactions are valued and equivocal. According to the homophily theory and similarity-attraction theory^[11], similarity is "the degree to which pairs of individuals who interact are similar with respect to certain attributes, such as beliefs, values, education, social status, etc."^[12]. The homophily theory was first developed in social network verifying that individuals are inclined to make friends and socialize with people to whom they consider similar. Then the homophily theory was applied in marketing when considering the role of consumer's similarity on their decision-making. Take the example of online review, when consumers perceive the demographic similarity with reviewers, they express more trust and are more likely to be persuaded^[13]. Based on the homophily theory, we propose that demographic similarity, conceptualized as the similarity between travelers and hosts in demographic dimensions such as age, hometown, and education, influence consumers' decision-making by providing a means of identification^[5] and a way to increase the sense of personal, sociable, and human contact^[6] in the P2P accommodation-sharing context. The demographic similarity shortens the social distance between two parties and makes interaction among travelers and hosts easier and less challenging. The following hypothesis is thus developed:

H1: The traveler-host demographic similarities (the similarity of age, H1a; the similarity of hometown, H1b; the similarity of education, H1c) will positively affect listing sales.

2.3 Review volume and demographic similarity

2.3.1 Review volume and listing sales

As noted above, another major factor influencing hospitality or tourism performance is the number of

reviews, which has been proved by literature^[14]. We summarize the influence of review volume on listing sales from following aspects. First, since the online rating of listings on room-sharing platform has the J-shaped distribution^[15] and nearly 95% of Airbnb properties boast an average user-generated rating of either 4.5 or 5 stars^[16], which means the lack of variance makes the online rating miss enough information value for travelers comparing to the number of reviews. Second, as an information source, one of the basic functions of online reviews is to provide information to travelers who have little prior knowledge about the listings^[17]. Therefore, review volume determines how much information can obtain. The higher review volume, the more uncertainty of consumers can be reduced from peer evaluation. Neirotti (2016) point that the number of reviews has positive and significant moderation effect on revenue growth in hospitality industry^[18]. In line with this empirical evidence, we hypothesize:

H2: The review volume of hosts received will positively affect listing sales.

2.3.2 Combining review volume and demographic similarity

Studies combining review volume and demographic similarity are surprisingly rare. We assert that online reviews and demographic similarity are both cues for consumers to evaluate the offering's quality although working in two different ways. The demographic similarity is the degree of a traveler's similarity perception to a host, which belong to the traveler's self-identification. Meanwhile, review volume of hosts received represents the predecessor's feedback to the listing and host. More reviews, more information may contain. Studies have shown that there is a substitution effect between self-identification and predecessor' feedback^[13]. When demographic information is ambiguous, review volume can play an important role in compensating for information lack in the online environment and providing helpful cues for consumers to make decisions. Therefore, we expect the review volume of hosts received to be the primary source for trust and to show a stronger effect than demographic similarity with regard to the formulation of trusting beliefs and trusting behavior like making a reservation. Hence, we assert the impact of demographic similarity to depend on the review volume of hosts received, leading to the hypothesis:

H3: The review volume of hosts received moderate the effect of traveler-host demographic similarities (the similarity of age, H3a; the similarity of hometown, H3b; the similarity of education, H3c) on listing sales.

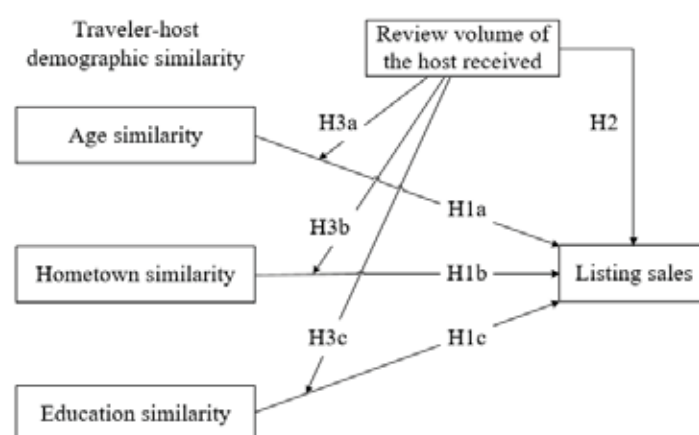


Figure 1. Research model

3. METHODOLOGY

3.1 Research context

To empirically test our research model shown in Figure 1, we collected data from *Xiaozhu* (xiao-zhu.com), a

clone of Airbnb that has led to a peer-to-peer accommodation-sharing platform in China. The online accommodation-sharing platform *Xiaozhu* is activated in August 2012 and now offers 200,000 online listings in more than 300 domestic cities^[19]. As one of the leaders in the accommodation-sharing market in China, it's worth making research on the real-world data generated from the platform to reveal the effects of demographic similarity of traveler-host and review volume of a host received and their interaction effect on listing sales. To fulfill the research requirement, the demographic information of travelers and hosts, review volume of hosts received, and listing sale history need to be obtained. Fortunately, *Xiaozhu.com* provides us all information we need.

3.2 Data and measures

Using an automated Python-based script, we developed a web-crawler to retrieve and collect information relating to all travelers, hosts, and listings. Our sample covers 104,416 listing sale history made by 65,543 unique travelers who stayed at 14,358 unique listings of 7,069 hosts in 18 cities between August 5, 2012, and September 16, 2016. The unit of analysis in our research is *Listing-Check in date*, which allows us to track every unique listing over multiple points of time. Our dependent variable, *Listing_cumOrder* is a numerical variable representing the cumulative number of listing sales. The independent variables are categorized into five dimensions, including *Traveler-host Similarities*, *Moderator*, *Traveler Control*, *Host Control*, and *Listing Control*. The specific variable definitions and descriptive statistics are listed in Table 1. The correlation values among explanatory variables are below 0.8 indicating that the estimation is unlikely to be biased by collinearity of variables.

Table 1. Variable Definition and Summary Statistics

Dimension	Variable	Definition	Obs	Mean	SD	Min	Max
Dependent Variable	Listing_cumOrder	Cumulative number of orders that a listing received	104,416	13.93	17.96	1	172
Demographic Similarities	Age_sim	Dummy variable of whether a host and a traveler are born in the same age, with values of 1=yes, -1=no, and 0=either of age information of hosts or travelers is missing	104,416	-0.01	0.20	-1	1
	Hometown_sim	Dummy variable of whether the hometown of a host and a traveler are the same provinces, with values of 1=yes, -1=no, and 0=either of traveler's or host's hometown information is missing	104,416	-0.03	0.20	-1	1
	Education_sim	Dummy variable of whether a host and a traveler have the same education level, with values of 1=yes, -1=no, and 0=either of traveler's or host's education information is missing	104,416	0.00	0.22	-1	1
Moderator	Host_cumReview	Review volume of the host has received	104,416	34.22	49.38	0	471
Traveler Control	Membership	Number of days since a traveler registered on Xiaozhu	104,416	88.64	156.40	0	1528
	PhoneVeri	Dummy variable of whether a traveler provides verified phone number, with values of 1=yes and 0=no	104,416	1.00	0.05	0	1

	EmailVeri	Dummy variable of whether a traveler provides verified email, with values of 1=yes and 0=no	104,416	0.11	0.31	0	1
	SocialVeri	Dummy variable of whether a traveler provides verified social media account such as WeChat, with values of 1=yes and 0=no	104,416	0.22	0.41	0	1
Host Control	Hmembership	Number of days since a host registered on Xiaozhu	104,416	247.43	269.49	0	1492
	ReplyRate	Number of online replies of a host versus number of inquiries of online shoppers	104,416	0.95	0.08	0	1
	ConfirmTime	Average number of minutes a host takes to respond to a reservation request	104,416	5.26	10.48	0	598
	AcceptRate	Number of accepted reservations versus number of reservation requests	104,416	0.89	0.11	0	1
	NameVeri	Dummy variable of whether a host provides verified name, with values of 1=yes and 0=no	104,416	0.32	0.47	0	1
	NumList	Number of listings owned by a host	104,416	5.65	7.07	0	82
	Gender	Dummy variable of host gender, with values of 1=female and 0=male	98,574	0.65	0.48	0	1
Listing Control	Area	Area of a listing in square meter (m ²)	104,416	56.13	47.61	1	1200
	Bedroom	Number of bedrooms of a listing	104,416	2.09	1.19	0	21
	Livingroom	Number of living rooms of a listing	104,416	1.12	0.70	0	5
	Bathroom	Number of bath rooms of a listing	104,416	1.28	0.71	0	22
	Kitchen	Number of kitchens of a listing	104,416	0.91	0.31	0	11
	Balcony	Number of balconies of a listing	104,416	0.95	0.72	0	18
	Bed	Number of beds of a listing	104,416	1.82	1.15	1	21
	Price	Rate per night of a listing in Chinese Yuan	93,642	286.30	238.64	28	8000
	Category	Nominal variable of listing type, with 1=listing shared with the host (base category), 2=listing shared with other travelers, and 3=private listing	104,416	2.47	0.83	1	3

3.3 Model specification

The overall aim of this study is to examine how traveler-host demographic similarity and host's review volume influence and interact on listing transactions. Because our sample is panel data, the Hausman test was performed before empirical analysis to decide whether the fixed effect should be applied or the random effect in our model. The Hausman result shows that the fixed effect fits our model more. Although the fixed effect model drops the time-invariant explanatory variables in the regression, it is still considered an effective estimation method for controlling the unobserved heterogeneity. We use the ordinary least squares (OLS) regression with fixed effect estimations to examine our regression model as shown in equation (1).

$$\begin{aligned}
 Listing_cumOrder_t = & \alpha + \beta_1 Age_sim + \beta_2 Hometown_sim + \beta_3 Education_sim + \beta_4 Host_cumReview_{t-1} * \\
 & Age_sim + \beta_5 Host_cumReview_{t-1} * Hometown_sim + \beta_6 Host_cumReview_{t-1} * Education_sim \\
 & + \psi Traveler + \phi LIST + \Gamma HOST + \varepsilon
 \end{aligned} \tag{1}$$

4. RESULTS

4.1 Empirical results

Collinearity check was performed to ensure the accuracy of estimation before analyzing the regression model. The collinearity indicator, variance inflation factors (VIF) values of all independent variables are below 5, which indicates there's no multicollinearity lie in our research model. We then ran the OLS model with fixed effect to analyze our model, as presented in Table 2.

The empirical result of our focal explanatory variables is presented in column 5 of Table 2. Time-invariant explanatory variables were dropped because of fixed effect. We can see that the empirical results of three traveler-host similarity variables partially support our hypothesis. To be specific, the *Age_sim* ($\beta=1.037$, $p=0.000$) and *Education_sim* ($\beta=0.558$, $p=-0.004$) show a significant and positive influence on listing sales. The travelers and hosts that were born in same age or educated to the same level can be more appealing to each other comparing to individuals that were not. Hence, H1a and H1c are supported. Differ from the above results, the *Hometown_sim* ($\beta=0.313$, $p=-0.133$) of travelers and hosts fails to have a positive effect on listing sales significantly, meaning that the travelers do not prefer the hosts who come from the same province, rejecting H1b. This may arise from traveler's motivation to accommodation-sharing platform, which largely includes the novelty-seeking and local information acquisition. Travelers tend to pursue unique and novel experience and get a deeper understanding of local customs, which foster them to reserve the local host's listings. In line with Hypothesis 2, we further find that the *Host_cumReview* positively influence the listing observations indeed ($\beta=0.232$, $p=0.000$). The more review volume of hosts received, the more trust that travelers build on hosts, the more reservations that travelers may contribute to. As expected in H3a and H3c, the *Review_age_sim* ($\beta=-0.023$, $p=0.000$) and *Review_education_sim* ($\beta=-0.006$, $p=-0.013$) pass the hypothesis test by showing a negative effect on listing sales, meaning that the review volume of hosts received can negative moderate the effect of traveler-host similarity of age and education on listing sales. That is, as the growth of host's review volume, the age and education similarities of traveler-host are less important in traveler's decision making. We further find that *Review_hometown_sim* ($\beta=-0.002$, $p=-0.489$) don't make significance on listing sales, indicating that there's no moderating effect between the review volume of hosts received and the hometown similarity of traveler-host. H3b was not supported.

Table 2. Estimation results and robustness check

Dependent Variable: <i>Listing_cumOrder</i>					
Independent variables	Model1	Model2	Model3	Model4	Robustness check
<i>Age_sim</i>		-0.178	-0.266*	1.037***	1.395***
		(0.312)	(0.077)	(0.000)	(0.000)
<i>Hometown_sim</i>		0.433**	0.214	0.313	0.173
		(0.016)	(0.163)	(0.133)	(-0.530)
<i>Education_sim</i>		0.399**	0.260*	0.558***	0.594**
		(0.015)	(0.063)	(0.004)	(-0.018)
<i>Host_cumReview</i>			0.232***	0.232***	0.167***
			(0.000)	(0.000)	(0.000)
<i>Review_Age_sim</i>				-0.023***	-0.024***
				(0.000)	(0.000)
<i>Review_Hometown_sim</i>				-0.002	0.001

				(0.489)	(-0.651)
Review_Education_sim				-0.006**	-0.007**
				(0.013)	(-0.013)
Traveler control					
Membership	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PhoneVeri	2.629***	2.650***	1.974***	1.976***	2.252***
	(0.000)	(0.000)	(0.001)	(0.001)	(-0.002)
EmailVeri	1.891***	1.934***	1.332***	1.347***	1.627***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SocialVeri	-0.196**	-0.194**	-0.136**	-0.136**	-0.086
	(0.016)	(0.017)	(0.050)	(0.049)	(-0.377)
Host control					
Hmembership	0.059***	0.059***	0.029***	0.029***	0.047***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-3.569***	-3.590***	-3.148***	-3.138***	-5.022***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
<i>R-square</i>	0.465	0.465	0.612	0.612	0.628
Observations	88469	88469	88469	88469	50286

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.2 Robustness check

In our empirical model, we did not consider the effect of location city of listings on listing sales. Therefore, robustness check was performed to verify that the effect of traveler-host similarity and host's review volume on listing sales across the different cities. We kept the listings from the top three marketplace of *Xiaozhu.com* which is Beijing, Shanghai, and Chengdu as our robustness check sample, and ran our estimation model again to exclude the effect of cities on our empirical result. As shown in column 6 of Table 2, the estimation results are highly consistent with our main results, which further provide a strong evidence for our estimated results.

5. CONCLUSIONS

In this paper, we focused on the traveler-host demographic similarity and host's review volume and investigated their effect on listing sales. We addressed the research question by proposing detailed hypotheses concerning the relationship between the traveler-host demographic similarity, host's review volume, and listing sales. The OLS model with fixed effects was employed to empirically validate our research model. Results show that the traveler-host demographic similarity of age and education have a significant positive effect on listing sales. Further, the host's review volume presents a positive effect on listing sales and a negative moderating effect on the relationship between the demographic similarity of age and education and listing sales. Based on the results, both theoretical and practical implications were presented above. Although this paper offers important contributions to both theory and practice, the limitations still exist. First, the host's characteristic data in our dataset was acquired from the personal homepage of them. However, not all hosts of *Xiaozhu.com* have opened his/her personal page, which makes our dataset may not contain all hosts and their listings. Second, we use the host's review volume as a moderator variable. We argue that the review score and valence of hosts received are also important indicators to represent online reviews, which can be studied as future directions.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71373194.

REFERENCES

- [1] PwC. (2015). The sharing economy. www.pwc.com/us/en/industry/entertainment-media/publications/consumer-intelligence-series/assets/pwc-cis-sharing-economy.pdf
- [2] Iresearch. (2017). 2017 China online accommodation-sharing industry research report. www.iresearch.com.cn/report/2944.html(in Chinese)
- [3] Finley, K. (2013) Trust in the Sharing Economy: An Exploratory Study. Centre for Cultural Policy Studies. Coventry: University of Warwick.
- [4] Bacharach, M., & Gambetta, D. (2001). Trust in Signs. Trust in society.
- [5] Liu, C. S. (2012). A couchsurfing ethnography: traveling and connection in a commodified world. *Student Pulse*, 4(07).
- [6] Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, 18(12), 1192-1217.
- [7] Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in internet transactions: empirical analysis of ebay's reputation system. *Advances in Applied Microeconomics*, 11(02), 127-157.
- [8] Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American Economic Review*, 92(3), 434-459.
- [9] Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: a review and assessment. *Journal of Management Official Journal of the Southern Management Association*, 37(1), 39-67.
- [10] Mavlanovaa, T., & Koufaris, M. (2012). Signaling theory and information asymmetry in online commerce. *Information & Management*, 49(5), 240-247.
- [11] Byrne, D. E. (1971). The attraction paradigm. Academic Press.
- [12] Mcpherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27(1), 415-444.
- [13] Chan, I. C. C., Long, W. L., Chow, C. W. C., Fong, L. H. N., & Law, R. (2017). The effect of online reviews on hotel booking intention: the role of reader-reviewer similarity. *International Journal of Hospitality Management*, 66, 54-65.
- [14] Xie, K. L., Chen, C., & Wu, S. (2016). Online consumer review factors affecting offline hotel popularity: evidence from tripadvisor. *Journal of Travel & Tourism Marketing*, 33(2), 211-223.
- [15] Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *ACM*.
- [16] Zervas, G., Proserpio, D., & Byers, J. (2015). A first look at online reputation on airbnb, where every stay is above average. *Social Science Electronic Publishing*.
- [17] Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: the impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1), 123-127.
- [18] Neirotti, P., Raguseo, E., & Paolucci, E. (2016). Are customers' reviews creating value in the hospitality industry? exploring the moderating effects of market positioning. *International Journal of Information Management*, 36(6), 1133-1143.
- [19] Iresearch. (2017). 2017 China online accommodation-sharing industry research report: the case of Xiaozhu. www.iresearch.com.cn/report/3051.html(in Chinese)

The Impact of Board Capital on Performance of China's Listed Tourism Companies Based on the Moderating Effect of Leadership Structure^①

Li Muchun, Xu Tianqi, Liu Tingting

School of Economics and Commerce, South China University of Technology,
Guangzhou, 510006, China

Abstract: Board capital shape how directors govern and offer advice to the firm and affect the ideas and resources that they provide. Based on data analyzed over a six-year period with a sample of 24 listed tourism companies in China, this paper examines the board capital and the moderating effect of leadership structure on firm performance. The results indicate that: Directors' educational level has a negative effect on firm performance, and board' leadership structure negatively moderates this effect; However, board's leadership structure positively moderate the correlation between the board's political resources and firm performance. The enlightenment of this article is that board capital in China's listed tourism companies has not been fully utilized, and leadership structure positively moderates the correlation between board's political resources and firm performance.

Keywords: Listed Tourism Companies, Board Capital, Leadership Structure, Firm Performance

1. INTRODUCTION

Since the Reform and Opening, China's tourism industry has witnessed rapid development with the continuous expansion of its industrial scale and the gradual improvement of its industrial system. It has become a major tourist country in the world and is moving toward tourist powerhouse. In the "13th Five-Year Plan for the Development of Tourism" issued by the State Council in 2016, the total size of the tourism market will reach 6.7 billion by 2020, the total investment in tourism will reach 2 trillion yuan, and the total tourism revenue will reach 7 trillion yuan, tourism industry's comprehensive contribution to the national economy reached more than 12%. In addition, the "Planning" also made an important strategic plan for the development of the tourism industry, which injected a great impetus into the development of the industry. As of the end of 2016, there were 46 listed tourism companies in China (including A shares, Hong Kong stocks and US stocks).

Qin et al.^[1] pointed out that corporate governance is the basic strategy to promote the development of the tertiary industry and is also one of the motivating mechanisms to promote the development of the entire tertiary industry in the market competitiveness. Corporate governance, in turn, can be seen as the board's assurance of management's long-term value through sustainable management. Lu^[2] also pointed out that the first principle of modern corporate governance is that the company's business and affairs are managed under the guidance of the board of directors. The key factor that a corporation can surpass natural persons as a legal person lies in the board of directors.

At present, the domestic research on the board of directors of tourism enterprises is mainly based on two aspects: one is the corporate governance level, the other is to study the governance of the board of directors as an independent research object. However, no matter from the level of corporate governance or board governance, most scholars neglect the important role of board capital, one of the important features in the governance of the

^① **Corresponding author:** Muchun Li, associate professor of South China University of Technology
E-mail: limch@scut.edu.cn

board of directors in the research process. In fact, under the framework of modern enterprise system, in addition to its board structure, behavioral characteristics and motivational characteristics, the effectiveness and role of the board must be based on characteristics such as the functional background and political resources. Zhou and Li^[3] also pointed out that the two companies may have different value performance due to different board skills and social networks, even if the size of the board of directors is the same as the proportion of independent directors and implementing similar incentive mechanisms.

At present, China is in the stage of economic transition. Listed companies face the intense pressure of market competition and their thirst for resources is getting stronger and stronger. And the human capital and social capital owned by the directors who are the company's decision makers have become a channel for enterprises to obtain resources. Therefore, based on the data of China's listed tourism companies from 2011 to 2016, this paper comprehensively examines the impact of board capital on the performance of China's listed tourism companies from the human resources and social capital of the board, and further explores whether the leadership structure of the board has a moderating effect.

2. LITERATURE REVIEW AND RESEARCH HYPOTHESIS

The board of directors, as a "passerby" of the company, plays an important governance and decision-making role in the company's response to the rapidly changing market competition^[4]. Before the concept of the board capital was proposed, scholars mainly studied the surface properties of the board of directors, that is, the influence of board size, composition, incentive mechanism and independence on firm performance. These studies did not break through the structural problems of corporate governance and were difficult to answer the questions such as how to influence the creation of enterprise value. From the external characteristics of the past to the internal substance, this paper examines the human resources provided by the board of directors and the resources provided by social capital for the company, which helps to uncover the "black box" of the board of directors.

The board capital is primarily used to measure the board's ability to provide resources to the company, including board human capital and board social capital^[5]. Among them, the board of directors' human capital includes the directors' age, educational level, occupational background, tenure as well as the heterogeneity of human capital, which is the collective ability of directors and individuals to provide the board with resources such as knowledge and skills^[6], including the internal and external networks owned by the directors and the various real or potential resources brought by these networks^[7]. Board capital can enhance the board's influence and independence, also the oversight of managers and control over the formulation and execution of corporate strategies^[8], bringing ideas, perceptions, experiences, business knowledge and diversified decision-making, which is conducive to the board of directors to make a clearer judgment on the complicated external environment and the formation of innovative thinking^[3], the board capital can also strengthen the connection between the enterprise and the external environment and help enterprises to develop their skills, business contracts, reputation and legitimacy and other key resources, and create a relatively superior institutional environment for the enterprise^[9]. Therefore, the board capital is conducive to improving the governance efficiency of the board and the improvement of firm performance.

2.1 Board of directors' capital and business performance

2.1.1 Educational level

Educational level refers to the knowledge base, cognitive ability, the sense of worth and the influencing effect on their behavior^[3]. Studies by Datta and Rajagopalan^[10] show that there is a positive correlation between educational level and knowledge, skills, resilience, creativity, and information processing abilities. The higher

educational level of the board of directors, the more comprehensive the knowledge structure of the board will get. Based on the Resource Dependence Theory, Chen Yue et al.^[11] explored the impact of board capital on firm performance from the perspectives of social capital and human capital. The results show that the educational level of the board of directors is positively correlated with the company ROA. Based on this, this paper proposes the following hypothesis:

H1: There is a positive correlation between educational level of the board and firm performance in China's listed tourism companies.

2.1.2 Occupational background heterogeneity

Hillman et al.^[5] argue that directors of different occupations play different roles in the board of directors and provide different resources to the company. Studies by Haynes and Hillman^[12] confirm that the larger occupational heterogeneity of the board, the more likely the company's strategy will be creative. According to the Resource Dependence Theory, Feng Ruixue et al.^[13] tested the relationship between board capital and corporate value based on the data of listed companies in Shanghai Stock Exchange from 2007 to 2011. The results show that the heterogeneity of the board of directors can significantly enhance the corporate value, and pointed out that in the economic situation tends to be complex, the diversification of human capital and experience of directors will help companies to avoid risks in various decision-making, and in the specific circumstances to use their rich experience and specific knowledge structure to help companies make the right choice, so as to promote the improvement of enterprise value. At present, China is in the period of economic transition, the pressure in the international market and the complexity in the capital market require that the directors of the enterprises should deal with the rapid changes through their rich experience besides being familiar with the business in this industry, to ensure that the board is capable of playing the role of the highest decision-making body in the crucial period of the company. Based on this, this paper proposes the following hypothesis:

H2: There is a positive correlation between occupational heterogeneity and firm performance in China's listed tourism companies.

2.1.3 Interlocking directorates

Interlocking directorates refers to the phenomenon that a director holds the position in two or more companies simultaneously^[14]. The network of interlocking directorates arising from the board's personal presence on the boards of two or more enterprises not only facilitates the exercise of the supervisory function of the board, but also the board's resource provisioning function and thus improves the efficiency of board governance. At present, China is in the period of economic transition and the capital market is still in the development stage^[13]. If under the conditions of a mature market economy, the relationship between enterprises is only an addition to the formal system, but under the condition of transition economy, the network relations among enterprises are not just an alternative to the formal system and play a leading role^[15]. Through empirical analysis, Chen et al.^[11] concluded that the proportion of interlocking directorates in listed companies in China is positively correlated with firm performance. That is, the more chain directors, the more resources the enterprise will obtain. Therefore, this paper proposes the following hypothesis:

H3: There is a positive correlation between interlocking directorates and firm performance in China's listed tourism companies.

2.1.4 Political resources

The political resources among board members not only reduces the uncertainty and complexity of the business environment in the external market of the enterprise, but also helps the enterprise to obtain extra scarce resources so as to help the enterprise to establish and maintain its competitive advantage and position^[11]. Agrawal and Knoeber^[16] demonstrated that politically-backed directors can advise companies how to deal with

the government. When new members with political ties join the board, the stock of the company will yield extraordinary rate of return. At present, the capital market of China is still in the stage of development. The special stage of its economic development determines that the connection with the government has become an important "resource" for ensuring the smooth conduct of business activities. Directors with political resources will play an important role in the enterprise. The communication with the government can help companies get government support and thus more resources^[13], thus contributing to the improvement of business performance. Based on this, this paper proposes the following hypothesis:

H4: There is a positive correlation between political resources of the board and firm performance in the China's listed tourism companies.

2.2 The moderating effect of leadership structure

Board leadership structure refers to whether the CEO should simultaneously serve as the chairman of the board. There is much controversy about the separation between the two roles of chairman and CEO. The hypothesis of "the Separation of the Roles of Chairman and CEO" based on the Principal-agent Theory holds that the selfishness and bounded rationality of the human being make it naturally lazy and opportunistic. The duality of leadership structure enables the board of directors to be controlled by the managers, which will have bad impact on the interests of the company, while the separation of two roles helps to strike a balance between rights, enhance accountability and improve board independence^[2]. Brickley et al. (1997) also pointed out that the cost of transferring a large amount of knowledge and experience owned by managers to the chairman of the board is enormous and the combination of two positions avoids such costs and thus helps improve company performance. It's contradictory to the hypothesis of "Separation of two Positions" of Principal-agent Theory. How the current leadership structure has an impact on the company's business performance depending on the relative intensity of the two positions in the company.

At present, China is in an economic transition period with a high environmental uncertainty. The leadership structure with one post helps to enhance the freedom of decision-making of the general manager, fully and effectively promotes and enhances its decision-making ability, and effectively promotes organizational learning ability. In addition, the leadership structure of one board with two posts has increased the rights of the board of directors. At the same time, it has also promoted the opportunities for shareholders and affiliates to provide sufficient external resources and information for the listed companies, thereby enhancing the organizational learning ability and the company innovation^[17]. During the period of institutional transformation, China is more suitable for the leadership structure of the board of directors with one post. First, the CEO as the chairman of the board can well overcome top management conflict, to ensure the consistency of decision-making orders; second, the combination of two positions gives the CEO greater rights and sense of responsibility, making investors increasingly believe that enterprises have clear leadership and development goals; third, the two positions in one can promote the exchange of information between the top management team and the board then make better decisions. Therefore, when the leadership structure of the board is a combination of the two posts of the chairman and CEO, it is conducive to strengthen the leadership of the board, thereby enhancing the cooperation effect and further stimulate the board's potential to create value. Based on this, this paper proposes the following hypothesis:

H5: Board leadership structure has a positive moderating effect between educational level and firm performance.

H6: Board leadership structure has a positive moderating effect between occupational heterogeneity and firm performance.

H7: Board leadership structure has a positive moderating effect between interlocking directorates and firm performance.

H8: Board leadership structure has a positive moderating effect between political resources and firm performance.

3. RESEARCH DESIGN

3.1 Sample selection and data source

This paper takes China's listed tourism companies as the research object and chooses China's listed tourism companies in Shanghai and Shenzhen Stock Exchange from 2011 to 2016 as samples, excluding ST, PT, B shares and other listed companies whose main businesses have changed. Finally, there are 24 listed companies that meet the requirements, for a total of 144 observations. All the financial data and capital related to the board required in this paper are mainly from the annual reports of listed tourism companies. The other data are from the websites of Shanghai and Shenzhen Stock Exchange, Juchao Website and Sina Finance Network. All the data are processed and calculated by EXCEL and SPSS18.0.

3.2 Variable selection and definition

(1) Dependent variables

This paper analyzes the profitability (earnings per share, total assets profit rate, net profit margin on sales, ROE), solvency(current ratio, quick ratio, cash ratio, property ratio), growth ability(main business revenue growth rate, total assets growth rate), operating ability(fixed asset turnover, total asset turnover, current assets turnover) and cashing ability (return on operating cash flow of assets, operating cash flow to debt ratio, operating cash flow ratio) through the Principal Component Analysis to a full measure of China's listed tourism companies operating performance.

First, the KMO and Bartlett tests should be performed on original variables. Among them, KMO is used to test the correlation between the original variables. In general, the closer the KMO value is to 1, the stronger the correlation between the original variables. In this paper, KMO = 0.750, indicating that the correlation between the original variables is strong and suitable for the Principal Component Analysis; Bartlett spherical test for the test of the original variables are independent, the spherical test approximate chi-square value of 2665.630, and the significance level of less than 1%, indicating that the original variable composition matrix is not an identity matrix. Therefore, the null hypothesis of spherical test can be rejected, and it is considered that the Principal Component Analysis is suitable for the original variable. Then, using the Principal Component Analysis of these 16 indicators of factor analysis, the interpretation of the total variance shown in Table 2. According to the principle that the eigenvalue is greater than 1, a total of five factors are extracted, and the contribution rates of the first to the fifth factors are 25.618%, 18.671%, 17.464%, 13.276% and 10.472% respectively. The total variance explained rate reached 85.501%, indicating that these five common factors contain most of the information reflected by the 16 original variables. Finally, the score of each common factor is calculated separately, and the total performance of each company is calculated by taking the weight of each public factor's contribution to the cumulative variance as the weight. The formula of comprehensive performance index is as follows: $P = (25.618 \times \text{fact_1} + 18.671 \times \text{fact_2} + 17.464 \times \text{fact_3} + 13.276 \times \text{fact_4} + 10.472 \times \text{fact_5}) / 85.501$.

(2) Independent variables

Educational level: The educational level of the board is measured by the average number of board members with the highest educational level. Assignment of variables are: 4 for doctor's degree, 3 for master's degree, 2 for bachelor's degree and 1 for junior college degree and others.

Occupational heterogeneity: using the Herfindahl Index which is:

$$H = 1 - \sum_{i=E}^n P_i^2 \quad (1)$$

In this formula, E refers to the professional background of the directors, including business experts, support specialists and social influencers; P_i refers to the proportion of directors with the i background in the board of directors; H represents the team differences and the qualitative level is between 0 and 1, and the greater the H value, the higher the heterogeneity will be.

Interlocking directorates: This variable is measured by the ratio of the number of directors serving in two or more companies to the size of the board.

Political resources: the proportion of directors who are in the work of the central government, the local government, National People's Congress and the CPPCC(the Chinese People's Political Consultative Conference), etc, as a percentage of the size of the board.

(3) Control variables

The control variables in this article include company size, debt asset ratio and board size. Table 1 shows the definition of all variables and their calculation.

Table 1 Research Variables and Definitions

Variable	Symbol	Variable Definitions
Performance Index	P	Based on Principal Component Analysis
Educational Level	X1	Average of the highest educational level of board members
Occupational Background Heterogeneity	X2	H Index
Interlocking Directorates	X3	Number of interlocking directorates / Total number of board members
Political Resources	X4	Directors who have political background/Total number of board members
Leadership Structure	X5	If chairman and CEO are in one positions, the value is 1;if not, the value is 0
Company Size	X6	Natural logarithm of total assets
Assets Liabilities Ratio	X7	Total liabilities /Assets
Board Size	X8	The total number of board directors

3.3 Model design

Based on the above analysis, this paper builds a data model of the influence of board capital and leadership structure on firm performance, examining the research hypothesis in this paper, and the basic form of the equation is as follows:

$$P = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_1 \times X_5 + \beta_7 X_2 \times X_5 + \beta_8 X_3 \times X_5 + \beta_9 X_4 \times X_5 + \beta_{10} X_6 + \beta_{11} X_7 + \beta_{12} X_8 + \xi$$

In this equation, β_0 is the intercept, $\beta_1 \sim \beta_{12}$ is the coefficient, $X_1 \times X_5 \sim X_4 \times X_5$ is the interaction term, and ξ is the residual.

4. EMPIRICAL ANALYSIS

4.1 Descriptive statistical analysis of variables

Through the descriptive statistical analysis of the board capital and operating performance, the results are shown in Table 2. As can be seen from Table 2, the minimum performance index of the company is -1.07, the maximum is 2.15 and the average is 0, indicating that most of China's listed tourism companies do not add value to the company's performance. The average educational level of the board of is 2.59. The standard deviation is 0.42, indicating that the educational level of the board members is relatively high and the average level is above the undergraduate level. This may because most independent directors in the board of China's listed tourism companies have well-educated college teachers who generally hold Ph.D., which to some extent raised the average educational level of directors; average occupational heterogeneity of the board is 0.53; the interlocking

directorates ratio of board is 0.52, which means that half of the people in the company have two positions. There is a huge network relationship between directors and enterprises; the minimum political resources owned by the board is 0, the maximum is 0.88, and the mean is 0.15, indicating that the average board political resources possessed in the research sample are low.

Table 2 Descriptive Statistics of Variables

Variable	Sample Size	Minimum	Maximum	Mean	Standard Deviation
Performance Index P	144	-1.07	2.15	0.00	0.47
Educational Level	144	1.56	3.44	2.59	0.42
Occupational Background Heterogeneity	144	0.00	0.67	0.53	0.12
Interlocking Directorates	144	0.00	0.93	0.52	0.25
Political Resources	144	0.00	0.88	0.15	0.18
Leadership Structure	144	0.00	1.00	0.12	0.32
Company Size	144	10.71	16.50	12.46	1.20
Assets	144	0.02	0.80	0.36	0.18
Liabilities Ratio	144	5.00	15.00	9.38	1.63
Board Size	144	5.00	15.00	9.38	1.63

4.2 Empirical results and analysis

In order to avoid multicollinearity problems among the variables in the regression analysis, correlation analysis was conducted among the variables, and the analysis results are shown in Table 3. The data show that the correlation between the variables is small, there is no serious multicollinearity problems, and multivariable regression can be carried out.

Table 3 Pearson Correlation Coefficient Matrix

Variable	P	1	2	3	4	5	6	7	8
Performance Index P	1.000								
1Educational Level	-.336**	1.000							
2Occupational Background Heterogeneity	.016	.030	1.000						
3Interlocking Directorates	.041	-.022	.122	1.000					
4Political Resources	-.023	-.263**	-.189*	.007	1.000				
5Leadership Structure	-.012	.132	-.156	-.065	.079	1.000			
6Company Size	-.118	.220**	-.298**	-.087	.237**	.171*	1.000		
7Assets	-.637**	.375**	-.191*	-.113	.229**	.006	.470**	1.000	
Liabilities Ratio									
8Board Size	-.321**	-.135	-.087	.360**	.146	-.086	.044	-.136	1.000

Note: **. The correlation is significant at a confidence interval (dual) of 0.01; *. The correlation is significant at a confidence interval (dual) of 0.05

Hierarchical Regression (Least-squares Method) is used to test the hypothesis. Prior to regression, all interaction terms are centrally processed. The regression results are shown in Table 4. Table 4 shows that the

VIF values of Model 1 to Model 3 are all less than 4, much lower than the threshold of $VIF = 10$, indicating that the models all pass the multiple collinearity test. In addition, with the gradual increase of the effect, the adjustment of Model 3 reached 56.8%, indicating that the model constructed in this paper has a good explanation.

Table 4 Multiple Regression Results

Variable	Model 1	Model 2	Model 3
Constant Term	.266 (0.734)	1.013** (2.195)	1.470*** (3.226)
Educational Level		-.165** (-2.101)	-.323*** (-3.774)
Occupational Background Heterogeneity		-.289 (-1.205)	-.533** (-2.190)
Interlocking Directorates		.183 (1.494)	.224 (1.370)
Political Resources		.184 (1.042)	.592*** (2.820)
Leadership Structure		.019 (0.213)	-.306** (-2.114)
Educational Level × Leadership Structure			-1.309*** (-3.151)
Occupational Heterogeneity × Leadership Structure			-.242 (-.242)
Interlocking Directorates × Leadership Structure			.571 (.541)
Political Resources × Leadership Structure			4.710*** (4.116)
Company Size	.079*** (2.944)	.068** (2.441)	.057** (2.161)
Assets Liabilities Ratio	-1.838*** (-10.140)	-1.681*** (-8.391)	-1.627*** (-8.421)
Board Size	-.062*** (-3.541)	-.085*** (-4.420)	-.073*** (-3.908)
F Value	45.376***	19.435***	16.664***
Adjusted R ²	0.482	0.508	0.568
VIF value	<2	<2	<4
Observations	144	144	144

Note: *, **, and *** represent significance levels of 10%, 5%, and 1% respectively; numbers in brackets are t value.

Model 1 examines the effect of control variables on dependent variables without adding independent variables. The regression results showed that all three control variables were related to the dependent variable at a significant level of 1% with an F value of 45.376, indicating that the selection of the control variables in this paper is valid. Among them, there is a significant positive correlation between company size and business performance, mainly because large companies have economies of scale and are easier to obtain external funds to improve their performance). Assets liabilities ratio is significantly and negatively related to business performance. The more debt, the less business performance improvement. This finding is in line with the conclusion of the research by Ma and Jin^[18]. The board size and operating performance are significant. The main reason for this result may be that the large-scale board make the decision-making relatively slow and their response to market information lags behind. Therefore, the size of the board in China's listed tourism companies may not be as large as possible.

Model 2 adds four board capital variable and the moderating variable on the base of model 1. From Model 2, we can see that among the board capital of China's listed tourism companies, only educational level of the

board and operating performance have passed a significant correlation test, with a significance level of 5% and a regression coefficient of -0.323, that is, the higher the average educational level of the board, the worse the operating performance will be, rejected the original hypothesis H1, the reason may be: on the one hand, there are a large number of independent directors of college teachers, they generally have a doctor's degree, which to some extent raising the overall education level; on the other hand, managers with higher education tend to adopt complex management methods. Therefore, the higher the education level, the greater the possibility that the team will have conflict. However, the heterogeneity of the professional background of the board of directors, the interlocking directorates and the political resources owned by the board of directors all failed to pass the test of significance. Chen and Fan^[19] give reasons: Currently, the board of Chinese companies tries to provide special information and knowledge by introducing directors of different backgrounds into the board, such as former government officials, university professors, industry association members, etc.

Model 3 adds four interaction terms between board capital and leadership structure. As can be seen from Model 3, the interaction term of the educational level and the leadership structure of the board and the interaction term between the political resources and the leadership structure have passed the test of significance. Among them, the educational level and the leadership structure of the interaction term is negative, indicating that the leadership structure negatively moderate the correlation between educational level and firm performance, rejected the hypothesis H5; political resources and leadership structure interaction term is 4.710, indicating that the leadership structure significantly enhances the influence of the political resources on the firm performance, and in the two-in-one leadership structure, the political resources owned by the board promote the performance of enterprises, H8 has been strongly supported. The interaction term between occupational heterogeneity and leadership structure, as well as the interaction term between interlocking directorates and leadership structure neither passed the test of significance level, indicating that the leadership structure has no moderating effect on both the correlation between occupational heterogeneity and firm performance, and the correlation between the interlocking directorates and firm performance.

In China, due to the special history of social development and the political and economic system, the government directly or indirectly controls a considerable number of enterprises through the administrative departments, which plays a decisive role in the operation and management of enterprises. At the same time, the comprehensiveness of tourism products is extremely strong, including six elements of food, lodging, transportation, travelling, shopping and recreation. These elements are closely linked to each other and form a complete tourism industry value chain. Tourism enterprises with political resources often have information superiority than those with no political resources, it is more accurate to grasp the macroscopic information and then improve the business performance of enterprises. The board of directors with two positions (CEO and chairman) will not only help reduce the internal management costs and decision costs, ensure the uniformity and timeliness of orders, but also help to establish a rapid response mechanism between the enterprise and the external environment, taking full advantages of market opportunities and avoid threats. Therefore, the leadership structure of two positions is more conducive to the positive impact of the political resources of the board in China's tourism listed companies on firm performance.

5. CONCLUSIONS AND RECOMMENDATIONS

This paper takes 24 listed companies in Shanghai and Shenzhen Stock Exchanges during the period of 2011-2016 as the research object and takes the leadership structure of the board as the moderating variable to make an empirical study on the impact of board capital (human capital and social capital) on firm performance. The results show that the current human capital of China's listed tourism companies does not play a positive role in promoting firm performance, while the social capital of the board, to some extent, has a slight positive effect

on firm performance. The leadership structure with two positions in China's listed tourism companies can better adapt to the uncertainty of the environment.

At present, China is in the stage of economic transition. Listed companies are facing intense market competition pressure, and the thirst for resources is more and more fierce. The human capital and social capital owned by the decision makers have become one of the resource obtain channels^[13]. Although the company hopes to capitalize on its board capital as much as possible to enhance its growth potential, it does not mean that the board of directors of some type will have as many capital as possible. Instead, it should pay attention to the portfolio effect because the combination of different human capital and social capital attributes can provide the company both the advantages and disadvantages^[6], a board with good combination of directors can become an important intangible asset for the company.

The shortcomings of this paper lies in the following aspects: First, only a few indicators are selected for the measurement of capital variables, which are not meticulous and perfect; second, whether more moderating variables can be used in the research are not considered in detail.

ACKNOWLEDGEMENT

This research was supported by Key Research Project of Philosophy and Social Sciences of Ministry of Education under Grant 17JZD020; Higher Education Reform Project of South China University of Technology under Grant gj2015003; Key Project of Guangdong Provincial Degree and Graduate Education Reform in 2017 under Grant 2017JGXM-ZD03

REFERENCES

- [1] Qin Rong, Zuo Li, Zhang Tongjian.(2014).Research on the Promotion Effect of the Third Industry Corporate Governance on Market Competence - Based on the Data Test of Listed Real Estate Companies Based on Tourism, Finance and Insurance. *Business Era*,(14):90-92. (in Chinese)
- [2] Lu Tong.(2012).The Trends and Inspirations of International Corporate Governance Reform after the Financial Crisis[. *International Economic Review*,(4):108-120. (in Chinese)
- [3] Zhou Jian, Li Xiaoqing.(2012)._An Empirical Study of the Impact of Board Cognitive Heterogeneity on Firm Innovation Strategy. *Management science*, (6):1-12.(in Chinese)
- [4] Feng Huiqun.(2014). Research on the Effect of Board of Directors' Capital on Corporate Governance. Ms D Thesis. Tianjin: Nankai University. (in Chinese)
- [5] Hillman A J, Dalziel T.(2003). Boards of Directors and Firm Performance: Integrating Agency and Resource Dependence Perspectives. *Academy of Management Review*,28(3):383-396.
- [6] Kor Y Y, Sundaramurthy C.(2009). Experience-based human capital and social capital of outside directors. *Journal of Management*, 35(4): 981-1007.
- [7] Kim Y, Cennella A J.(2008).Toward a social capital theory of director selection. *Corporate Governance: An International Review*,16 (4) :282-293.
- [8] Shamsul Abdullah (2014) .The causes of gender diversity in Malaysian large firms[J].*Journal of Management and Governance* ,4(18):1137-1159.
- [9] David M. Reeb ,Wanli Zhao. (2013) .Director capital and corporate disclosure quality[J]. *Journal of Accounting and Public Policy* ,4(32):191-212.
- [10] Datta D K,Rajagopalan N. (1998) .Industry structure and CEO characteristics: An empirical study of succession events[J].*Strategic Management Journal*,19(9) : 833-852.
- [11] Chen Yue, Zhu Xiaoyu, Liu Zeyuan.(2015). Empirical Study on the Relationship between Board of Directors Capital and Corporate Performance. *Journal of Dalian University of Technology (Social Sciences Edition)*, (04): 23-29.(in

Chinese)

- [12] Haynes K T., Hillman A (2010) . The Effect of Board Capital and CEO Power on Strategic Change[J]. Strategic Management Journal,31(11): 1145-1163.
- [13] Feng Ruixue, Li Wei, Sun Changjiang.(2013). Will the capital of the board of directors increase corporate value?——Based on the empirical data of listed companies in China's Shanghai stock market. Friends of Accounting, (34):65-71(in Chinese)
- [14] Koka B R,Prescott J E (2002) .Strategic alliances as social capital: a multidimensional view[J]. Strategic Management Journal, (9): 795-817.
- [15] Zhu Xiumei, Li Mingfang. (2011). The dynamic influence of entrepreneurial network characteristics on resource acquisition[J]. Management World, (6):105-115.(in Chinese)
- [16] Agrawal A ,Knoeber C R (2001) . Do Some Outside Directors Play a Political Role?. Journal of Law and Economics, 1 (44) : 179-99.
- [17] Yan Zichun, Xue Youzhi.(2015). The impact of board social capital and corporate leadership structure on the R&D investment of firms. Chinese Journal of Management, (04):509-516.(in Chinese)
- [18] Ma Jing, Qi Tianting.(2017). Research on Capital Structure and Operational Performance of Tourism Listed Companies[J]. Finance and Accounting Communications, (11):39-42.(in Chinese)
- [19] Chen Huaichao, Fan Jianhong.(2015). Research on the Influence of Human Capital and Internal Social Capital on the Creative Decision-making of Board of Directors: Based on the Perspective of Knowledge Creation. China Human Resources Development,(07):53-59.(in Chinese)
- [20] Tina Yang, Shan Zhao. (2014) .CEO duality and firm performance: Evidence from an exogenous shock to the competitive environment[J].Journal of Banking & Finance, (49):534–552

The Business Talents Cultivation of Cross-border E-commerce under “the Belt and Road Initiative”

Ai Xu^{1}, Shi Tang², Yijia Gao³, Zongqing Zhou⁴*

^{1,2}International Business Faculty, Beijing Normal University Zhuhai, China

³Business School, Chung Yuan Christian University, China

⁴College of Hospitality and Tourism Management, Niagara University, USA

Abstract: Along with “the Belt and Road Initiative”, China’s cross-border e-commerce has been developed rapidly, which will certainly bring new challenges for talents cultivation in the Higher Education. The paper analyzes the new requirements for business talents under cross-border e-commerce. And then a cultivation system was proposed by examining the five major programs in our university. The cultivation objectives, the curriculum and the practical module are elaborated. Finally, some suggestions on the implementation of this cultivation system are put forward.

Keywords: talent cultivation, cross-border e-commerce, business talents, curriculum module

1. INTRODUCTION

Nowadays, cross-border e-commerce has become the focus of all walks of life, and has shown an explosive growth momentum. In order to ensure the steady growth of foreign trade and provide advice and guarantee for the development of cross-border trade of e-commerce, the Office of State Council announced "the implementation of policies to support cross-border e-commerce retail outlet" in August, 2013. Two years later, the Political Bureau of CPC Central Committee approved the overall scheme of Guangdong Pilot Free Trade Zone, Tianjin Free Trade Zone and Fujian Free Trade Zone to further deepen the reform of Shanghai Free Trade Testing Zone. These policies have promoted the development of cross-border e-commerce greatly. Along with “the Belt and Road Initiative”, China’s cross-border e-commerce entered a new era. The total transaction of cross-border e-commerce is estimated to be 8 trillion RMB by the end of 2018 and will reach 12 trillion RMB in 2020. The speed of growth will remain at an average of 20% for the further few years. The cross-border e-commerce will become an “Online Silk Road” that cannot be ignored. This not only has a great meaning to the nation’s economy and international trade, but also addresses a new challenge of cultivation of related talents.

The management and transactions of cross-border e-commerce involve knowledge and skills in many fields, such as economics, management, accounting, e-commerce and so on. However, this industry is still short of the talent and personnel who have the knowledge and skill^[1]. This is the problem that universities, especially those with related majors, have to solve. Therefore, universities need to design new curriculum modules based on the perspective of cross-border ecommerce. This will also strengthen the construction of related majors in universities and promote the transformation and development of cross-border e-commerce for both universities and enterprises under the background of “internet plus”.

2. NEW REQUIREMENTS FOR BUSINESS TALENTS UNDER CROSS-BORDER E-COMMERCE

The development of cross-border e-commerce rapidly transforms the traditional trade modes^[2]. These transformations not only put new requests to the traditional international trade professionals, but also challenge other personnel within the trading activities. The business talents need to have new knowledge and competency in the new cross-border e-commerce environment.

* Corresponding author. Email: gdxuai@163.com (Ai Xu)

2.1 Capabilities to execute e-marketing function

Traditional enterprises usually build foreign marketing channels through international exhibitions or agents^[3]. This model is unsuitable for miniaturized businesses that directly serve the consumers due to the high cost. Instead, using internet to launch a low cost marketing model to overseas consumers is more feasible. Compared with the traditional way, the internet model requires cross-border e-commerce practitioners to master the basic knowledge of e-marketing, understand the consumer behavior through analysis of the overseas markets and refined products selling points and promotion according to consumer demand. At the same time, they are also required to effectively use target countries' search engines to carry out promotional activities, such as SEM, SNS and foreign medium PR, on the internet.

2.2 Capabilities to operate cross-border e-commerce platforms

The use of cross-border platform, such as Alibaba and DHgate, to carry out e-commerce activities is a convenient way for small and medium-sized enterprises to expand overseas market. Therefore, cross-border e-commerce requires personnel to operate business platforms, such as export portal of Alibaba and Aliexpress, to implement the practices of import and export, which include inquiry management and transformation, customer relationship and product management. The business personnel also need to develop foreign markets by using global search engines and e-marketing strategies. On the other hand, personnel in other business areas, such as finance, accounting and business administration, also need to master the financial management, auditing, accounting and process management on the platform to deal with cross-border e-commerce related businesses.

2.3 Capabilities to run cross-border e-commerce websites

In addition to using the existing cross-border e-commerce platform, some competent enterprises also need to build their own e-commerce websites for overseas market and providing services for overseas consumers^[4]. This requires the relevant personnel to have the experience of coordinating supply chain, logistics, website promotion, order processing and after sale service. They need to improve brand awareness in the network environment through effective website operation and management.

2.4 Capabilities to manage the logistics and supply chain

Logistics play an important role in the development of cross-border e-commerce. Unlike ordinary international trade, the logistics of cross-border e-commerce is mostly done by direct mail, express, aviation special line, various entrainment and shipping bulk cabinets, which is different from liner shipping or ship booking. There have been great changes in commodity inspection, transportation insurance, which are closely related to logistics. All these require the business personnel in cross-border e-commerce environment to handle the issues of internal and external logistics of bulk commodities and effectively coordinate all aspects of supply chain under the situation of small batch and multi batches.

3. CULTIVATION SYSTEM OF BUSINESS TALENTS FOR CROSS-BORDER E-COMMERCE

A cultivation system is proposed here by examining the five major programs in our university in order to make our students adapt the requirements of the cross-border e-commerce environments. We set up cultivation objectives according to the analysis of the requirements for business talents under cross-border e-commerce, and then design curriculum module and practice module respectively. Students could select corresponding courses in the module after completing the study of core courses of their own major so as to obtain and improve their capabilities to cross-border e-commerce. (Figure1)

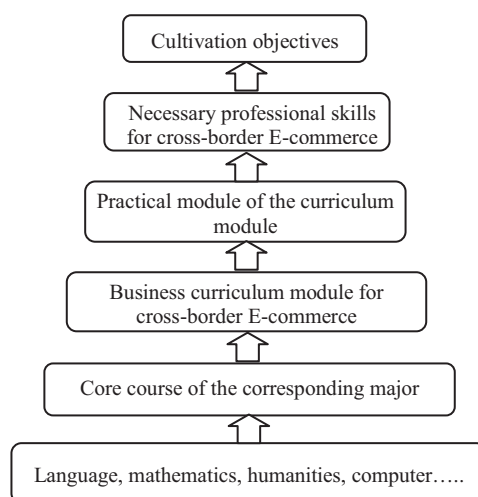


Figure 1. Structure of Business Talents Cultivation for Cross-border e-commerce

3.1 Cultivation objectives

The cultivation objectives for cross-border e-commerce business talents, must meet the needs of economic and social development of cross-border business environment as well as enterprises to business professionals^[5]. In general, these talents should have a certain structure of knowledge, competency and quality. They should include following elements:

3.1.1 Knowledge structure

- Master English and knowledge of computer application.
- Master the completed knowledge of the principal of modern management and economics.
- Master basic theories, knowledge and skills of their own major.
- Master the basic theories and knowledge of cross-border e-commerce, such as trade, business management, international settlement, accounting, financial management and logistics management.
- Have certain knowledge of business law.
- Master the qualitative and quantitative analysis methods required by the major.
- Be familiar with the principles, policies, regulations and international custom and rules.
- Understand the theoretical frontiers and trends of this subject.

3.1.2 Competency structure

- Master the basic skills of practical operation in this field.
- Master computer skills and be able to use information equipment, network tools and business management application.
- Master the skills to deal with related business in the field of cross-border e-commerce.
- Have a strong ability to listen, speak, write, translate and read of a foreign language.
- Develop their own analytical, innovative and decision-making abilities through lectures, class discussions, writing papers, participating in seminars and practical activities.
- Have strong abilities of oral and written language expression, communication, analysis and problem solving.
- Understand the basic methods of information retrieval and have the abilities for preliminary scientific research and practical work.

3.1.3 Quality structure

- Have good moral and ethical quality.

- Have Good physical and mental health.
- Have enterprising spirit, responsibility and dedication.
- Have a positive attitude and ability to face difficulties.
- Have a certain sense of innovation and creative potential.
- Have teamwork spirit, tolerance and affinity.

3.2 A business curriculum module for cross-border E-commerce

According to the current requirements for business talents in cross-border electric business development, a curriculum module was proposed by examining the five major programs in our university, i.e. e-commerce, international economics and trade, finance, accounting and business administration^[6]. Individual majors need to create knowledge for cross-border e-commerce development through related elective courses in addition to the original teaching plans. The module should include such courses as Introduction of E-Commerce, E-Marketing, Online Payment, Mobile Commerce, ERP Simulation, Supply Chain Practices, International Logistics, Practice of Financial and Business Integration, International Settlement, Theory and Practice of International Trade, International Business Communication, Import and Export Trade, Simulation of International Cargo Transportation and Operation, Insurance, Business Negotiation, Online Retail, Customer Relationship Management and Comprehensive Training of Cross-Border E-Commerce.

Each major in our university needs to build corresponding system according to the requirements of cross-border e-commerce. Students who study in these majors should have the ability to work in cross-border e-commerce environment through proper practice and training.

3.3 The practical module of the curriculum module

Practices are critical in the curriculum module, because business talents need to have the ability to operate on the e-business platform in the environment of cross-border e-commerce. These practices could be divided into three types, i.e. single practice, comprehensive practice of the course and the comprehensive training.

3.3.1 Single practice in one course

Single practice mainly refers to the practical projects within the courses. It can be conducted by such manners as case studies, enterprise field visits, industry experts' presentation, software demonstration, and group design. Teachers need to clearly identify the type and length of projects. For operational practice projects, in order to receive better results, it is recommended that teachers make their own practical guidance books that students can follow. At the same time, students can also complete these practice items on their own, if the open laboratory is possible. This would help to design reasonable and feasible single practices as well as strengthen the supervision, management and assessment of the courses in the module^[7]. Therefore, it can reduce the randomness of implementation process and ensure the effectiveness of practice.

3.3.2 Comprehensive practice of one course

The comprehensive practice of the course is aim to make the entire course a workshop of the practices. Teachers should write detailed syllabus and guidance books to fulfill students' practical tasks and achieve the goal of practice. The courses of ERP Simulation, Supply Chain Practice, Practice of Financial Business Integration, Import and Export Trade Simulation are this kind of practices.

3.3.3 Comprehensive Training for the module

In order to enable students to obtain practical operation ability under the cross-border e-commerce platform, a comprehensive training is necessary. Faculties also need to design the knowledge framework and training content on the base of the major domestic cross-border e-commerce platform. The knowledge framework should include the product selections, commodity presentation, commodity promotion, online payment and settlement, distribution and logistics, customer service and communication, cross-border business rules and intellectual

property rights, job and career development, innovation and entrepreneurship contest^[8].

It is necessary to use Aliexpress, Amazon, DHgate, Wish and other major cross-border electronic business platforms to design learning modules. So that students can practice on registration and certification, platform operations, the use of marketing tools, order processing and data analysis. They also can understand operational method and basic business process of these platforms through the modules.

4. SUGGESTIONS ON THE IMPLEMENTATION OF THE SYSTEM PROPOSED

In order to better implement the cultivation system of business talents for cross-border e-commerce, the following suggestions could be very helpful.

4.1 Integrating the cultivation program with the cross-border e-commerce curriculum module

It is planned to embody this module in the teaching plans of five majors of our university when the talent cultivation program is revised in 2018. The curriculum guidance and syllabuses need to be improved by compiling instruction documents for practical teaching. Provide instructions of course selection for students according to their majors. Encourage students to select courses according to the module.

4.2 Strengthening teacher training to improve teachers' comprehensive quality and ability

Teachers will play variety of roles, such as reviewers, guides, analytical commentators and business advisers, in the teaching of cross-border e-commerce modules^[9]. That addresses high requirements for teachers to meet. Firstly, teachers are required to have excellent knowledge structure and good knowledge of economic management, especially the knowledge in the aspect of cross border e-commerce. Secondly, teachers should have good organizational and coordinating ability, class control ability and strain capacity. They need to master various teaching tools and software and have experience to answer the questions that students ask and solve the problems encountered. Thirdly, teachers are required to have good personality. They need to encourage students to be creative and enterprising. They should be patient and loving in order to educate students to be good people. Finally, the teachers who are engaged in the comprehensive training course must have spirit of dedication, because there is often no concept of time in practical training courses. Workshops often cannot be finished on time as students have different personal characteristics and groups have different speed of finishing tasks. Therefore, teachers must have responsibility to be able to undertake such courses and ensure teaching quality.

We should fully use the resources of enterprises along with strengthening the construction of teaching staff. We can invite business professionals to make reports for students, so that students can better understand the reality of enterprises and deepen their understanding of practical courses. Enterprise personnel can also be the instructor of the courses and guide the students' complete training practice.

4.3 Improving the teaching method and introducing modern information technologies

Teachers should be encouraged to actively adopt advanced teaching methods and means. Students can use the resources from the uniformed comprehensive training platform of cross-border e-commerce in the university for autonomous learning and training.

4.4 Increase investment in teaching funds and strengthen the construction of practical training base and practical teaching base outside school.

We need to quickly build our cross border e-commerce training platform. We should strive for further financial support from the university in addition to the existing laboratory resources to build a simulation training base, so that students can simulate the main business processes of cross-border e-commerce on campus. At the same time, we must continuously build off-campus practice teaching base to strive for more internship

opportunities for students and enhance their competitiveness.

4.5 Innovating the teaching management and introducing the incentive mechanism

The teaching management needs to be further developed to make the module work as soon as possible. The academic affair office sets the standard requirements while the faculty is responsible for monitoring and assessment. We also notice that teachers need to spend extra time and effort to add practical contents into courses and develop corresponding experimental projects. Therefore, the incentive mechanism must be introduced. Teachers with outstanding effect should be rewarded. We need to promote their teaching experience for demonstration and inspiration.

5. CONCLUSION

In short, along with the rapid development of cross-border e-commerce, having the related professional skills and application ability is critical to business talent cultivation. The design of the curriculum system is a complex work that needs to integrate with professional practice and student characteristics. We have already begun to deploy the curriculum modules in our university and keep improving it during the practice.

ACKNOWLEDGEMENT

This research was supported by Seentao Ltd under the Grant 201701063019 due to a cooperative program from the Ministry of Education of the People's Republic of China. The research was also supported by our university under Grant CXQX2017022.

REFERENCES

- [1] Giuffrida M, Mangiaracina R, Perego A, Tumino A. (2017). Cross- border B2C e-commerce to Greater China and the role of logistics: a literature review. *International Journal of Physical Distribution & Logistics Management*, 9(47):772-795
- [2] Wang F. (2015). Study on the Teaching Reform of International Trade of Cross-border E-commerce. *Journal of Hubei Correspondence University*, (21):100-101 (in Chinese)
- [3] Yu K, Gong R. (2016). Analysis on Cultivation and Construction of Talent Team in Library and Information Science Specialty. *Open Journal of Social Sciences*, (4):161-166
- [4] Gomez-Herrera E, Martens B, Turlea G. (2014). The drivers and impediments for cross-border e-commerce in the EU. *Information Economics & Policy*, (28):83-96.
- [5] Yuan J, Liang J. (2018). Cultivation for the Cross-border E-commerce Talents for International Trade Majors. *Information Recording Materials*, (2):166-167 (in Chinese)
- [6] Wang Y, Zhang S, Fu J. (2017). A Study on the Optimization Design of the Training Program of Curriculum System for Cross-border E-commerce Talents in Higher Vocational Colleges Based on the CIPP Model. *Revista de la Facultad de Ingenieria*, 8(32):589-596
- [7] Ping P. (2017). Research on the Cultivation Model of Cross-border E-commerce under the "Internet Plus". *Market Modernization*, (23):29-31 (in Chinese)
- [8] Su Q. (2016). Research on the Cultivation of International Trade Professionals in Independent Colleges under the Background of "Cross-border E-commerce". *Economic Research Guide*, (17):84-87 (in Chinese)
- [9] Fan W. (2015). The New Models of Cross-border E-commerce and International Trade. *Foreign Economic Relations and Trade*, (1):4-7 (in Chinese)
- [10] White A. (2017). Driving Business Retention, Expansion and Attraction Program to A Higher Level Through Talent Pipeline Management. *Economic Development Journal*, 4(16):28-36

Study on Social Network for College Students' Job Hunting

LiLi¹, Zhenyan Hu^{1}, HuimingXiao¹, MeilinChen¹, YitangZeng¹*

¹Wuhan Donghu University, Wuhan, 430212, China

Abstract: SNSs recruitment has caused the change of the traditional recruitment mode due to its advantages such as broad audience, quick information transmission, low recruitment cost and good interpersonal interaction. Through questionnaire survey, the paper implements on a study on the situation that the college students use social networks to find jobs. It is found that social network platforms are popular with college students, but it is not widely used in social job hunting platforms. What is more, the job hunting effect is not obvious. Meanwhile, the user information is prone to leakage, and the job hunting information provided is in poor quality. The paper proposes that the social network job hunting platforms should take measures to perfect the service content, strengthen the technical level and improve the information security for the college students. Additionally, the college students should also make reasonable choices and correctly utilize the social recruitment platform to protect the personal information security and improve the job hunting efficiency.

Keywords: social networking sites, recruitment, college Students, job hunting

1. INTRODUCTION

The advent and widespread use of mobile Internet and Web 2.0 technologies have promoted the development of social networking sites (SNSs)^[1]. To date, a majority of Internet users no longer passively receive the publishers' information unidirectionally. Instead, they can communicate bidirectionally with publishers and other users. The development of social networks has provided a new platform for talent recruitment and job hunting, such as Facebook, LinkedIn, Twitter, Weibo^[2], Renren, Dajie, WeChat and so on^[3], making HR establish contacts with a large number of potential job applicants easily.

According to iResearch data, the number of online employers of recruitment industry has reached 4.282 million, and the number of job seekers has amounted to 144.495 million in 2016. In the first half of 2017, the amount of China's online recruitment market was 2.71 billion yuan^[4]. According to a survey released by the Global Employee Index Survey of Kelly Services, the success rate of job-seekers in China was ranked as follows^[5]:

1. Job website (38%)
2. "Recommended" (25%)
3. "Headhunting" (18%)
4. Employers take the initiative to seek talents (12%)
5. Offline meetings (4%)
6. Paper media advertisements (3%)
7. Get jobs directly from the social networks (1%).

Social networks have deepened the academic penetration to traditional job hunting and recruitment methods, especially the No. 2 "Recommended" (25%), No. 3 "Headhunting" (18%) and No. 4 Employers take the initiative to seek talents (12%), reaching 55% in total^[5].

iResearch has classified China's online recruitment markets into four categories: one is the comprehensive recruitment websites represented by 51job and Zhaopin. The second is the vertical recruitment websites represented by Lagou and BOSS, which are based on some specific industry to provide job hunting and recruitment services for employers and job-seekers. The third is occupational social recruitment website

* Corresponding author. Email: 286475904@qq.com(Zhenyan Hu)

represented by Dajie, which learns from the success model of the American business social network LinkedIn and hopes to occupy a position in the form of business SNS in the online recruitment market. Finally, the fourth is classified information websites represented by 58.com and Ganji.com, which provides a series of life services such as renting, job seeking, group buying and traveling around^[4].

Faced with a large number of recruitment sites and recruitment channels, how should college students choose among them, and what are the prominent problems in the specific applications of social networks for job seeking? What are the underlying reasons? In view of this, the paper conducts a questionnaire survey on this year's graduates and graduates within two years through SO JUMP to analyze the application status quo and existing problems of social networks in college students' job hunting, so as to provide reference for the smooth and reasonable employment of the college students under the severe employment situation.

2. THE CURRENT SITUATION OF SOCIAL NETWORK IN COLLEGE STUDENTS' JOB HUNTING

A total of 150 questionnaires were distributed in the survey, and 128 valid questionnaires were collected, the effective rate reaching 85.3%, of which 50 were men, accounting for 39.06%. The 78 women accounted for 60.94%. There were 96 with bachelor degree or above, making up 75%, 32 with college degree, accounting for 25%, and 59 students, accounting for 46.09%, 26 graduates, accounting for 20.31%, 31 graduates within two years accounting for 24.22%, 12 graduates more than two years, accounting for 9.38%.

According to the statistics of the results of the questionnaire, the application status quo of social network for college students can be summarized as follows:

2.1 Most college students are willing to use social networks for job search

Amongst the 128 respondents surveyed, 73 said they were willing to choose social networks for job search, accounting for 57.03% of all respondents, while only 10.16% were unwilling to use social job search platforms and 42 would choose social networks for appropriate mode of job hunting, accounting for 32.81%. It indicates that more than half of college students will choose social networks for job search.

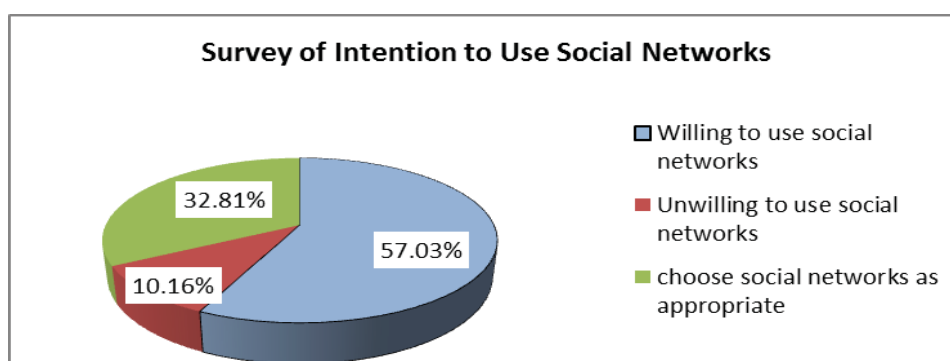


Fig 1. Survey of Intention to Use Social Networks

2.2 The diversity of college students' participation in social network for job hunting

About the methods that students interact with these Social network sites, 34.38% of the students chose to download the app; 33.59% of the students chose to login the website, 28.13% of the students preferred the WeChat concern, and 3.91% of the students opted for Weibo. It can be seen from the data that college students participate in social networks in a variety of forms. Most recruitment sites have set up their own Wechat Official Account, Weibo and APP to increase customer viscosity, so that customers can log on both the PC client and mobile client. For example, Zhaopin, 51job and chinahren and other comprehensive recruitment websites have

set up their Wechat Official Account and Weibo; Dajie, Lagou, Liepin and Neitui have established their own APP. Mobile phone users mainly participate in the interaction of social platforms through the download APP.

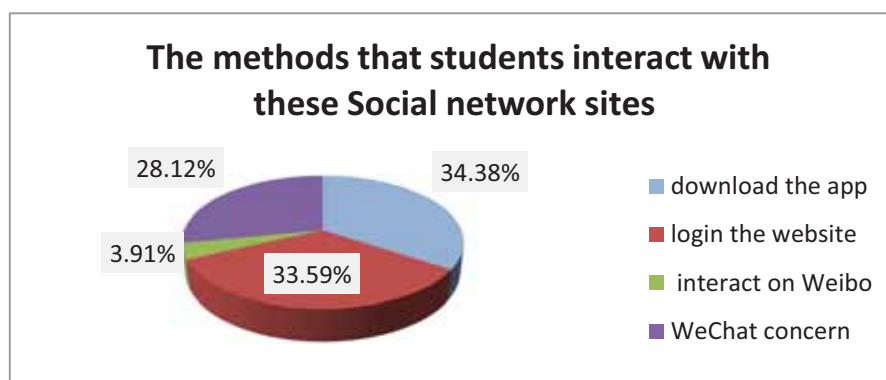


Figure 2. The methods that students interact with these Social network sites

2.3 The reasons for college students to consider using social networks for job hunting

According to iResearch statistics, in 2017, in China's online job seekers, the population with bachelor degree or higher education accounted for 57.4%. In recent years, the educational background structure of job seekers in the online recruitment industry has been basically stabilized. The proportion of highly-educated people has basically sustained at around 60%, while the proportion of people with low education levels has been kept at about 40%^[6].

The paper also investigates the reasons why college students choose social networks for job hunting through online questionnaires. The survey results show that 74.22% of respondents deem it to be convenient to use social networks for job hunting; 64.13% of respondents consider the interaction is strong and the information exchange is rapid; 57.03% of respondents think social job search can provide more job opportunities; 32.57% of respondents choose social network for job hunting to get more authentic information because social network sites are generally the career circles based on acquaintances or interests to establish and it can provide more reliable job seeking information. We can understand the dynamic information of the workplace through interpersonal interaction, for instance, you can understand the salary level of same position in different enterprises, and we can get credible information from the interaction with real working people.

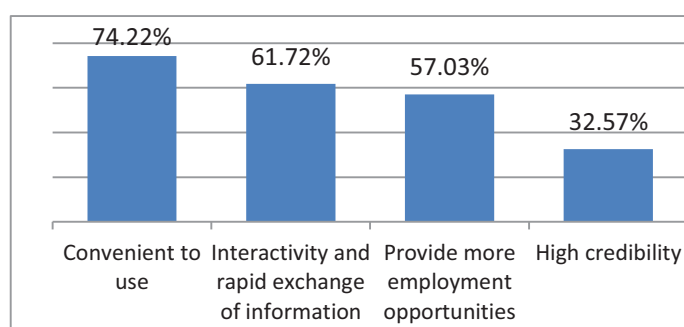


Figure 3. Reasons why college students are willing to choose social networks

3. THE PROBLEM OF THE APPLICATION OF SOCIAL NETWORK FOR CLLEGE STUDENT

3.1 Social networks are not widely used in college students' job hunting

The survey also delves into college students' favorite job search channels. Of the 128 people surveyed,

those who like comprehensive recruitment websites account for 39.84%. The proportion of those who like job fair rank second is 33.59%. Those who like to get recruitment information through newspapers, paper media advertisement makes up 7.03%. People who like social networks for job hunting account for 15.41%, and the proportion of other platforms is 4.13%. It justifies that the public favorite online recruitment channel acts as a comprehensive recruitment website.

In addition, throughout the survey of college students, many students remarked that they do not know how to fill in the questionnaire, because they understand so little about job seeking on social network. Many students have never heard of many social network job seeking platforms such as Wealink, indicating that social network job hunting is not widely used by college students.

Table 1. College students' favorite job search channels

Job-hunting channel	Examples	proportion
comprehensive recruitment website	51job , Zhaopin, chinahr	39.84%
job fair	Campus Recruiting	33.59%
social networks for job hunting	Dajie, wealink ,weibo, weixin,renren, Facebook, LinkedIn, Twitter	15.41%
Paper media advertisements	newspapers, TV and radio station	7.03%
Other channels	Lagou , BOSS	4.13%

The survey results are consistent with the survey results of usage of the entire network job search platforms in iResearch "2017 China Online Recruitment Industry Half-Year Report". IResearch statistics show that on a variety of online job search platforms, the turnover of the comprehensive recruitment websites was ranked the first, superior to social network recruitment platforms. The total turnover of 51job and Zhaopin ranked Top Two in the comprehensive recruitment websites account for 62.5% of the entire online job search industry, other job search platform sites only accounting for 37.5%, indicating that the number of users of social network platforms is far less than comprehensive recruitment platforms.^[6]

3.2 The effect of social network job hunting is not significant

Social network job search is a kind of reform and innovation of traditional recruitment methods. Compared with other job search methods, social network job search has the advantages of better interpersonal interaction and faster information feedback^[7]. However, respondents' feedback reveals that the actual job search effect is not significant. According to the network survey, only 13.44% of college students claimed they found the favorite job by means of social networking job search platform, and only 30.63% of university students determined they will continue to use the social network job search platform to find work, suggesting that most college students are dissatisfied with the effectiveness of social network job search. Social network job search effect is not significantly better than other ways. Despite that social network job search is still mainly used as a channel for information acquisition, it cannot completely replace the traditional job search methods.

4. THE REASONS FOR THE POOR EFFECT OF SOCIAL NETWORK SITES IN COLLEGE STUDENTS JOB HUNTING

4.1 Lack of social network job hunting sites for college students

Social networking sites are also divided into entertainment social networking sites and professional social networking sites. Entertainment social networking sites, such as Renren, Pengyou, QQ space, and Kaixin001 are commonly used in China^[8], the recruitment functions of which are weak. Professional social networking sites such as Wealink and Liepin are mainly for workplace users, which rarely set up job-specific module for college students. Taking Renren that covers most social users of college students as example, Renren sets up the

Xiaozhao Renren module, but recruitment is not its main business. Therefore, the recruitment information and quality of service provided are not high. In recent years, Renren's service area has become overly complicated, such as Renren shopping, Jiexi for wedding celebration service and so on. Renren has deviated from the original intention of serving student groups.

4.2 Lower level of information management and serious risks of information security

In the satisfaction survey of the use of social networking sites, 82.03% of interviewees thought the number of false information has been increasing; 67% of respondents said personal information disclosure has caused various harassment; 51.56% of respondents believed there are more invalid social contacts; 2.34% of the respondents thought spam is flooded; 39.06% of the respondents supposed the information is invalid; 30.47% of the respondents complained that the classification of job information is not clear, while 25% of users commented the social network still cannot recommend personalized information. From the survey results, the reasons for the low satisfaction of social networking sites are concentrated in the information. With the development of the Internet and the way and characteristics of viral information dissemination, social networking sites have become a platform for information exchange favoured by enterprises and job seekers^[9]. However, information processing technology is not mature enough and the internal management procedures of social network platforms are not standardized, while the relevant government departments' supervision is not in place. Furthermore, social security technology system itself needs to be improved. Among all the problems, the major ones are the timely processing and protection of user information, resulting in unsatisfactory use of social networking sites^[10].

4.3 Insufficient management information ability of college students

Insufficient management information ability of college students is manifested in three aspects: First of all, the ability to collect information is not strong. Many college students do not know how to capitalize on various social networks to collect recruitment information and post job search information. In the survey, many college students have not heard of some well-known job search sites, indicating that the ability of students to take the initiative to obtain information is not strong. Secondly, the ability to identify the information is insufficient. Through interviews with college students, many college students are easily tricked by fake recruitment information and reveal personal information and suffer harassment. Thirdly, the processing ability of information is not strong, including image processing, text editing, interactive communication skills, affecting the personal impression in the mind of employers, especially the abilities in the production of personal resume and in the display of individual specialties are not strong, which seriously impact on the effectiveness of job search.

5. HOW TO IMPROVE THE USING EFFECTIVENESS OF SOCIAL NETWORKS IN COLLEGE STUDENTS ' JOB HUNTING

5.1 Social networking sites should set up job search modules for college students to provide quality of service

Social networking job search platform should subdivide customer groups, invest professional development teams and service teams for the job hunting characteristics of college students, so as to provide job search services suitable for college students or establish job search module for college students. For example, they should establish campus social networking platforms cooperating with colleges and universities. The service content should take into account the exchange of learning, academic exchanges, school-enterprise exchanges,

student internships, job search, starting business services^[11] and so on to increase student user coverage rate.

5.2 Social network service providers should raise awareness of information security and improve security technologies

For starters, social networking sites should pay attention to user safety, and invest R & D efforts to improve safety technology, for example, preventing the site from being attacked by viruses and hackers and reducing vulnerabilities in social networking sites^[12]. Secondly, we should improve the user information security management system and procedures, refine user privacy protection terms, introduce website security access guidelines. At the same time, we should remind users to take care of their personal information on the full position of the website, and provide "bank-level" security measures for the information content involving the privacy and property of users^[13]. Finally, we should play the regulatory role of the network security department, and learn from the advanced experience of European and American countries, such as the United States' protecting the privacy of the network through the industry self-protection model. For example, when people see the well-known "trust" certification mark of third-party certification, which implies that the site has taken some measures to protect personal information, you can rest assured and use it^[14].

5.3 Improve data processing technology and improve the accuracy of information services

Data processing is the process of extracting valuable information from a large amount of raw data and converting the data into information, including data collection, storage, processing, classification, merging, computing, sorting, transformation and so on. It is the inevitable trend of development in the future to collect and process the information of social networking job search platforms with big data technology, and it is also the protection that timely provides users with effective information. In this sense, social networking service providers must pay attention to data processing technology to improve the accuracy of information services.

5.4 Improve the management process, improve user satisfaction

First of all, when selecting a recruiting company or a headhunter, it is necessary for social networking sites to review the recruiting enterprises and classify the recruiting enterprises, so that different groups of users can efficiently collect their own enterprises. Second, we should review the distribution of recruitment information, excluding false information and fraud information, delete outdated information and merge duplicate information to provide users with a cleaner online job search environment and product-oriented evaluation framework for system users/buyers. The common concept of system performance connects the two fields: generally designers intend to produce high performance systems, and likewise users want to buy them.

5.5 Strategies for the effective use of social networks for college students

College students should strengthen career planning, and select the appropriate social job search platform for job search. Only when one is clear about your career planning and goals, and identify the targeted position or the targeted company, you can use the appropriate social networking job search sites to find your favorite job for a better choice^[15]. In general, the efficiency and effect on professional social networking sites is higher than entertainment social networking sites.

College students should strengthen the abilities of information management and network behavior management.

As a first step, college students should develop the ability to gather information and evaluate information, learn to observe the concerns of business recruitment, and create rich content on social media personal home pages to show themselves. In social media recruitment, the employers usually deduce candidates' personality,

interests, specialties and so on based on information presented by job seekers' social media, and conduct background checks on job seekers^[16]. College students should check the content, form and frequency to leave good impression on employers. Secondly, college students should develop their ability to handle information. For example, you can learn some essential computer skills and be able to demonstrate your learning achievements and practical experiences in the form of words, videos, pictures, anime and more. Thirdly, college students should learn the ability of information identification, avoid the loss caused by false information and fraudulent information, and publish healthy content in the dissemination of information and interpersonal interaction.

College students should establish an effective interpersonal circle, and communicate with businessman with the same interest and professional to avoid invalid social contacts. College students should make time management. Many college students use social networking sites for job search, and they should pay attention to screening effective recruitment information and do not get caught in the other entertainment or invalid chat of the social network.

College students should protect themselves and information security, and do not disclose their personal information, paying attention to screening job information, job search sites and job search authenticity to avoid being deceived. College students should develop good Internet habits, such as personal privacy settings, avoid using a mailbox to register multiple social networking sites in case of any chain reaction or harm to other social networking sites once one of the website account is leaked^[17]. College students also should protect their bank account and other financial information.

ACKNOWLEDGEMENT

This work was supported by the grants from Hubei Provincial Collaborative Innovation Centre of Agricultural E-Commerce(Wuhan Donghu university research [2016] No. 9 Document).

REFERENCES

- [1] Lukasz Sulkowski.(2014) .The use of Social Networking Sites in modern recruitment – empirical study . *Przedsiębiorczosci Zarzadzanie*,15(2):36-73.
- [2] Barbara Lohse .(2013).Facebook Is an Effective Strategy to Recruit Low-income Women to Online Nutrition Education. *Journal of Nutrition Education and Behavior*, (10):69-76.
- [3] Fangling Li. Research on User-oriented Website Evaluation of Recruitment Nanjing University.2016.5(In Chinese).
- [4] iResearch Consulting China Online Recruiting Industry Development Report .2016.Q1
- [5] Jin Li.(2013).Analysis of Social Network Job Search Behavior of Graduates in Campus Recruitment. Ms D Thesis. Chongqin:Southwest University. (In Chinese).
- [6] iResearch Consulting. China's Online Recruitment Industry Semi-annual Report.2017.
- [7] Jianlin Xu.Research on Use and Satisfaction of Social Recruitment Website. Ms D Thesis. Shanghai: Shanghai Jiaotong University. (In Chinese).
- [8] ChenjiaoHe.(2015) Exploration of Job Searching Patterns for College Students Based on Social Media. *Higher Education Exploration*.(3):114-118. (In Chinese).
- [9] YuelinLi, XiminYan.(2016).College Students Information Seeking Behavior during Job Hunting : Selection and Use of Information Source. *Documentation,Information & Knowledge*, (5):57-65.
- [10] BoWu. (2017)Research on Corporate Socialization Recruitment Under Background of Big Data [J], *Modern Business*, 4:86-89 (In Chinese).
- [11] XiaoWang.(2015)Research of Innovative Social Networking Platform Base on the Connection between Studies and Employments in Campus. Ms D Thesis.Hangzhou: Zhejiang University of Media and communication. (In Chinese).

- [12] Kumud Sagar,Vijaya Waghmare.(2016).Measuring the Security and Reliability of Authentication of Social Networking Sites. *Procedia Computer Science*.79(2):668-679.
- [13] Shailendra Rathore, Pradip Kumar Sharma, Vincenzo Loia .(2017).Social Network Security: Issues, Challenges, Threats, and Solutions. *Information Sciences*.421(8):43-69.
- [14] Joaquim Marques.(2014). Improving user content privacy on social networks using rights management systems. *annals of telecommunications* .69(2)37-45.
- [15] FanHe,YanLi.(2014).Analysis on social network job hunting behavior of graduates in Campus Recruitment.*Business*,(11):31-32. (In Chinese).
- [16] Lihao, Liyan Bao.(2013).The Practice and Thinking of Organizational SocializationRecruitment Methods in Age of WEB2.0. *Human Resources Development of China*,10:43-48. (In Chinese).
- [17] Daniela Popescu,Mircea Georgescu.(2016)Social Networks Security in Universities: Challenges and Solutions. *Annals of the Alexandru Ioan Cuza University – Economics*.62(1):53-63.

INDEX OF AUTHORS

- Ai Xu 583
 Bin Chen 252
 Bin Xu 145
 Bo Yang 298
 Caiyan Lin 406
 Chaofan Yang 476
 Cheng Hu 531
 Chengyi Yue 260
 Chong Zhang 204
 Chuan Luo 109
 Chuan Luo 341
 Chunmei Gan 415
 Chunxiang Xue 449
 Chunxiu Qin 22
 Dan Fang 48
 Daqing Zheng 432
 Ding Feng 490
 Ding Zhang 490
 Dongqing Luan 211
 Edward Kasabov 227
 Fan Feng 473
 Fanbo Meng 321
 Fanbo Meng 539
 Feifei Huang 56
 Fengfeng Lou 499
 Gang Ren 22
 Ge Zhang 173
 Hao Guo 173
 Hong Peng 235
 Hong Peng 243 484
 Hongwei Wang 330
 Hongxiu Li 546
 Hu Yang 252
 Huiming Xiao 145 589
 Huizhong Tan 555
 Huosong Xia 165
 Jason Cohen 31
 Jian Chai 9 155
 Jian Mou 22 31 515
 Jiang Wu 564
 Jiangping Wan 63 358 457
 Jianping Peng
 Jiaqi Pang 382
 Jiarui Li 9
 Jiawen Huang 63 457
 Jiaxin Tong 406
 Jing Quan 398
 Jing Wu 252
 Jingwen Liu 366 375 391
 Jinlin Wan 382
 Jishun Wang 173
 Jun Feng 124
 Jun Ma 48 490
 Jun Mi 235
 Kaijuan Xing 196
 Kee-hung Lai 321 539
 Kegui Liao 99
 Kerry Kurcz 22
 Lei Zhu 449
 Leimin Ye 523
 Leqi Xie 358
 Li Junbo 1
 Li Li 145 219 310 312 476 589
 Li Muchun 572
 Liang Chen 268
 Lin Huang 432
 Ling Ding 132 523
 Ling-wu Zeng 71
 Liqiang Wu 333
 Liu Tingting 572
 Lixin Gao 406
 Liyi Zhang 219 507
 Lu Chen 56
 Lu Sun 260
 Luo Manman 423
 Man Chen 79
 Meilin Chen 589
 Mengmeng Jin 564
 Mengqi Fei 555
 Mengzhi Xian 290
 Min Zhang 440
 Ming Li 196
 Mingli Zhang 79
 Minhoo Ju 515
 Ning Zhang 382
 Peijian Song 473
 Peiwen Guo 398
 Peng Guo 211
 Pengfei Xie 298
 Ping Wang 546
 Qihua Liu 507
 Qing Xu 555
 Qing Zhu 9 155 276
 Qingfang Sang 465

- Qu Fei 1
 Quansheng Wang 473
 Renxian Zuo 276
 Rong Zhang 298
 Rui chen 335
 Ruibo Yao 310 312 476
 Ruijun Zhang 406
 Samer Ben Issaa 88
 Shan Huang 507
 Shan Wang 440
 Shangyang Yu 406
 Shenhao Zhou 109
 SHI Ming-liang 189
 Shi Tang 583
 Shiyu Liu 290
 Shizhong Ai 39
 Shuang Tao 484
 Si Fan 366 375
 Silan Li 290
 Taiye Luo 398
 Ting He 335
 Ting Wang 155
 Ting Zhu 375 391
 Tingting Li 415
 Tingting Xiao 39
 Weijia You 140
 Weili Zhang 39
 Weiwei Liao 312
 Wen Guang Qu 499
 Wenmin Qi 310
 Wensheng Yang 48
 Xiang Yuan 330
 Xiaoli Yu 109
 Xiaoyu Qi 196
 Xin Jin 252
 Xin Zhang 173
 Xing Wan 117
 Xing-jian Zhou 71
 Xini Yuan 333
 Xinli Zhao 321 539
 Xinxin Zheng 406
 Xinyan Liu 124 531
 Xitong Guo 321 539
 Xiulian Jiang 132
 Xiuzhi Wu 449
 Xu Tianqi 572
 XU Ying-fan 189
 Xue Feng 117
 Xuefeng Li 350
 Xuewen Gui 99
 Xun Wang 341
 Xuwei Pan 523
 Yalan Li 341
 Yan Tu 350
 Yanan Xu
 Yang Li 145
 Yang Zhao 333
 Yanhong Chen 366
 Yanhong Chen 391
 Yanhui Zhang 211
 Yanli Pei 440
 Yanling Fan 79
 Yanping Liu 31
 Yi Chen 181
 Yi Cui 31
 Yi Hu 71
 Yi Zhang 196
 Yijia Gao 583
 Yiliu Tu 48
 Yiliu Tu 88 490
 Ying Wu 341
 Yingnan Zhao 473
 Yining Song 132
 Yiqiong Wu 155 276
 Yitai Yang 165
 Yitang Zeng 589
 Yong Liu 546
 Yu Wang 432
 Yu Zhao 333
 Yuan Meng 330
 Yuanheng Li 350
 Yuqi Zhang 440
 Yushan Yusufujiang 243
 Yuze Li 276
 Yuzhen Zhang 99
 Zhang Jue 1
 Zhang Ling 423
 Zhao Yuehan 1
 Zhenyan Hu 589
 Zhong Wang 63
 Zhu Lijun 423
 Zihan Jiang 132
 Zongqing Zhou 583
 Zongwei Li 211

Alfred University

Our Mission:

In all of our actions, we seek to assure that our students and faculty attain distinction to their personal, public and professional lives

Our primary vehicle for this is the development and delivery of programs of instruction and scholarship in professional management. Recognizing that we live in world in which technical advances, political shifts and social changes are all occurring rapidly, our principal objectives is to be a dynamic learning organization that prepares our students for leadership roles in their professions. We will design our systems and provide for incentives that build ongoing commitment to improving the quality of what we do.

In support of this mission our educational objectives are to:

1. Provide undergraduate and graduate programs that are excellent in quality, innovative in delivery, and relevant to current business practices.
2. Develop leadership and lifelong leadership skills.
3. Provide an environment, which fosters understanding and appreciation of cultural diversity and ethical conduct.
4. Support, conduct and disseminate scholarship in business.
5. Serve the community through programs and partnerships that enhance the intellectual quality of the region, and that enhance our core intellectual activities.
6. Provide active learning opportunities, which develop distinction through the acquisition of professional business skills.

Phone 607-871-2124

Fax 607-871-2114

Email businesscollege@alfred.edu

Website <http://business.alfred.edu>

China University of Geosciences

China University of Geosciences has gone through many changes and significant difficulties since 1952. The academic culture of universities like Beijing University, Qinghua University, etc. has greatly influenced the creation and the direction of this university. There are successively over 60 academicians who have ever industriously carried out teaching in the university; even more middle-aged and young experts have started their careers here. The university became one of the key universities in 1960, and was authorized to set up Graduate School in 1986. In 1997, it entered the 211 Project as one of the first group of universities in our country. It is therefore a key university being mainly created in China.

The university has splendid glories brought by outstanding records. It has shaped its own traditions and demeanors of being self-supporting and realistic. Up to now, the university has cultivated over 80,000 graduates. They are engaged in their own fields and dedicated themselves to the development of the country. Many of them have become academic elites, political genius, business tycoons, and excellent athletes. Premier Wen Jiabao and 21 academicians of Chinese Academy of Sciences and Engineering are typical representatives among them.

Over more than fifty years, people of CUG have been forging ahead, keeping pace with the times, and carrying out innovations. They have caught hold of the opportunities brought about by the reforms, and achieved rapid progress with their intelligence and efforts. Today, CUG has changed from just a geological college with single discipline to a comprehensive university offering multi-disciplines with geology, resources, environment, and geological engineering technology as the main features along with the development of sciences, engineering, liberal arts, management, economics and law.

Presently, the university has raised the goal of constructing itself into a world-class the field of earth sciences. In order to achieve the goal, we will inherit and carry forward the excellent traditions of CUG. We will regard people as the resource, discipline construction as the guide, improvement of teaching strength as the core, and talent cultivation as the lifeline. At the same time, we will intensify the conception of competition, quality and services, carry out reforms while remaining innovative, and aim to achieve outstanding credits in order to meet the demands of both the history and the future.

Yanxin Wang

President of China University of Geosciences, Wuhan, China