Marzia Zaman   Yawei Liang
Sohail M Siddiqui   Tim Wang
Vincent Liu   Ceecee Lu (Eds.)

Communications in Computer and Information Science 113

E-business Technology and Strategy

International Conference, CETS 2010
Ottawa, Canada, September 2010
Proceedings

Springer

Communications
in Computer and Information Science 113

Marzia Zaman   Yawei Liang
Sohail M Siddiqui   Tim Wang
Vincent Liu   Ceecee Lu (Eds.)

E-business Technology and Strategy

International Conference, CETS 2010
Ottawa, Canada, September 29-30, 2010
Proceedings

Springer

Volume Editors

Marzia Zaman
Queen's University
Kingston, Ontario, Canada
E-mail: zamanm@queensu.ca

Yawei Liang
Royal Military College of Canada
Kingston, Ontario, Canada
E-mail: liang-y@rmc.ca

Sohail M Siddiqui
Université du Québec en Outaouais
Gatineau, Québec, Canada
E-mail: sohail.siddiqui@uqo.ca

Tim Wang
CeBA Canada
Ottawa, Ontario, Canada
E-mail: tim.wang@cebacanada.com

Vincent Liu
Carleton University
Sprott School of Business
Ottawa, Ontario, Canada
E-mail: liuyide@connect.carleton.ca

Ceecee Lu
University of Toronto
Toronto, Ontario, Canada
E-mail: ceecee.lu@utoronto.ca

Library of Congress Control Number: 2010936132

CR Subject Classification (1998): K.6, K.4.4, H.4, H.3, J.1, H.2

ISSN 1865-0929
ISBN-10 3-642-16396-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-16396-8 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

The International Conference on E-business Technology & Strategy (CETS) provides a peer-reviewed forum for researchers from across the globe to share contemporary research on developments in the fields of e-business, information technology and business strategy. It seeks to promote effective and vibrant networking among researchers and practitioners from around the world who are concerned about the effective management of information technology in organizations.

This network of researchers views fostering the development of emerging scholars in the information technology and e-business fields as its primary task. Consequently the conference is designed to provide a venue for researchers to get substantive and beneficial feedback on their work.

There were 134 contributions submitted to CETS 2010. After in-depth discussions, 29 high-quality contributions were selected for publication in this volume. The authors are from Canada, USA, China, Japan, India and Malaysia.

We thank all the authors who submitted papers, the Program Committee members, and the external reviewers. We also thank all the local people who were instrumental in making this edition of CETS another very successful event. In particular, we are very grateful to Ying Xie, who was responsible for the local arrangements. Special gratitude goes to the publishing editor, Leonie Kunz, who managed the complexity of information and communication aspects. Furthermore, we thank the many students who volunteered on the organization team, as well as the IT services of Carleton University.

August 2010

Ceecee Lu
Marzia Zaman
Sohail M Siddiqui
Tim Wang
Vincent Liu
Yawei Liang

Organization

Conference Chair

Yawei Liang	Royal Military College of Canada

Program Committee Co-chairs

Mazia Zaman	Queen's University, Canada
Sohail M Siddiqui	Université du Québec en Outaouais, Canada
Tim Wang	Ceba Canada, Canada
Vincent Liu	Carleton University, Canada

Pogram Committee

Africa Arino	University of Navarra, Barcelona, Spain
Alain Sandoz	Vauban Technologies, Switzerland
Amit Bhardwaj	Meerut Institute of Engineering and Technology, India
Anne-Françoise Le Meur	Université de Lille, France
Ceecee Lu	University of Toronto, Canada
Christian Zirpins	University College London, UK
Craig Kuziemsky	University of Ottawa, Canada
David Douglas	University of Arkansas, USA
Darlene Lui	Ericsson Canada Inc, Canada
Emmanuel A. Abegunrin	Cappella University, USA
Gi Wu	Alcatel-Lucent Canada, Canada
Fahim A. Bhuiyan	Strayer University, Maryland, USA
Ferhat Khendek	Concordia University, Canada
Gilbert Babin	HEC Montréal, Canada
Guoping Xia	Beihang University, China
Guy Tremblay	LATECE – UQAM, Canada
Hafedh Mili	LATECE – UQAM, Canada
Hugh Deng	Agriculture Canada, Canada
Jacknan Wang	University of Toronto, Canada
Joseph Bonnici	Bryan College, Rhode Island, USA
Kamel Adi	Université du Québec en Outaouais, Canada
Karen Fernandez	University of Waikato, New Zealand
Kenneth Kraemer	University of California, Irvine, USA
Liam Peyton	University of Ottawa, Canada
Lizhi Long	University of Rhode Island, USA
Lyn S. Amine	Saint Louis University, Missouri, USA
Lorin M. Hitt	University of Pennsylvania, USA

Lu Liu	Beihang University, China
Nigel Melville	Boston College, USA
Peng Zhou	Carleton University, Canada
Petko Valtchev	LATECE – UQAM, Canada
Pravat K. Choudhury	Howard University, USA
Roy Rada	University of Maryland, USA
Rex Dumdum	Marywood University and State University of NY at Binghamton, USA
Shari Shang	The University of Melbourne, Australia
Stéphane Gagnon	Université du Québec en Outaouais, Canada
Sohail M. Siddiqui	Université du Québec en Outaouais, Canada
Soku Byoun	University of Southern Indiana, USA
Teck-Yong Eng	Aston University, UK
Tim Wang	CeBA Canada, Canada
Yan Zhuang	Notel Canada, Canada
Yawei Liang	Royal Military College of Canada
Vijay Gurbaxani	University of California, Irvine, USA
Vincent Liu	Carleton University, Carnada
Yi Wu	Agriculture Canada, Canada

Workshops and Tutorials Committee

Jacknan Wang	University of Toronto, Canada
Sohail M. Siddiqui	University of Quebec, Canada
Tim Wang	CeBA Canada, Canada

Publicity and Sponsorship Committee

Ceecee Lu	University of Toronto, Canada
Li Zhu	CeBA Canada, Canada
Vincent Liu	Carleton University, Canada

Local Arrangements Committee

Hongtao Li	Carleton University, Canada
Kyoko Hirano	CeBA Canada, Canada

Table of Contents

Research on Group Decision-Making Mechanism of Internet Emergency Management 1
Kefan Xie, Gang Chen, Wu Qian, and Zhao Shi

The Challenges and Issues Regarding E-Health and Health Information Technology Trends in the Healthcare Sector 23
Pouyan Esmaeilzadeh, Murali Sambasivan, and Naresh Kumar

The Study of Address Tree Coding Based on the Maximum Matching Algorithm in Courier Business 38
Shumin Zhou, Bin Tang, and Wen Li

Customer Credit Scoring Method Based on the SVDD Classification Model with Imbalanced Dataset 46
Bo Tian, Lin Nan, Qin Zheng, and Lei Yang

K-BRiR: An Efficient and Self-organized Routing Mechanism for Structured Peer-to-Peer Service Networks 61
Zhenhua Tan, Guangming Yang, Zhiliang Zhu, Wei Cheng, and Guiran Chang

Study on Amortization Time and Rationality in Real Estate Investment 75
Yancang Li, Shujing Zhou, and Juanjuan Suo

Risk Evaluation of Business Continuity Management by Using Green Technology 86
Chen Gang

Automated Methodologies for the Design of Flow Diagrams for Development and Maintenance Activities 93
Handigund Shivanand M. and Bhat Shweta

ACO-Based Projection Pursuit: A Novel Clustering Algorithm 105
Yancang Li, Lina Zhao, and Shujing Zhou

Ant Colony Algorithm for Multiple-Depot Vehicle Routing Problem with Shortest Finish Time 114
Jianhua Ma and Jie Yuan

Chaos Caused by a Transitive Map in Information 124
Lidong Wang, Jiao Peng, and Heng Liu

Comparative Study of Methods for Monitoring and Controlling a Regional Economic System 130
Shi Zheng, Wen Zheng, and Xia Jin

Energy Levels and Co-evolution of Product Innovation in Supply Chain Clusters 140
Guojun Ji

A Typology Framework of Loyalty Reward Programs 159
Yuheng Cao, Aaron L. Nsakanda, and Inder Jit Singh Mann

The Knowledge Transfer Factors under Different Perspectives 171
Yuanhua Hu and Lili Xue

IT Operational Risk Measurement Model Based on Internal Loss Data of Banks 180
Xiaoling Hao

Prediction Task Assignment of Multi-UAV Approach Based on Consensus 192
Chen Chen, Zheng Qin, and Jian-kuan Xing

A Macroscopic Timing Optimization Model for Signalized Intersections Based on an Extended Internal State Node Model 200
Ximin Liu, Shoufeng Lu, and Shiqiang Dai

A Research Methodology for Green IT Systems Based on WSR and Design Science: The Case of a Chinese Company 215
Yinghong Zhong and Hongwei Liu

Research on the Hotel Image Based on the Detail Service 226
Ban Li, Zheng Shenghua, and Yi He

Impact of Network Information on the Management of Library E-Resources 234
Zhixiang Yin and Hua Chen

Technology Route and System Design of Green IT 238
Juchi Hou, Ying Liang, and Jianli Wei

The Tail of BPM 245
Steve Kruba and Jim Meyer

The Aggregation-Diffusion Model of Conduction of Technology Value in the Large-Scale Sports Events 258
Chuansheng Dong

Specification Patent Management for Web Application Platform Ecosystem 269
Yoshiaki Fukami, Masao Isshiki, Hideaki Takeda, Ikki Ohmukai, and Jiro Kokuryo

Evaluation of the Acceptance of Audience Response System by Corporations Using the Technology Acceptance Model 281
Hsing-Hui Chu, Ta-Jung Lu, and Jong-Wen Wann

Identification of Variables and Factors Impacting Consumer Behavior in On-Line Shopping in India: An Empirical Study 292
Sudesh Chhikara

Knowledge Model of Emergency Decision-Making Based on Knowledge Supply and Demand 305
Bing Wu and Lin-du Zhao

The Consistency and Ranking Method Based on Comparison Linguistic Variable 318
Qisheng Zhao, Fajie Wei, and Shenghan Zhou

Author Index 327

Research on Group Decision-Making Mechanism of Internet Emergency Management*

Kefan Xie, Gang Chen, Wu Qian, and Zhao Shi

Management School, Wuhan University of Technology, 125 LuoShi Road,
Wuhan, Hubei 430070, China
xkf@whut.edu.cn

Abstract. With the development of information technology, internet has become a popular term and internet emergency has an intensive influence on people's life. This article offers a short history of internet emergency management. It discusses the definition, characteristics, and factor of internet emergency management. A group decision-making mechanism of internet emergency is presented based on the discussion. The authors establish a so-called Rough Set Scenario Flow Graphs (RSSFG) of group decision-making mechanism of internet emergency management and make an empirical analysis based on the RSSFG approach. The experimental results confirm that this approach is effective in internet emergency decision-making.

Keywords: Group Decision-making, Internet Emergency Management, Rough Set Scenario Flow Graphs.

1 Introduction

The information age is coming and information is one of indispensable resources to people. With the development of information technology, the application of computer network has deeply touched various fields, which plays a very important role in the development of politics, economy, technology and society. Therefore, network security becomes a crucial issue to social stability and economic development.

In December 2006, a computer virus named "Worm Nimaya" was rapidly spreading nationwide in China at the speed of attacking three to five hundred thousand computer systems per day, and produced 700 new heterogeneous viruses in just two months. Until January 2007, over millions of personal computers had been attacked in China, and plenty of personal and commercial files had been damaged. The direct and indirect economic losses for it were out of calculation. Moreover, the horrible virus caused a serious panic. Many netizens and commercial organizations stopped using network in order to protect their computers from the virus so that some online communication was broken. Hereafter, network security organization and some anti-virus software companies were developing anti-virus software to fight off it. At the same time, police office was investigating its human factor with the help of

* This research is supported by National Natural Science Foundation of China, project No. 90924010.

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 1–22, 2010.

national Computer Network Coordination Emergency Response Technical Team (CNCERT). The network order had been not restored until the virus-maker was under arrest in February 2007. But its “sequela” had lasted for one year. The process to deal with “Worm Nimaya” virus emergency is shown in figure 1 as follows.

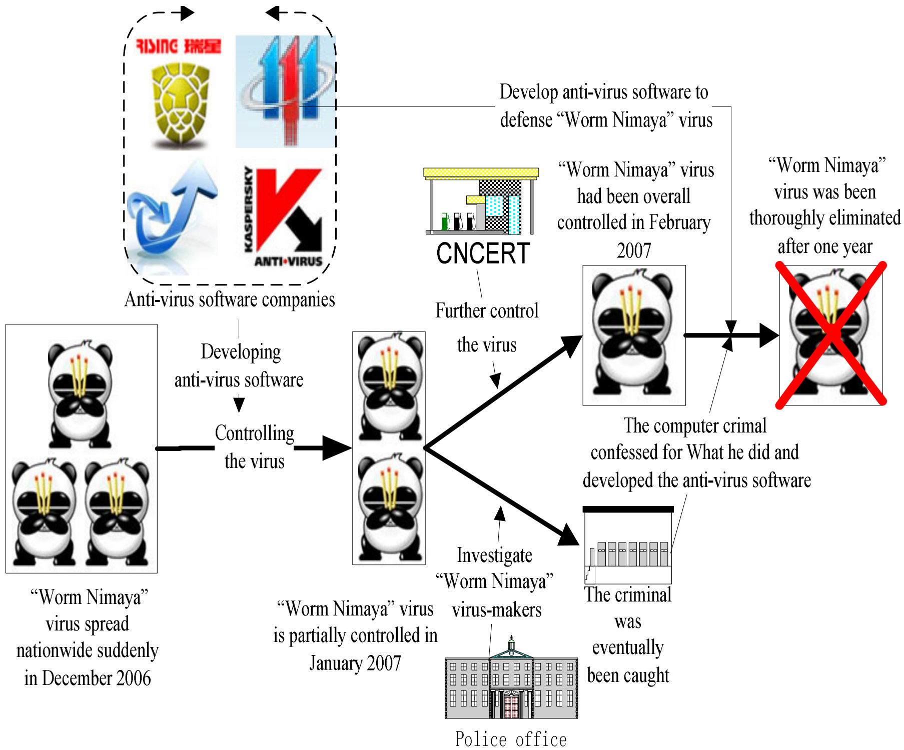


Fig. 1. The process of “Worm Nimaya” virus emergency

An intensive earthquake struck the ocean space southern of Taiwan on Dec. 26, 2006, which caused six national submarine optical cables interrupted, so that some international and regional communications ware severely affected. Most of communication lines among Hong Kong, Macao and Taiwan were interrupted, and the communication quality of the Mainland China could not escape from the disaster also, so that it caused many enterprises and individuals suffering great loss. At the same time, the other internet channels can not avoid serious congestion. So, the optical cable maintenance departments from related countries were forced to repair the cables immediately with the help of some maritime operation institutions. The optical cables had been repaired to normal one month later.

According to the two cases, it was found that internet emergency has very intensive and extensive influence and asks several organizations to deal with it in time. Take “Worm Nimaya” emergency for example, it not only involves some anti-virus software companies, but also CNCERT and police office etc.; in the process of dealing with cable interruption emergency, there are many departments to deal with it such as maintenance department of related country, maritime operation institutions, internet management department and communication corporations. In addition, internet emergency breaks out so suddenly that it can cause online communication interruption in short time and lead to enormous direct/indirect economic losses at the same time. The influence of internet emergency could increase quickly with time passing by. So decision must be made quickly to cope with internet emergency. Moreover, internet emergency is usually so complicated that it has great uncertainty to make decision. It usually needs to meet multiple goals when making decision with little information. Therefore, the process of internet emergency management is a multi-department decision-making in urgent situation with little information. In a word, normal emergency decision-making system can not satisfy internet emergency need. It should establish a group decision-making system with multi-organization and multi-department

At present, several scholars have done plenty of research works on the emergency decision-making system. Reference [1] first developed an emergency management evaluation model, and then proposes an extended fuzzy multi-criteria group evaluation method, which can deal with both subjective and objective criteria under multi-levels by a group of evaluators, for emergency management evaluation. Reference [2] discusses the applicability of management models on-scene an accident based on 20 years of experience as the commanding officer of a Norwegian fire department, and then they proposed some ideas on future research designs applied to real-time emergency management. Reference [3] proposed the social emergency early warning system to cope with the emergency. Reference [4] developed a dynamic decision making model to support the normative problem-solving process during emergencies. Reference [5] argued that increasingly the term virtual emergency operations centre was used to describe everything from an emergency operations centre with computers, through facilities that offer public information to internet users, to truly virtual facilities that apply incident command system management principles to supply electronic services. Reference [6] propose that enhanced learning of the non-technical skills, through experience and directed practice following repeated exposure to TDGs (tactical decision game), will lead to more efficient emergency management, particularly when dealing with hazardous materials. Reference[7] propose Multiple average-multiple threshold (MAMT) active queue management(AQM) as a solution for providing available and dependable service to traffic from emergency users after disasters. Reference [8] presents a dynamic relief-demand management model for emergency logistics operations under imperfect information conditions in large-scale natural disasters.

According to the above related researches, though many scholars have researched on emergency decision-making system, they mainly concerned on individual or little group decision-making and seldom on large group decision-making of multi-organization and multi-department. So, the paper studies large group decision-making of internet emergency according to its characteristics.

2 Internet Emergency Management

Internet Emergency is one kind of unexpected event based on internet media, which includes the subjects such as natural force and human power, caused by network media, and effects the objects such as network-users, internet organization and computer equipment. Therefore, Internet Emergency means an unexpected event which is caused by natural factors or human factors, and is destroyed to some important computer network or to large scope of computer network, seriously threats the security of country, society, person and property.

2.1 Classification of Internet Emergency

According to the cause, objects and affected region, internet emergency events can be categorized as follows:

a. According to the cause, they are divided into two categories, hardware-damaged internet emergency and software-destroyed internet emergency. The former refers to computer network interruption and network service termination for widespread damage of communication lines and equipment. The latter refers to such event as abnormal operation of network terminals and losing of stored data as computer software is destroyed;

b. According to the objects, three categories are divided as commercial internet emergency, government internet emergency and civil internet emergency. Commercial network is mostly used by enterprises, so commercial internet emergency mainly affects enterprise network and causes economic losses, even economy paralysis in one region. Governments are mainly users of government network, so government internet emergency can cause governmental information flow to interrupt, government administration system to fail and national secret information to betray. The users of civil network are mainly social citizens, so civil internet emergency can bring threats to the security of person and property, then give birth to social panic and unrest;

c. According to the affected region, there are two categories as follows: fixed-region internet emergency and non-fixed-region internet emergency. The former refers to an unexpected event that happens in one or several regions and hardly affects other areas. The latter refers to an unexpected event that its affected areas are uncertain.

2.2 Characteristics of Internet Emergency

As an unconventional event, besides internet emergency owns the general characteristics of emergency events, there are special characteristics as follows:

a. Unconventional and unexpected. It means internet emergency is very rare and its symptom is very unobvious before happening, the process of happening and evolution are irregular, so it is very difficult to be forecasted and controlled according to past experience, which means it's very difficult to early-warning and effective action.

b. Dependency of media. From beginning to end of internet emergency, computer network is a very important part of internet emergency, even internet emergency doesn't exist without media.

c. Extensive influence. Internet connects many areas and fields, so the influence of internet emergency usually crosses the boundary of one area, region or field. In addition, computer network is a net-like structure, so, when one side of terminal is affected, the other sides of terminal in the whole computer network also can be affected.

d. Intense destructive force. The destructive force of internet emergency is reflected in three respects as follows: First, from economy perspective, internet emergency could cause internal information interruption in enterprises, betray of business secrets etc., which may lead to economy losses of individuals and enterprises. Second, from social perspective, internet emergency probably cause social panic. Third, from maintainability perspective, the losses for internet emergency destruction are difficult to be repaired especially when computer stored data is lost and computer hardware equipment is destroyed;

e. Its source is difficult to be discerned. Since computer network is extensively used and network is widely distributed, it is hard to discern its initiating trigger factors once internet emergency happens.

2.3 The Three Factors of Internet Emergency

Internet emergency consists of three factors: cause factor, media factor and subject factor. The cause factor of internet emergency refers to the factor that leads to internet emergency events. The cause factor includes human factor, natural factor and other factors. The human factor refers to individual or organizational unconscious or intentional damaging activities to computer network. The natural factor refers to natural disasters or natural power which destroys computer network. Other factors refer to the special factors that cause computer network destructive. The cause factor directly determines how to find the reason of internet emergency and how to category internet emergency, which plays an important role in the process of internet emergency management. The media factor of internet emergency refers to the damaged parts of computer network, including network optical cables, network service, optical network terminal and optical network terminal software etc. It is the basis of internet emergency scheme and determines the orientations of internet emergency response. The subject of internet emergency consists of two parts as follows: affected subject and decision-making subject. The former can be divided into four levels: First level is national emergency mechanism including police office and security department etc. ; Second level is social infrastructure including bank, water supply system and power supply department etc. ; Third level is social organizations including schools, scientific research institutes and enterprises etc. ; Fourth level is personal computer users. The decision-making subject refers to the departments of internet emergency response, for example, police office, network management department, news media, and network technology department etc. Moreover, spread speed and influence scope are two attribute factors of internet emergency. The former refers to the changing number of affected individual, regional area and losses per unit time, which decides diffusion ability of internet emergency and is the vital basis of internet emergency action. The latter includes the affected areas and objects and its spread tendency, which is the basis of the control measures in the process of internet emergency response. The relationship among the three internet emergency factors is shown in figure 2 as follows:

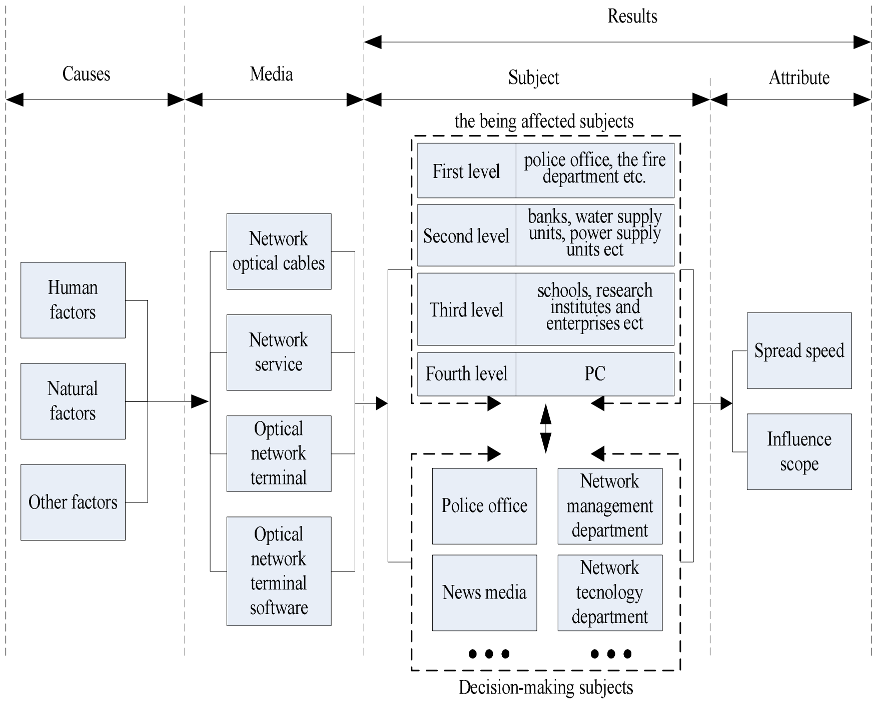


Fig. 2. The factor relationship of the internet emergency

3 Group Decision-Making Mechanism of Internet Emergency Management

3.1 General Operation Mechanism of Group Decision-Making in Internet Emergency Management

Based on the essential characteristics of internet emergency, the process of the group decision-making in internet emergency is quite different from the common decision-making process. The differences are listed as follows:

a. Decision makers should response emergency events more rapidly and take decision schemes more quickly. Internet emergency usually has extensive effect and destructive force. Therefore, internet emergency events will become too severe to control if the decision makers do not make the best decision early enough;

b. Internet emergency events are so complicated that many groups, including departments, enterprises and other individuals, will all participate in the decision-making process. Therefore, group decision-making requires a considerable level of cooperation to protect internet emergency action from disorder;

c. Internet emergency is highly unpredictable and changeable. Therefore, in order to make adaptable decisions, decision makers should track the emergency events unceasingly, test and correct decision schemes without delay.

Consequently, definition of group decision-making of internet emergency management is making choices in joint actions among several decision-making groups which consist of different decision-making subjects who act for the common purpose or benefit, or make profits for their own in the process of internet emergency.

To achieve the goal of agile response and adaptable decision, three systems are needed to combine group decision-making mechanism of internet emergency. They are internet emergency early warning system, internet emergency response system, and supervisory control system. The chief responsibilities of internet emergency early warning system are to predict the possible emergency and to sound the alarm. Internet emergency response system is in charge of analyzing the present state of emergency events, proposing internet emergency decision schemes, making decisions, and then executing the schemes by taking full advantage of all kinds of resources. Internet emergency supervisory control system is in charge of supervising the whole process of emergency decision-making, dealing with the subsequent controlling work, and sending feedback the decision results. The general mechanism of group decision-making in internet emergency can be shown in figure 3. The figure shows the laws and principles of group decision-making in internet emergency.

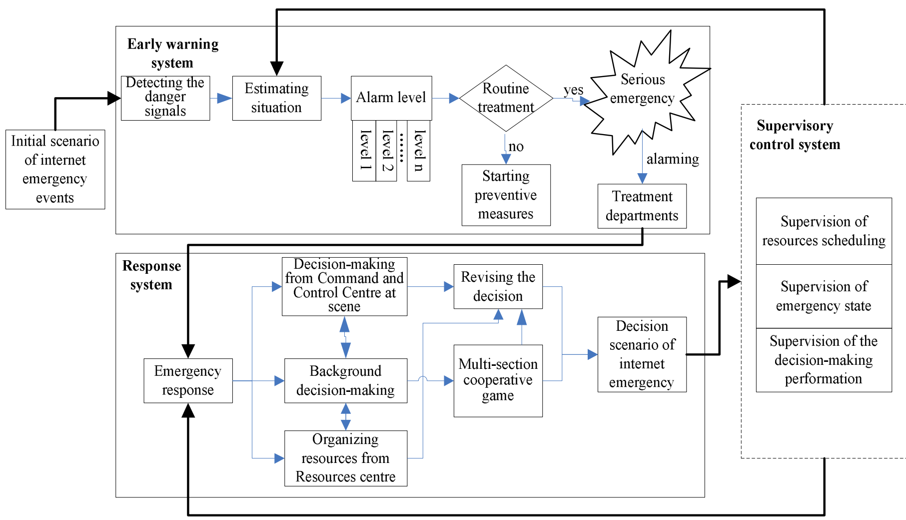


Fig. 3. Group decision-making systems of internet emergency

1. *Information flow is an important clue*

Information plays an important role in the whole decision mechanism. Firstly, information is the dominant characterization of internet emergency, and internet emergency can be analyzed qualitatively according to its showed information. Secondly, it can reflect that whether internet emergency decision result is good or not. For decision makers can obtain some important information of internet emergency process which is very important to group decision-making. Thirdly, the information is used by three systems of group decision-making mechanism in internet emergency.

So, the quicker the information flows and the sooner it updates, the quicker the systems will response internet emergency for decision-makers to take actions.

2. *The decision-making process of internet emergency is complicated*

In one way, because internet emergency happens suddenly and has a lot of uncertainty to the future, so its decision makers should adjust their decision in time according to the updated information. In other way, internet emergency impacts intensively and spreads extensively, so the affected subjects involve many departments, organizations and enterprises. When they are facing with internet emergency events, they usually either cooperate together or fighting for their self-interest. Therefore, decision-making game, conflict and cooperation are usually found in emergency management actions. Those emergency decisions and its actions become a driving force to promote the complicated evolution of internet emergency.

3. *The three systems are cooperative*

The group decision-making mechanism of internet emergency consists of early warning system, response system and supervisory control system. The three systems combine into a comparatively complete mechanism. Early-warning system is the prerequisite of the mechanism. Once early-warning system detects danger signals, response system and supervisory control system will be activated immediately. Response system is the core of the mechanism. In one way, response system takes the corresponding level of emergency scheme according to the severity that early-warning system provided. In other way, response system adjusts its decisions according to the emergency information or the analysis results of supervisory control system. Supervisory control system is the guarantee of the mechanism. It supervises the whole emergency management actions, feedbacks the results to early-warning system in order to re-evaluate the emergency severity, and feedbacks the latest situation of internet emergency and the results to response system in order to adjust its decision.

3.2 Analysis on the Three Systems of Internet Emergency Group Decision-Making

1. *The process analysis of internet emergency early-warning system*

Early-warning system consists of detecting information, estimating situation, alarm level and responding emergency. Firstly, it observes emergency information factors. The subject factors include internet optical cable, internet server, internet terminal equipment, etc. The environmental factors include weather, economic condition, social condition and other important factors. If the system detects danger signals from these factors, it will be reported to the detect centre of early-warning system. Secondly, the system makes a judgment to the observed events and determines the alarm level. Generally, the higher the alarm level, the more urgent and intensive it has. Thirdly, it will start the emergency scheme of the same level if the emergency can be solved by existing resources and abilities. If not, then it will start the alarm immediately and inform the related departments to take action.

2. *The process analysis of internet emergency response system*

Response system is the core part of the whole decision-making mechanism. Internet emergency is so serious that many corresponding departments and individuals will participate in the emergency management actions. Therefore, there are many groups to make decision. It forms multi-schemes by the groups in the decision process. The final decision is been made by the game process of the groups. Internet emergency response system is divided into three parts as follows: decision-making from the command and control centre at scene, background decision-making, and resource allocation decision-making. While, foreground decision-making is the execution unit of internet emergency response, which is under the direction of the background decision-making, is bound by resource allocation decision-making and is interfered by emergency situation. Resource allocation decision-making is the insurance unit of the needed resources to internet emergency response. It is under the direction of background decision-making. Background decision-making is the core unit of emergency response, which is based on the foreground information and resource information. The three decision-making processes have interaction to each other and is usually a multi-agent decision-making. Multi-agent decision-making may lead to the different decision schemes which can result in internal decision-making game.

3. *The process analysis of supervisory control system*

Supervisory control system is in charge of supervising the whole emergency management actions, having the subsequent control, and feedbacking the results and existing problem. When internet emergency is confirmed by early-warning system, supervisory control system starts to detect and track. In emergency management action, it collects some valuable information from emergency response system, such as resource usage, resource allocation, emergency situation and decision effect. In one way, it transfers them to early-warning system at intervals to observe whether internet emergency is perfectly controlled or not. If internet emergency has been controlled to the special level, the alarm will stop. In other way, it transmits those information to emergency response system at intervals in order to update emergency information of emergency response system which can response agilely and adjust the decision schemes to have the better effect.

4 Models of Group Decision-Making Mechanism of Internet Emergency Management

Flow Graph was proposed by Z. Pawlak in 2003[9], which was used to analyze the information flow decision. Flow graph can measure the relationship of nodes by using flow distribution among them. That is, it analyzes and reasons the data from the view of quantization, and it requires stable structure of the flow internet. However, there are still shortcomings. For example, it is not fit to dynamic update and self-learning because the results of credibility and coverage are stable; it can’t reflect the consistency of the knowledge system, and can’t accurately depict the relationship among the nodes; it not only can’t predict the future, but also can’t extend its structure. So, it is obvious that it isn’t suitable for internet emergency decision-making that is an unconventional decision-making, but if Scenario Analysis[10] is introduced to flow graph, the new method will have the qualitative prediction function and structure extension function, so it could be employed in group decision-making of internet emergency management and it is defined as Rough Set Scenario Flow Graph.

1. *Rough Set Scenario Flow Graph*

Definition 1: Rough Set Scenario Flow Graph (RSSFG for short) is directed acyclic graph $G = (N, B, \varphi)$, in which, N is node set, $B \subseteq N \times N$ is directed arc set, $\varphi : B \to 2^E$ is object set who flows arc.

Definition 2: In RSSFG, node set N refers to the set of the nodes through where all objects flow in G. The nodes include situation scenario node, situation response node, and situation system scenario node.

Assume x_{ti}, y_{tj} and z_{tk} respectively denote situation scenario node, situation response node and situation system scenario node; x_t, y_t and z_t respectively denote situation scenario node set, situation response node set and situation system scenario node set, $x_{ti} \in x_t \subset B$, $y_{tj} \in y_t \subset B$, $y_{tj} \in y_t \subset B$. Here situation scenario node x_{ti} is defined as the input of situation response node y_{tj}, situation scenario node y_{tj} is defined as the output of situation scenario node x_{ti}, $x_{ti} \to y_{tj}$ is situation rough set response arc, $\varphi(x_{ti}, y_{tj})$ is the flow rate of $x_{ti} \to y_{tj}$, then,

$$\varphi(x_{ti}, y_{tj}) = CD(x_{ti} \to y_{tj}) \cdot \tag{1}$$

Where, $CD(x_{ti} \to y_{tj})$ refers to the confidence of decision rule $x_{ti} \to y_{tj}$. Then, the probability of the situation response y_{tj} in situation scenario x_{ti} is $\varphi(y_{tj} | x_{ti}) = \varphi(x_{ti}, y_{tj})$.

Moreover, situation response node y_{tj} is defined as the input of situation system scenario node z_{tk}, situation scenario node z_{tk} is defined as the output of situation response node y_{tj}, $y_{tj} \to z_{tk}$ is situation response scenario analysis arc. In an open system, situation scenario x_t may trigger several situation responses y_{tj} $(j = 1, 2, \cdots, m_t)$. According to scenario analysis theory, several situation responses y_{tj} combine randomly based on different levels, which can lead to situation more complex and present several situation system scenarios z_{tk}. In a situation system scenario z_{tk}, any situation response y_{tj} may be the "major factor" or "minor factor", so m_t situation responses y_{tj} can combine randomly into 2^{m_t} situation system scenario z_{tk}. Obviously, whether a situation response y_{tj} is "major factor" or "minor factor" is correlated with its probability $\varphi(y_{tj} | x_{ti})$. When the probability $\varphi(y_{tj} | x_{ti})$ is very big, the situation response y_{tj} is more likely the major factor of situation system scenario z_{tk} ; otherwise, it is more likely the minor factor of situation system scenario z_{tk} . Therefore, in the situation system scenario z_{tk} , the probabilities that situation response y_{tj} is a major factor or minor factor can be described as $\varphi\left(y_{tj} \middle| x_{ti}\right)$ and $1-\varphi\left(y_{tj} \middle| x_{ti}\right)$. Moreover, situation system scenario contains some elements such as situation scenario and situation response, so, the probability of situation system scenario z_{tk} is related with whether its inputting situation responses y_{tj} are major factors or minor factors and their combination, and it can be solved as formula (2).

$$\varphi\left(z_{tk} \middle| x_{ti}\right) = \underset{j=1,2,\cdots,m_t}{f}\left(\varphi\left(y_{t,1} \middle| x_{ti}\right), \varphi\left(y_{t,2} \middle| x_{ti}\right), \cdots, \varphi\left(y_{t,m_t} \middle| x_{ti}\right)\right) \tag{2}$$

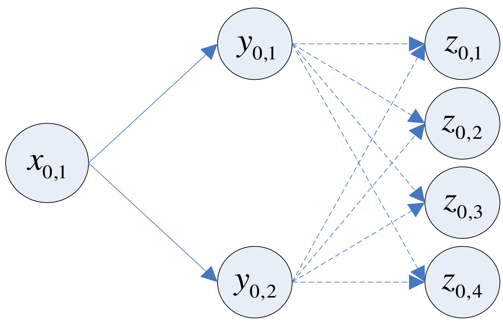


Fig. 4. An example of Rough Set Scenario Flow Graph

For example, as shown in Figure 4, there are two situation response node $y_{0,1}, y_{0,2}$ after situation scenario node $x_{0,1}$, then $2^2 = 4$ situation system scenario nodes z_t can be formed by random combination of situation responses $y_{0,1}, y_{0,2}$ as following: $z_{0,1} = \left\{x_{0,1}, \underline{y_{0,1}}\, \underline{y_{0,2}}\right\}$, $z_{0,2} = \left\{x_{0,1}, \overline{y_{0,1}}\, \overline{y_{0,2}}\right\}$, $z_{0,2} = \left\{x_{0,1}, \overline{y_{0,1}}\, \underline{y_{0,2}}\right\}$, $z_{0,2} = \left\{x_{0,1}, \underline{y_{0,1}}\, \overline{y_{0,2}}\right\}$, $\overline{y_{0,j}}$ and $\underline{y_{0,j}}$ denote respectively as " $y_{0,j}$ is a major factor" and " $y_{0,j}$ is a minor factor". Therefore, their probabilities can be solved according to formula (2) as following: $\varphi\left(z_{0,1} \middle| x_{0,1}\right) = \left[1-\varphi\left(y_{0,1}\right)\right]\left[1-\varphi\left(y_{0,2}\right)\right]$; $\varphi\left(z_{0,2} \middle| x_{0,1}\right) = \varphi\left(y_{0,1}\right)\varphi\left(y_{0,2}\right)$; $\varphi\left(z_{0,3} \middle| x_{0,1}\right) = \varphi\left(y_{0,1}\right)\left[1-\varphi\left(y_{0,2}\right)\right]$; $\varphi\left(z_{0,4} \middle| x_{0,1}\right) = \left[1-\varphi\left(y_{0,1}\right)\right]\varphi\left(y_{0,2}\right)$.

Definition 4: situation system scenario could promote the evolution of situation, that is $z_{tk} \rightarrow x_{t+1,i}$. So there are different evolution tendencies in different situation system scenarios. Here the path between situation system scenario node z_{tk} and situation scenario node $x_{(t+1),i}$ in the next stage is defined as situation evolution arc. Supposing that $\varphi\left(z_{tk}, x_{t+1,i}\right)$ is the flow rate of $z_{tk} \rightarrow x_{t+1,i}$, then,

$$\varphi\left(z_{tk}, x_{t+1,i}\right) = CD\left(z_{tk} \rightarrow x_{t+1,i}\right) \tag{3}$$

Where, $CD\left(z_{tk} \rightarrow x_{t+1,i}\right)$ refers to the confidence of decision rule $z_{tk} \rightarrow x_{t+1,i}$. Then, the probability of $x_{t+1,i}$ who is forced by situation scenario z_{tk} is as follows: $\varphi\left(x_{t+1,i} \middle| z_{tk}\right) = \varphi\left(z_{tk}, x_{t+1,i}\right)$. So, based on situation scenario $x_{t,i}$, its situation evolves to situation scenario $x_{t+1,i}$ in the next stage, its probability is as follows:

$$\varphi\left(x_{t+1,i} \middle| x_{t,i}\right) = \varphi\left(x_{t+1,i} \middle| z_{tk}\right) \cdot \varphi\left(z_{tk} \middle| x_{t,i}\right) = \varphi\left(z_{tk}, x_{t+1,i}\right) \cdot \varphi\left(z_{tk} \middle| x_{t,i}\right) \tag{4}$$

In formula (4), $\varphi\left(x_{t+1,i} \middle| x_{t,i}\right)$ is the flow rate of $x_{t,i} \rightarrow x_{t+1,i}$, which reveals the basic rules the situation evolves from scenario $x_{t,i}$ to scenario $x_{t+1,i}$.

So, RSSFG is a reasoning method to qualitative data by using rough set and scenario analysis, and can dig into the evolution rules of the general situation and draw the whole rough set scenario flow graph of situation evolution. The general rules of situation evolution can be dug by the ways of following methods:

a. By comparing between situation scenarios of two continuous stages, the deteriorated, stable or optimized trend can be examined.

b. By comparing situation scenario x_t with the early-warning threshold λ_t in the same stage, which is an early-warning to the situation evolution.

c. By comparing the situation scenario x_{t^*} in the target time t^* with the final aim, which can help select the best situation evolution paths and determine the best control variable z_{tk} in every stage.

2. *The RSSFG of internet emergency group decision-making*

In RSSFG, $x_{ti} \rightarrow y_{tj}$ and $y_{tj} \rightarrow z_{tk}$ can respectively represent a situation response rule and a situation evolution rule. In the process of internet emergency group decision-making, different internet emergency scenarios ask for different emergency decision; Even under the same internet emergency scenarios, different groups may have different emergency decisions. That is to say that there is game among group decision-making, which could affect the evolution of internet emergency. Therefore, the RSSFG of internet emergency group decision-making can be drawn based on the above method. Concretely speaking,

a. Every decision-maker has his decision in different situation scenario of internet emergency, which can form some decision information system.

$S_{x_t \to y_t} = \left(U, C_{x_t}, D_{y_t}\right)$. All decision rules $x_{ti} \to y_{tj}$ and their confidences $\varphi\left(x_{ti}, y_{tj}\right)$ can be determined by analyzing the information systems.

b. Based on scenario analysis theory, some situation system scenarios z_t are formed by random combination to all correlative decision rules. The probability of any situation system scenario z_{tk} can be obtained according to formula (2).

c. internet emergency can evolve to different situation scenario x_{t+1} under the forces of situation system scenario z_t, the information system $S_{z_t \to x_{t+1}} = \left(U, C_{z_t}, D_{x_{t+1}}\right)$ between z_t and x_{t+1} can be formed by collecting the information. Hereby, all evolution rules $z_{tk} \to x_{t+1,i}$ and their confidences $\varphi\left(z_{tk}, x_{t+1,i}\right)$ can be determined by analyzing the information systems.

d. Turn to step a. and analyze the next stage of internet emergency until meet the following condition, then the rough set flow stop and the RSSFG isn't extended again.

i. Exceed the research time;
ii. Achieve the control goal;
iii. Make the risk transfer.

3. *Group decision-making rule of internet emergency and situation evolution path*

Based on the RSSFG of internet emergency group decision-making, several evolution paths of internet emergency $x_0 \to x_1 \to \cdots \to x_{t^*}$ can be determined on the condition of multi-stage group decision-making. Supposing there are Γ paths which can achieve the control goal, every path depends on t_l. If $t_{l'} = \min\limits_{l=1,2,\cdots,\Gamma}\left(t_l\right)$, internet emergency evolution path $x_{0,i'} \to z_{0,k'} \to x_{1,i'} \to z_{1,k'} \to \cdots z_{t',k'} \to x_{t',k'}$ is the best control path, $z_{0,k'}, z_{1,k'}, \cdots, z_{t',k'}$ are respectively the best decisions of every stage in $\left[0, t'\right]$.

5 Example

Virus spread is a very important internet emergency, its spread speed and effect range are key factors to lead emergency unit to pay attention to. At the beginning of virus outbreak, most information about that computers is infected is feed backed to anti-virus software companies. So the initial scenario that its spread is very rapid and its effect is very extensive is firstly inputted to them, and anti-virus software companies are the first unit to emergency decision-making. On basic of the initial scenario ($x_{0,1}$ = {rapid spread, extensive effect}), they have two emergency countermeasures as follows: $y_{0,1}$ ={develop virus interception software}, $y_{0,2}$ ={develop anti-virus software}. According to investigation, in order to deal with Worm Nimaya virus, the decision rules of anti-virus software are explored by rough set as table 1.

Table 1. Emergency decision in initial scenario

Situation scenario			Decision	Confidence	Support
Code	Spread speed	Effect extension			
$x_{0,1}$	rapid	extensive	$y_{0,1}$	70%	42%
$x_{0,1}$	rapid	extensive	$y_{0,2}$	30%	18%

According to table 1, in the initial scenario $x_{0,1}$ ={rapid spread, extensive effect }, anti-virus software companies have two countermeasures, and the confidence and support of the decision rule $x_{0,1} \to y_{0,1}$ are respectively 70% and 42%, the confidence and support of the decision rule $x_{0,1} \to y_{0,2}$ are respectively 30% and 18%. On the effect of these two countermeasures, the virus spread event are in four scenarios as follows: $y_{0,1}$ and $y_{0,2}$ aren't major countermeasures to $x_{0,1}$ and the new countermeasure is continued to look for; $y_{0,1}$ and $y_{0,2}$ are major countermeasures to $x_{0,1}$; only $y_{0,1}$ is only a major countermeasures to $x_{0,1}$; only $y_{0,2}$ is only a major countermeasures to $x_{0,1}$. These form four situation system scenarios of the virus spread as following $z_{0,1}$, $z_{0,2}$, $z_{0,3}$ and $z_{0,4}$, and their probabilities can be solved by formula (2) as following: $\varphi\left(z_{0,1} \middle| x_{0,1}\right)=0.21$, $\varphi\left(z_{0,2} \middle| x_{0,1}\right)=0.21$, $\varphi\left(z_{0,3} \middle| x_{0,1}\right)=0.49$ and $\varphi\left(z_{0,4} \middle| x_{0,1}\right)=0.09$. On basic of the different situation system scenarios, the virus spread event evolves to some new orientations, which can be explored by rough set as table 2.

According to table 2, in the scenario $z_{0,1}$= { $x_{0,1}$, $y_{0,1}$ and $y_{0,2}$ aren't effectively carried out}, the virus spread event makes further worse and its situation scenario evolves to $x_{1,1}$= {very rapid spread, very extensive effect}, its confidence and support are respectively 100% and 25%. In the scenario $z_{0,2}$ = { $x_{0,1}$, $y_{0,1}$ and $y_{0,2}$ are major countermeasures and are carried out}, the virus spread event is controlled partially or completely, so its situation scenario could evolve to $x_{1,2}$ ={slow spread, not widespread} or $x_{1,3}$ ={very slow spread, small extent}, and their confidences are

Table 2. The new evolution scenarios of the virus spread event

Situation system scenarios			New scenarios	Confidence	Support
Code	Initial situation	Decision			
$z_{0,1}$:	$x_{0,1}$ ={rapid spread, extensive effect}	$\underline{y_{0,1}}\,\underline{y_{0,2}}$	$x_{1,1}$	100%	25%
$z_{0,2}$:	$x_{0,1}$ ={rapid spread, extensive effect}	$\overline{y_{0,1}}\,\overline{y_{0,2}}$	$x_{1,2}$	63%	15.75%
$z_{0,2}$:	$x_{0,1}$ ={rapid spread, extensive effect}	$\overline{y_{0,1}}\,\overline{y_{0,2}}$	$x_{1,3}$	37%	9.25%
$z_{0,3}$	$x_{0,1}$ ={rapid spread, extensive effect}	$\overline{y_{0,1}}\,\underline{y_{0,2}}$	$x_{1,2}$	79%	19.75%
$z_{0,3}$	$x_{0,1}$ ={rapid spread, extensive effect}	$\overline{y_{0,1}}\,\underline{y_{0,2}}$	$x_{1,3}$	21%	5.25%
$z_{0,4}$	$x_{0,1}$ ={rapid spread, extensive effect}	$\underline{y_{0,1}}\,\overline{y_{0,2}}$	$x_{1,2}$	83%	20.75%
$z_{0,4}$	$x_{0,1}$ ={rapid spread, extensive effect}	$\underline{y_{0,1}}\,\overline{y_{0,2}}$	$x_{1,3}$	17%	4.25%

respectively 63% and 37%. In the scenario $z_{0,3}$= { $x_{0,1}$, $y_{0,1}$ and $y_{0,2}$ are respectively major and auxiliary countermeasures and are carried out},the event is controlled partially or completely, so its situation scenario could evolve to $x_{1,2}$ ={slow spread, not widespread} or $x_{1,3}$={very slow spread, small extent}, and their confidences are respectively 79% and 21%. In the scenario $z_{0,4}$ = { $x_{0,1}$, $y_{0,1}$ and $y_{0,2}$ are respectively major and auxiliary countermeasures and are carried out}, the event is controlled partially or completely, so its situation scenario could evolve to $x_{1,2}$ ={slow spread, not widespread} or $x_{1,3}$={very slow spread, small extent}, and their confidences are respectively 83% and 17%. So, after the emergency action of anti-virus software companies, the event evolves to three results respectively $x_{1,1}$, $x_{1,2}$ and $x_{1,3}$, and their probabilities can be solved by formula (4) as following: $\varphi\left(x_{1,1}\middle|x_{0,1}\right)=0.21$, $\varphi\left(x_{1,2}\middle|x_{0,1}\right)=0.5941$, $\varphi\left(x_{1,3}\middle|x_{0,1}\right)=0.1959$.

On basic of the new situation scenarios, emergency units can take some new countermeasures, especially, when the event makes further worse, police and

CNCERT must take part in the emergency action. So, there are three countermeasures which respectively come from police, CNCERT and anti-virus software companies, and their countermeasures are respectively: $y_{1,1}$={search criminal and control virus spread by solving the case}, $y_{1,2}$={diagnosis virus mechanism and search anti-virus countermeasures} and $y_{1,3}$={upgrade anti-virus software}. Their decision rules can be explored by rough set as table 3.

Table 3. Emergency decision in scenario of t=1

Situation scenario			Decision	Confidence	Support
Code	Spread speed	Effect extension			
$x_{1,1}$	very rapid spread	very extensive effect	$y_{1,1}$ $y_{1,3}$	64%	25.6%
$x_{1,1}$	very rapid spread	very extensive effect	$y_{1,2}$ $y_{1,3}$	36%	14.4%
$x_{1,2}$	slow spread	not widespread	$y_{1,1}$ $y_{1,3}$	46%	18.4%
$x_{1,2}$	slow spread	not widespread	$y_{1,2}$ $y_{1,3}$	54%	21.6%
$x_{1,3}$	very slow spread	small extent	y *	100%	20.0%

According to table 3, whether the event is controlled completely or partially, anti-virus software upgrade is very necessary, its credibility and support are 100%. However, in the scenario $x_{1,1}$= {very rapid spread, very extensive effect}, emergency decision is more likely to $y_{1,1}$ than $y_{1,2}$, and the credibility of the former decision rule is 64% and the one of the latter is 36%. In the scenario $x_{1,2}$ = {slow spread, not widespread}, emergency decision is more likely to $y_{1,2}$ than $y_{1,1}$, and the credibility of the former decision rule is 54% and the one of the latter is 46%. In the scenario $x_{1,3}$= {very slow spread, small extent}, there isn't more subsequent countermeasures. In order to improve the effect of the emergency decision, it is necessary to coordinate the three countermeasures. After coordination, there are eight situation system scenarios as following: $y_{1,1}$ and $y_{1,2}$ aren't major countermeasures to $x_{1,1}$ and only $y_{1,3}$ is adopted; $y_{1,1}$ and $y_{1,2}$ are major countermeasures to $x_{1,1}$ and $y_{1,3}$ is adopted too; only $y_{1,1}$ and $y_{1,3}$ are major countermeasures to $x_{1,1}$; only $y_{1,2}$ and $y_{1,3}$ are major countermeasures to $x_{1,1}$; either $y_{1,1}$ or $y_{1,2}$ isn't major countermeasures to $x_{1,2}$ and only $y_{1,3}$ is adopted; both $y_{1,1}$ and $y_{1,2}$ are major countermeasures to $x_{1,2}$ and

$y_{1,3}$ is adopted too; only $y_{1,1}$ and $y_{1,3}$ are major countermeasures to $x_{1,2}$; only $y_{1,2}$ and $y_{1,3}$ are major countermeasures to $x_{1,2}$, which form eight situation system scenario nodes as following: $z_{1,1}, z_{1,2}, z_{1,3}, z_{1,4}, z_{1,5}, z_{1,6}, z_{1,7}, z_{1,8}$, and their probabilities can be solved according to formula (2) as following: $\varphi\left(z_{1,1} \mid x_{1,1}\right) = 0.2304$, $\varphi\left(z_{1,2} \mid x_{1,1}\right) = 0.2304$, $\varphi\left(z_{1,3} \mid x_{1,1}\right) = 0.4096$, $\varphi\left(z_{1,4} \mid x_{1,1}\right) = 0.1296$, $\varphi\left(z_{1,5} \mid x_{1,2}\right) = 0.2484$, $\varphi\left(z_{1,6} \mid x_{1,2}\right) = 0.2484$, $\varphi\left(z_{1,7} \mid x_{1,2}\right) = 0.2116$ and $\varphi\left(z_{1,8} \mid x_{1,2}\right) = 0.2916$. On the conditions of the different situation system scenarios, the virus spread event could evolve to different orientations, which can be explored by rough set as table 4.

According to table 4, in the effect of situation system scenario $z_{1,1}$, the event could evolve to $x_{2,1}$ ={spread speed and effect extension are out of control} or $x_{2,2}$ = {spread speed and effect extension are partially controlled}, and their confidences are respectively 87% and 13%, that is $\varphi\left(z_{1,1}, x_{2,1}\right) = 0.87, \varphi\left(z_{1,1}, x_{2,2}\right) = 0.13$; In the effect of situation system scenario $z_{1,2}$, the event could evolve to $x_{2,2}$ ={spread speed and effect extension are partially controlled} or $x_{2,3}$ = {spread speed and effect extension are completely controlled} , and their confidences are respectively 24% and 76%, that is $\varphi\left(z_{1,2}, x_{2,2}\right) = 0.24, \varphi\left(z_{1,2}, x_{2,3}\right) = 0.76$. Similarly, $\varphi\left(z_{1,3}, x_{2,2}\right) = 0.32$, $\varphi\left(z_{1,3}, x_{2,3}\right) = 0.68$, $\varphi\left(z_{1,4}, x_{2,2}\right) = 0.37$, $\varphi\left(z_{1,4}, x_{2,3}\right) = 0.63$, $\varphi\left(z_{1,5}, x_{2,1}\right) = 0.78$, $\varphi\left(z_{1,5}, x_{2,2}\right) = 0.22$, $\varphi\left(z_{1,6}, x_{2,2}\right) = 0.16$, $\varphi\left(z_{1,6}, x_{2,3}\right) = 0.84$, $\varphi\left(z_{1,7}, x_{2,2}\right) = 0.29$, $\varphi\left(z_{1,7}, x_{2,3}\right) = 0.71$, $\varphi\left(z_{1,8}, x_{2,2}\right) = 0.33$, $\varphi\left(z_{1,8}, x_{2,3}\right) = 0.67$. So, when t=2, the event evolves to three situation scenarios $x_{2,1}$, $x_{2,2}$ and $x_{2,3}$, and their probabilities can be solved according to formula (4) as following: $\varphi\left(x_{2,1} \mid x_{1,1}\right) = 0.2$, $\varphi\left(x_{2,2} \mid x_{1,1}\right) = 0.2643$, $\varphi\left(x_{2,3} \mid x_{1,1}\right) = 0.535$, $\varphi\left(x_{2,1} \mid x_{1,2}\right) = 0.194$, $\varphi\left(x_{2,2} \mid x_{1,2}\right) = 0.252$, $\varphi\left(x_{2,3} \mid x_{1,2}\right) = 0.554$

After several emergency actions, the virus spread event could still evolve to $x_{2,1}$ = {spread speed and effect extension are out of control} and $x_{2,2}$ = {spread speed and effect extension are partially controlled}. So, in order to control the event quickly, emergency units must carry out more intense and uniformed action. Anti-virus software regeneration is a necessary measure. In the process of anti-virus software

Table 4. The new evolution scenarios when t=2

Situation system scenarios			New scenarios	Confidence	Support
Code	Initial situation	Decision			
$z_{1,1}$:	$x_{1,1}$ ={very rapid spread, very extensive effect}	$\underline{y_{1,1}}\,\underline{y_{1,2}}\,y_{1,3}$	$x_{2,1}$	87%	10.88%
$z_{1,1}$:	$x_{1,1}$ ={very rapid spread, very extensive effect}	$\underline{y_{1,1}}\,\underline{y_{1,2}}\,y_{1,3}$	$x_{2,2}$	13%	1.63%
$z_{1,2}$:	$x_{1,1}$ ={very rapid spread, very extensive effect}	$\overline{y_{1,1}}\,\overline{y_{1,2}}\,y_{1,3}$	$x_{2,2}$	24%	3.00%
$z_{1,2}$:	$x_{1,1}$ ={very rapid spread, very extensive effect}	$\overline{y_{1,1}}\,\overline{y_{1,2}}\,y_{1,3}$	$x_{2,3}$	76%	9.50%
$z_{1,3}$	$x_{1,1}$ ={very rapid spread, very extensive effect}	$\overline{y_{1,1}}\,\underline{y_{1,2}}\,y_{1,3}$	$x_{2,2}$	32%	4.00%
$z_{1,3}$	$x_{1,1}$ ={very rapid spread, very extensive effect}	$\overline{y_{1,1}}\,\underline{y_{1,2}}\,y_{1,3}$	$x_{2,3}$	68%	8.50%
$z_{1,4}$	$x_{1,1}$ ={very rapid spread, very extensive effect}	$\underline{y_{1,1}}\,\overline{y_{1,2}}\,y_{1,3}$	$x_{2,2}$	37%	4.63%
$z_{1,4}$	$x_{1,1}$ ={very rapid spread, very extensive effect}	$\underline{y_{1,1}}\,\overline{y_{1,2}}\,y_{1,3}$	$x_{2,3}$	63%	7.88%
$z_{1,5}$:	$x_{1,2}$ ={slow spread, not widespread}	$\underline{y_{1,1}}\,\underline{y_{1,2}}\,y_{1,3}$	$x_{2,1}$	78%	9.75%
$z_{1,5}$:	$x_{1,2}$ ={slow spread, not widespread}	$\underline{y_{1,1}}\,\underline{y_{1,2}}\,y_{1,3}$	$x_{2,2}$	22%	2.75%
$z_{1,6}$:	$x_{1,2}$ ={slow spread, not widespread}	$\overline{y_{1,1}}\,\overline{y_{1,2}}\,y_{1,3}$	$x_{2,2}$	16%	2.00%
$z_{1,6}$:	$x_{1,2}$ ={slow spread, not widespread}	$\overline{y_{1,1}}\,\overline{y_{1,2}}\,y_{1,3}$	$x_{2,3}$	84%	10.50%
$z_{1,7}$	$x_{1,2}$ ={slow spread, not widespread}	$\overline{y_{1,1}}\,\underline{y_{1,2}}\,y_{1,3}$	$x_{2,2}$	29%	3.63%
$z_{1,7}$	$x_{1,2}$ ={slow spread, not widespread}	$\overline{y_{1,1}}\,\underline{y_{1,2}}\,y_{1,3}$	$x_{2,3}$	71%	8.88%
$z_{1,8}$	$x_{1,2}$ ={slow spread, not widespread}	$\underline{y_{1,1}}\,\overline{y_{1,2}}\,y_{1,3}$	$x_{2,2}$	33%	4.13%
$z_{1,8}$	$x_{1,2}$ ={slow spread, not widespread}	$\underline{y_{1,1}}\,\overline{y_{1,2}}\,y_{1,3}$	$x_{2,3}$	67%	8.38%

regeneration, anti-virus software companies is a predominate role, CNCERT provides technical support and the maker of virus is forced to it by police. So their uniformed action is $y_{2,1}$= {anti-virus software regeneration}, there are two situation system scenarios when t=2 as following: $z_{2,1}$ ={ $x_{2,1}$, $y_{2,1}$ }, $z_{2,2}$ ={ $x_{2,3}$, $y_{2,1}$ }, and $\varphi\left(z_{2,1} \middle| x_{2,1}\right) = \varphi\left(z_{2,2} \middle| x_{2,2}\right) = 1$. On the effect of the two situation system scenarios, the event could evolve to two situation scenarios $x_{3,1}$={the virus spread event is completely controlled} or $x_{3,2}$ ={the virus spread event is partially controlled}, the two results are explored by rough set as table 5. According to table 5., in the scenario of $z_{2,1}$ ={ $x_{2,1}$, $y_{2,1}$ }, the virus spread still could be partially controlled, its confidence is 4%, that is $\varphi\left(x_{3,2} \middle| z_{2,1}\right) = 0.04$ .similarly, $\varphi\left(x_{3,1} \middle| z_{2,1}\right) = 0.96$, $\varphi\left(x_{3,1} \middle| z_{2,2}\right) = 1$.

Table 5. The new evolution scenarios when t=3

Situation system scenarios			New scenarios	Confidence	Support
Code	Initial situation	Decision			
$z_{2,1}$:	$x_{2,1}$ ={spread speed and effect extension are out of control}	$y_{2,1}$	$x_{3,1}$	96%	38.4%
$z_{2,1}$:	$x_{2,1}$ ={spread speed and effect extension are out of control}	$y_{2,1}$	$x_{3,2}$	4%	1.6%
$z_{2,2}$:	$x_{2,2}$ = {spread speed and effect extension are partially controlled}	$y_{2,1}$	$x_{3,1}$	100%	60%

The evolution process of the virus spread event could be explored by above scenario analysis and rough set, and its rough set scenario flow graph could be drawn as Figure 5. According to the graph, some group decision mechanisms of internet emergency management could be explored.

In order to control the virus spread before t=1, some countermeasures which could lead $x_{0,1} \rightarrow x_{1,3}$ can be adopted, situation system scenarios $z_{0,2}$, $z_{0,3}$ and $z_{0,3}$ can lead the event to evolve from scenario $x_{0,1}$ to scenario $x_{1,3}$, and their probabilities are respectively $\varphi\left(x_{1,3} \middle| z_{0,2}\right) = 0.37$, $\varphi\left(x_{1,3} \middle| z_{0,3}\right) = 0.21$, $\varphi\left(x_{1,3} \middle| z_{0,4}\right) = 0.17$. So, when t=0, the best decision scenario is $z_{0,2}$, that is both $y_{0,1}$= {develop virus interception software} and $y_{0,2}$ = {develop anti-virus software} are major countermeasures to $x_{0,1}$ when t=0.

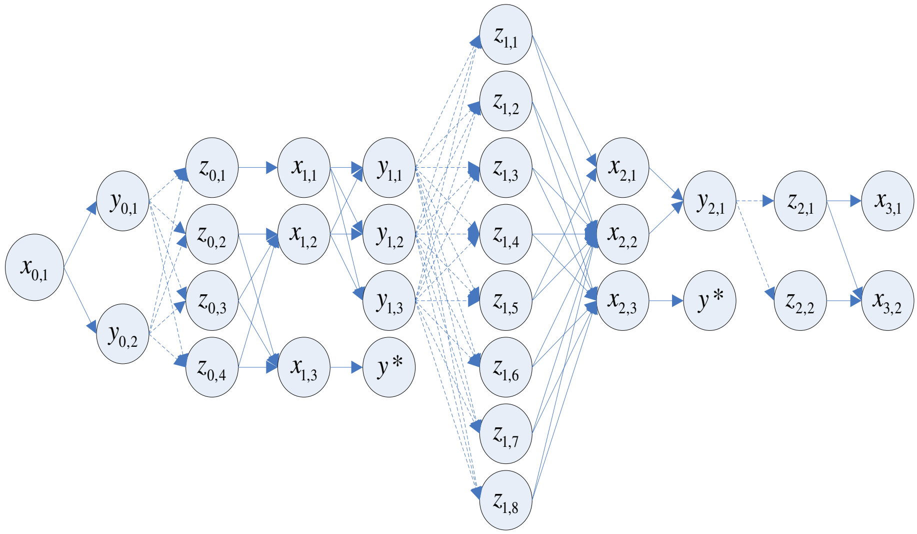


Fig. 5. Rough set scenario flow graph of virus spread

In order to control the virus spread before t=2, some countermeasures which could lead $x_{0,1} \longrightarrow x_{2,3}$ can be adopted. When the situation system scenario is $z_{0,1}$, $z_{0,2}$, $z_{0,3}$ or $z_{0,4}$ when t=0, the probabilities which the virus spread can be controlled before t=2 are respectively: $\varphi\left(x_{2,3}\left|z_{0,1}\right.\right)=\varphi\left(x_{1,1}\left|z_{0,1}\right.\right)\cdot\varphi\left(x_{2,3}\left|x_{1,1}\right.\right)=0.535$; $\varphi\left(x_{2,3}\left|z_{0,2}\right.\right)+\varphi\left(x_{1,3}\left|z_{0,2}\right.\right)=0.71902$; $\varphi\left(x_{2,3}\left|z_{0,3}\right.\right)+\varphi\left(x_{1,3}\left|z_{0,3}\right.\right)=0.64766$; $\varphi\left(x_{2,3}\left|z_{0,4}\right.\right)+\varphi\left(x_{1,3}\left|z_{0,4}\right.\right)=0.62982$ .So, the best decision scenario is $z_{0,2}$ when t=0, that is, both $y_{0,1}$ and $y_{0,2}$ are major countermeasures to $x_{0,1}$ when t=0. when the situation system scenario is $z_{1,1}$, $z_{1,2}$, $z_{1,3}$, $z_{1,4}$, $z_{1,5}$, $z_{1,6}$, $z_{1,7}$ or $z_{1,8}$ when t=1, the probabilities which the virus spread can be controlled before t=2 are respectively: $\varphi\left(x_{2,3}\left|z_{1,1}\right.\right)=0$, $\varphi\left(x_{2,3}\left|z_{1,2}\right.\right)=0.76$, $\varphi\left(x_{2,3}\left|z_{1,3}\right.\right)=0.68$, $\varphi\left(x_{2,3}\left|z_{1,4}\right.\right)=0.63$, $\varphi\left(x_{2,3}\left|z_{1,5}\right.\right)=0$, $\varphi\left(x_{2,3}\left|z_{1,6}\right.\right)=0.84$, $\varphi\left(x_{2,3}\left|z_{1,7}\right.\right)=0.71$, $\varphi\left(x_{2,3}\left|z_{1,8}\right.\right)=0.67$. So, the best decision scenario is $z_{1,2}$ when t=1, that is, $y_{1,1}$ and $y_{1,2}$ are major countermeasures to $x_{1,1}$ and $y_{1,3}$ is adopted too when t=1; the best decision scenario is $z_{1,6}$ when t=1, that is, $y_{1,1}$ and $y_{1,2}$ are major countermeasures to $x_{1,2}$ and $y_{1,3}$ is adopted too when t=1.

On basic of above method, the rough set scenario flow graph of virus spread can be extended, the further decision rules can be explored too. So, rough set scenario flow graph is an important method to improve group emergency decision.

6 Conclusions

Comparing with the common emergency, internet emergency has some characteristics, such as medium dependence, widespread impact, great destruction, and origin enshrouding. Based on these characteristics, internet emergency management system should be constructed to deal with internet emergency and it should include early-warning system, emergency response system, and supervisory control system. The internet emergency management system is also based on the group decision-making, but it needs not only the independent decision-making but also the collective decision-making. The influences of decision-making on the evolution of internet emergency must be discussed in order to make scientific decision from independent decision, collective decision and their combining result. According to that reason, the paper proposes the RSSFG method. The method can dig and predict the evolution rules of internet emergency management, which is important to research on the whole process control method.

As an important unconventional emergency, network emergency has growing impact in people's lives. Different from other general emergency, network emergency has several characteristics, including media dependence, extensive affect, strong destructive, hidden source etc., which determine that early warning system, emergency monitoring system and network control system should be established as an comprehensive system of three integrated emergency response system in the network emergency management process. The comprehensive system is built on the basis of group decision, which demands decision-makers independent in decision-making as well as complement each other. Timing of independent or coordinated decision-making is very important, and it needs of emergency decision-making on the evolution of the network, which depends on the method of rough set scenario flow graph the paper proposed. This method combines the rough set, the flow graph and scenario analysis, which can effectively mine and forecast network emergency evolution in the process of emergency response. The method provides useful guidance for researching the control of the entire process.

References

1. Zhang, G., Ma, J., Lu, J.: Emergency Management Evaluation by a Fuzzy Multi-criteria Group Decision Support System. Stochastic Environmental Research and Risk Assessment 23(4), 517–527 (2009)
2. Rake, L.: Emergency Management and Decision Making on Accident Scenes: Taxonomy, Models and Future Research. International Journal of Emergency Management 1(4), 397–409 (2003)
3. Xiao, Z., Wen, D., Na, P.: Study on the Mode of Social Emergency Early Warning Management based on Grid. In: Proceedings of 2007 International Conference on Management Science and Engineering, pp. 2463–2468 (2008)
4. Pan, T., Hong, W., Wei, Z.: Decision Making based on Emergency Plan Templates. In: Yu, W., He, H., Zhang, N. (eds.) ISNN 2009. LNCS, vol. 5551, pp. 1181–1190. Springer, Heidelberg (2009)

5. Green, W.: E-emergency Management in the USA: A Preliminary Survey of the Operational State of the Art. International Journal of Emergency Management 1(1), 70–81 (2003)
6. Crichton, M., Flin, R.: Training for Emergency Management: Tactical Decision Games. Journal of Hazardous Materials 88(1), 255–266 (2001)
7. Joshi, M., Mansata, A., Talauliker, S., Beard, C.: Design and Analysis of Multi-level Active Queue Management Mechanisms for Emergency Traffic. Computer Communications 28(2), 162–173 (2005)
8. Sheu, J.: Dynamic Relief-demand Management for Emergency Logistics Operations under Large-scale Disasters. Transportation Research Part E: Logistics and Transportation Review 46(1), 1–17 (2010)
9. Pawlak, Z.: Flow Graphs and Decision Algorithms. In: The Ninth International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, pp. 2463–2468 (2003)
10. Kahn, H., Wiener, A.: The Year 2000. MacMillan, New York (1967)

The Challenges and Issues Regarding E-Health and Health Information Technology Trends in the Healthcare Sector

Pouyan Esmaeilzadeh, Murali Sambasivan, and Naresh Kumar

Graduate School of Management, University Putra Malaysia (UPM), 43400 UPM SERDANG, Selangor, Malaysia
pouyanes@yahoo.com

Abstract. Like other industries, the utilization of the internet and Information Technology (IT) has increased in the health sector. Different applications attributed to the internet and IT in healthcare practice. It includes a range of services that intersect the edge of medicine, computer and information science. The presence of the internet helps healthcare practice with the use of electronic processes and communication. Also, health IT (HIT) deals with the devices, clinical guidelines and methods required to improve the management of information in healthcare. Although the internet and HIT has been considered as an influential means to enhance health care delivery, it is completely naive to imagine all new tools and mechanisms supported by the internet and HIT systems are simply adopted and used by all organizational members. As healthcare professionals play an important role in the healthcare sector, there is no doubt that mechanism of newly introduced HIT and new application of the internet in medical practice should be coupled with healthcare professionals' acceptance. Therefore, with great resistance by healthcare professionals new mechanism and tools supported by IT and the internet cannot be used properly and subsequently may not improve the quality of medical care services. However, factors affecting the healthcare professionals' adoption behavior concerning new e-health and HIT mechanism are still not conclusively identified. This research (as a theoretical study) tries to propose the source of resistance in order to handle the challenges over new e-technology in the health industry. This study uses the involved concepts and develops a conceptual framework to improve overall acceptance of e-health and HIT by healthcare professionals.

Keywords: E-health, Health Information Technology, Technology Acceptance Model, Level of interactivity, Computer self-efficacy, Future development.

1 Introduction

Organizations are becoming bigger and more complex everyday and at the same time information technology (IT) has undertaken a very vital role within the organizations [36]. A broad range of IT applications has been developed in service business. For instance, technological advances are facilitating the tasks which conducted by

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 23–37, 2010.

individuals every day ([43], [63], [69]). Currently, the prevalent role of technology is affecting the design of businesses to come across new ways of delivering services to customers [64]. The last few years have been witness for a rapid development of information and communication technologies in medical practice. According to Istepanian [51], the number of established telemedicine programs have increased all over the world. According to Bates et al. [15], the health industry has used IT to reduce the frequency of errors in medical practice. As the share of health care in GDP is gradually increasing, the health sector is trying to utilize the internet and IT to enhance health care services. On the other hand, as the cost of health care keeps increasing, the health industry concentrates on the application of HIT and the internet-supported mechanism more than before. The target of IT utilization in the design of medical practice is to reconstruct health care for the 21st century [40]. As stated by Andrews [9], medical care practices are undergoing massive changes by using different types of technology (HIT and e-health). The health sector and medical practice is one of the deep-rooted professions and IT (in different forms) has penetrated into professional setting to support highly specialized tasks [24].Also the health sector is considered as an individualized profession, relatively independent and entrepreneurial [45].

According to Pain et al. [59], latest technologies can help companies improve competitiveness in their turbulent business environment only when they are fully integrated into companies' practice patterns. As mentioned by Mathieson, [55] and chang et al [22], although a new technology promises potential merits, it cannot be effective for organizations if remains unused. However, in the health sector, evidence shows that healthcare professionals have not fully adopted HIT ([74], [83], [25], [35], [57], [81], [82]). According to Walter and Lopez [75], only greater acceptance of healthcare professionals can lead to improving health care delivery.

Based on a body of literature, the main concern in the successful adoption of a new IT has been user acceptance ([32], [34], [33], [49], [50]). A number of IT adoption models have been developed to explain and predict user acceptance of a new IT. ([32], [34], [33], [49], [50]). The models are general and applicable for different users. From an empirical view, IT adoption models have been tested empirically among students in academic contexts, knowledge workers from different industries, customers and etc. ([3],[49], [70], [42],[61]).

Hu and colleagues have studied healthcare professionals' IT adoption behavior regarding telemedicine ([23],[47],[48]). Their findings show that healthcare professionals' behavior toward a new IT is not the same as other IT users studied by past studies. Chau and Hu [23] stated that the differences are due to distinctive characteristics of healthcare professionals such as specialized training, autonomous practices, and professional work settings. Paul and McDaniel [60] have supported that healthcare professionals behave differently toward adoption of a new IT compared to other user groups.

In this study, we have integrated theories of intention and IT adoption, the professional characteristics of healthcare professionals and interactivity theory. Integration of these three theories helps us propose a modified model explaining factors affecting healthcare professionals' intention to use future e-health and HIT. This study also contributes to knowledge by expanding the current understanding on IT adoption behavior in professional context.

2 Development of a Theoretical Framework

2.1 Theories of Intention and IT Adoption

Based on the increasingly significant impact of emerging e-service and various types of IT systems in organizations, managers are expected to track new technologies' trends and introduce new IT to organizations in order to facilitate organizational functions ([13], [20]). Although management support is essential for newly introduced IT, another concern arises at the time of IT introduction in organizations. The following challenge would be the future reaction of employees who undergo a new change in the design of their work activities. If introduction of a new IT in organizations is followed by acceptance of the potential users, the systems can be fruitful and may improve productivity [80]. As the users accept the new IT, they become more willing to change their existing work routines. Also they are more likely to integrate a new IT into the flow of their everyday work practices [75].

According to King and He [52], in recent years, much interest paid to the identification of factors cause people to accept and take advantage of systems. By taking individual intention into consideration, several studies have been conducted. The results have been the development of eight theoretical models. The eight competitive models are: Theory of Reasoned Action (TRA), Technology Acceptance Model (TAM), Motivational Model (MM), Theory of Planned Behavior (TPB), a combined theory of planned behavior/technology acceptance model (C-TAM-TPB), Model of PC utilization (MPCU), Innovation Diffusion Theory (IDT), and Social Cognitive Theory (SCT). Each of these models explains individual willingness to accept new information systems and technology [31].

According to a wide variety of literature in IT adoption behavior, technology acceptance model (TAM) is found to be the most widely accepted models in information systems researches. As indicated by Wang et al. [84], the TAM's popularity is due to its parsimony, simplicity, understandability and empirical support. Davis [33] adapted the technology acceptance model (TAM) from the theory of reasoned action (TRA) to mainly explain the potential user's behavioral intention to use a technological innovation.

The original TAM embeds two independent variables for individual intention to accept new IT. The first variable namely perceived ease of use is defined as the extent to which a person believes that using a particular system would be effortless [33]. The second one namely perceived usefulness deals with the degree to which a person believes that by using a particular system he can improve his job performance [33]. With reference to Triandis [73], there are three dimensions to determine actual behavior: intention, facilitating conditions, and habit. Among these components, intention is considered as the best predictor for actual usage (behavior) in IT adoption models [78]. Intention is concerned with the individual's motivation to perform a given behavior.

According to the TAM, users are not willing to support those types of IT that interferes with their traditional work routines, unless they perceive that use of new IT is easy as well as the new system can improve their job performance in the organizations. Also it should be mentioned that this model has been extended and modified a lot over years. The following section concentrates on the modified version in the health industry.

2.2 IT Acceptance in Healthcare Profession

TAM has been widely tailored to a wide range of technologies, cultures and users. Some studies have used TAM and its modified models in the field of health care to explain and predict healthcare professional's IT adoption behavior. For instance, Aggelidis and Chatzoglou [4] emphasized that when it comes to comparison of IT users, healthcare professionals are different from other user groups in light of their special perceptions toward the use of technology. According to Lowenhaupt [53], physicians are very slow in terms of accepting clinical information systems. Based on a body of literature, healthcare professionals are not willing to integrate new IT with their day-to-day work activities if they perceive new IT as interfering with their traditional work routines ([7],[8]). Hu et al. [48] have studied in the context of telemedicine and proposed the difference of healthcare professionals from other types of IT users with respect to IT adoption behavior. Based on their findings, healthcare professionals are more pragmatic and realistic. Moreover they place more emphasis on perceived usefulness than perceived ease of use. As noted by Anderson [7], healthcare professionals are unwilling to support and use the IT that prevents them from continuing their long-established practice patterns.

Chau and Hu [23] proposed that a set of characteristics makes healthcare professionals different in IT adoption behavior compared to other user populations. Nevertheless, far too little attention has been paid to examining whether and how such differences influence healthcare professional's acceptance of a new IT [75].

2.3 The Distinctive Characteristics of Healthcare Professionals

Investing in HIT systems and e-health can provide benefits only when these systems are used by healthcare professionals. Therefore, healthcare professional adoption has long been identified as the main goal that is sought after for its great significance in support of these systems. Nevertheless, the literature indicates that healthcare professionals' adoption of new IT systems is known to be different from other IT users and they respond to IT differently [60]. Thus, this evidence made a motivation for researchers to identify the source of this difference.

Based on a rich body of literature, professionals (such as healthcare professionals) have some unique and professional characteristics and they are considered different from other non-professionals based on these unique characteristics. In this study, the special characteristics of healthcare professionals are put at the center of attention. As indicated by Brennan and Coles [16], healthcare professional's professionalism has long been based on a set of values. The most important characteristic has been healthcare professional autonomy and the other ones are patient sovereignty, healthcare professional confidentiality, and the habits of learning. According to an exploratory study by Chau and Hu [23], some special characteristics are attributed to healthcare professionals. Three proposed characteristics in this study are:

(a). Specialized training that reveals their mastery over medical knowledge which has been obtained during a lengthy period. As stated by Watts [76], they devote a considerable portion of their youth preparing for the profession. Their body of knowledge is directly link to the lives of patients. In this profession even a slightest mistake can be fatal.

(b). The second characteristic is professional autonomy. Based on the autonomous practice, healthcare professionals proclaim that they are in the best position to run, organize, and regulate their own practice. They are judged mainly through a peer review process in which professionals evaluate each other. As mentioned by Zuger [85], professional autonomy is clearly the most important value. This provided healthcare professionals with a sense of pride, and accomplishment.

(c). With reference to Watts [76], and Montague et al. [56], the third characteristic is professional work arrangements where healthcare professionals become health care providers, hospitals became health care facilities, and patient acts as both the product and the client in such a system. Based on dominating over medical knowledge, three occupational groups are available in a healthcare organization like a hospital. The healthcare professionals that include all the specialty doctors involve in treatment of patients. The para-professional group, such as medical assistants, possesses only partial professional competence and involve in assisting the healthcare professionals in performing their healthcare practices. And the last group is non-professionals who are just trained to take on clerical and office work to run a hospital's administrative affairs.

According to Abbott [1], being members of a profession is certainly conducive to professional autonomy. Based on Walter and Lopez [75], it seems that professional autonomy is viewed as a precious privilege in the hands of professionals so that they don't want to lose it in their workplace. Professionals have power over non-professional and paraprofessionals and can control the tasks carried out by them [38]. It should be added that the advantage of having control over subordinate groups is more pronounced in those organizations with existing hierarchies among various working groups. In this study, the most important characteristics of healthcare professionals, namely professional autonomy is considered as a construct that can affect healthcare professional's IT adoption behavior.

2.4 The Different Types of E-Health and HIT in the Professional Context

With reference to medical literature, e-health is a general term includes the intersection of healthcare practices, the internet and communication. It covers a wide range of services designed for medicine using IT. These services are telemedicine, m-health, knowledge management and so on. Also HIT deals with information science, computer science, clinical guidelines, and health care. The main purpose of HIT is to optimize the use of information in the health industry as well as assisting healthcare professionals in medical decision making. Two main types of HIT are Electronic Medical Report (EMR) systems, and Clinical Decision Support (CDS) systems. The main focus of this study has been placed on three IT systems available in the health industry: telemedicine, EMR and CDS.

1. Telemedicine systems make use of medical information which is transferred via electronic communications from one place to another, in order to make better health care services. For instance, in a form of telemedicine, videoconferencing is used to provide a consultation between medical specialists in different parts of word. According to Sood et al. [68], telemedicine is a developing tool in medical discipline

which can recover issues of irregular care planning, unorganized labor allocation as well as infrastructure deficiencies.

2. EMR systems are computer systems that allow a healthcare professional to create, store, and retrieve patient charts on a computer. So, these systems facilitate the rapid retrieval of information related to patients' problem lists, allergies, and socio-demographic data. Concisely, EMR is regarded as one health information technology (HIT) product that is able to improve health care practice through increasing the quality and efficiency of results ([21], [66], [71], [72]).

3. CDS systems are computer systems that provide professional advice. These systems are knowledge-based systems that use patient data and series of reasoning techniques to generate diagnostic and treatment options and care planning [75]. Pain et al. [59] have defined CDS as a system that helps healthcare professionals by providing some medical options with correct dosage and minimum possible side effects.

2.5 The Effect of E-Health and HIT on Healthcare Professional's Behavior

As mentioned previously, professional autonomy is very valuable for healthcare professionals. Furthermore, healthcare professionals support factors that protect their professional autonomy and react toward the elements may invalidate their professional autonomy and traditional work practice [75]. Not only healthcare professional's IT adoption behavior is influenced by their special characteristics, but also their IT acceptance is affected by organizational context as well as features and characteristics of HIT systems that could change their long-established work activities ([47], [23]).

According to Varonen et al. [77] and Sittig et al. [67], the culture of medicine practice has always given emphasis to physician autonomy. Invalidating the autonomy causes that system changes are not always well-received by healthcare professionals and becomes one of the biggest challenges for HIT implementation. Also, concerns about overreliance on the HIT and e-health equipments, makes healthcare professionals worry about losing their autonomy. Moreover, feature and nature of instructions, guidelines and recommendations that given by HIT in terms of problem-solving process can change healthcare professionals' work routines and may be considered as a factor that erodes their professional autonomy [18].These rules, procedures and recommendations designed in IT can weaken healthcare professionals' claim on possession of special competence in problem solving and invalidate their decision making skills in terms of deciding what the best thing to do for treatment of patients. As stated by Harrison et al. [44], healthcare professionals feel uncomfortable when they face regulations and instructions generated by a clinical decision system when advising them what to do. Because they believe that they are able to treat their patients based on their specialized knowledge, experience, skills and competence. According to Lowenhaupt [53], healthcare professionals become more anxious when someone or something (such as a computer system) shows he/it has more knowledge than them regarding treatment of their patients. Therefore, healthcare professionals perceive HIT as threatening to their professional autonomy.

2.6 Healthcare Professionals' Perceived Level of Interactivity

As mentioned earlier, the antecedent of healthcare professionals' perceived threat to professional autonomy is the rules, instructions and diagnostic options provided by the clinical IT. A general view toward new computerized systems is that any IT systems can reduce dependence on specific personnel and experts ([17], [58]). In medical context, new HIT and e-health can violate the healthcare professionals' professional autonomy and this can have an adverse effect on the acceptance behavior.

One feature of clinical information systems that influences perceived control is their level of interactivity. Perceived interaction is defined as the level of interactivity that a user perceives while working with the computerized system, and the extent to which the system is perceived to be responsive to the user's needs. Three levels of interactivity can be perceived from a medical technology [86]. At the first level, healthcare professionals use the technology as a means to generate data so the experts can make a diagnostic decision. So at this level of interaction, the medical IT can be considered as an enabler. At the second level, the technology is more complicated and acts as a partner with the professionals. At this level, the tasks of both healthcare professionals and technology have the same weight. At the third level, the role of healthcare professionals is summarized as an operator of the technology. At the third level, the technology takes on decision making process and recommend course of action and the operators just control the process. Thus, different levels of interactivity with IT systems are conducive to different perception toward using that system.

Interactivity theory applies to both the technologies and the user relationships with technology. McMillan [87] has referred to product interactivity as a type of user-to-system interaction. Perceived level of interactivity is largely based on the belief that the interactive nature of the clinical system can assist in creating cooperation between the healthcare professional and the IT system. If healthcare professionals perceive that the nature of new HIT and e-health is interactive, they perceive more control and involvement in the process of decision making. Consequently they perceive less threat from new HIT and e-health to their professional autonomy.

To sum up, the main goal of new HIT and e-health is to interact with healthcare professionals and provide them with consultation and assist them in making the best care planning and diagnosis analysis. In this human-machine interaction, both the healthcare professional's knowledge and the HIT and e-health function are required to better analyze the patients' data rather than relying on either human or the systems. In the interactive relationship between HIT and health care professionals, the system doesn't make decisions for healthcare professionals telling them what to do. Also, the process of interaction with HIT and e-health can be perceived more interactive when the possibility of adapting and customizing the system is possible in case of a patient.

Some studies have shown that healthcare professionals are unwilling to use computer systems for two main reasons; first for the fear of the new applications and mechanism of systems and second for the fear of bringing changes in their core care practices ([6], [2], [26], [10]). Gagnon et al. [41] have explained that the behavior of healthcare professionals to accept a new technology (like telemedicine) can be affected by their comparatively low computer literacy, which rooted in their established work routines, and the high professional autonomy they possess.

2.7 Computer Self-efficacy

Social cognitive theory implies two main cognitive factors affecting individuals' behavior: self-efficacy and outcome expectations [29]. One application of social cognitive theory that developed by Bandura [11], is in the context of IT usage. Based on the social cognitive theory, the perceptions held by people about their efficacy impact their outcome expectation. Thus, a person with high level of self-efficacy is more likely to envision success picture. The positive relationship between self-efficacy and outcome expectation are applied in the context of using computer, knowledge sharing and internet usage ([29], [46]). Self-efficacy refers to the belief that a person holds about his ability to do a particular job [11]. Conner and Armitage [30] have defined self-efficacy as an individual's confidence to conduct the behavior in question. With high level of self-efficacy a person believes that issues over the behavior can be managed easily.

More specific, computer self-efficacy is defined as the confidence of individuals in their own ability to work with a computer or any specific programs [27]. There is enough evidence that computer self-efficacy affects computer use, through its effect on relieving users' computer anxiety [54]. Pain et al. [59] have indicated that the reason for the low acceptance of HIT systems by doctors is the fault and malfunction of applying the system. As stated by Gagnon et al. [41], physicians' decision making to accept a new technology (such as telemedicine) can be problematic due to their relatively low computer literacy. In other words, computer literacy generally plays a powerful weapon in the interaction between the IT systems and individuals and in the health industry can help healthcare professionals interact with HIT and e-health systems. Although healthcare professionals are professional in the medical practice, they are in short of knowledge on how to work well with the applications of HIT and e-health. Computer self-efficacy helps healthcare professionals better know the codes, rules, regulations and mechanisms of new HIT and e-health. As a result, healthcare professionals may not be likely to use HIT and e-health which threaten their professional autonomy, yet this fear may be moderated by their computer self-efficacy.

3 Conceptual Framework and Propositions Development

3.1 Perceived Threat to Professional Autonomy vs. Intention to Use IT

As shown in Fig. 1, intention to use new technology is the dependent variable which refers to individual willingness to accept a new technology [34]. This study proposes that if healthcare professionals believe that using new HIT and e-health would decrease their control over the processes and content of their, they become less likely to use the new systems. Therefore, the first proposition is as follows:

Proposition1. There is a negative relationship between healthcare professionals' perceived threat to professional autonomy and their intention to use new HIT and e-health.

3.2 Interactivity Perception vs. Perceived Threat to Professional Autonomy

Based on the interactivity theory which explains human – computer perceived interaction; a high level of interactivity can be manifested by immediate, reactive and continuous exchange of information between the users and system [88]. A higher perceived level of interactivity with HIT and e-health would lead to high degree of control that healthcare professionals perceive during an interaction with the system. It consequently would result in the less perceived threat to their professional autonomy and finally healthcare professionals become more likely to use the new HIT and e-health. Therefore, the propositions are presented as follows:

Proposition2. There is a negative relationship between healthcare professionals' interactivity perception and their perceived threat to professional autonomy.

Proposition3. Perceived threat to professional autonomy mediates the relationship between healthcare professionals' interactivity perception with HIT as well as e-health and their intention to use the new system.

3.3 The Moderating Role of Computer Self-efficacy

In this study, Computer self-efficacy referees to "an individual judgment of one's capability to use a computer" ([28], p.192). Computer literacy can alleviate the anxiety in the interaction between human and applications of computerized systems. It also helps users interact easily with the systems. Therefore, when healthcare professionals believe in their computer knowledge, skills and ability in running clinical guidelines, rules and regulations, they perceive more control over the interaction with new HIT and e-health. As a result, their threat perceived from the guidelines and procedures given by new HIT and e-health would be moderated and they become more willing to use the new HIT and e-health. Thus, the next proposition is:

Proposition4. Healthcare professionals' computer self-efficacy moderates the relationship between their perceived threat to professional autonomy and intention to use new HIT and e-health.

3.4 Perceived Usefulness and Intention to Use IT

Based on the literature, perceived usefulness is found as one of the strongest predictors of intention to use new IT [80]. The TAM embeds this variable as "the degree to which a person believes that using a particular system would enhance his or her job performance" [34]. Thus, as far as new HIT and e-health mechanisms are perceived as useful tools, healthcare professionals are willing to accept and use them. Therefore, the next proposition is as follows:

Proposition5. There is a positive relationship between healthcare professionals' perceived usefulness and their intention to use new HIT and e-health.

Also, it is proposed that healthcare professionals may not use new HIT and e-health which encroaches on their professional autonomy, even though the new technology is perceived useful. So, the respective proposition is as follows:

Proposition6. There is a negative relationship between healthcare professionals' perceived threat to professional autonomy and perceived usefulness.

Furthermore, perceived threat to professional autonomy is proposed to leave its mark on intention to use new HIT and e-health through perceived usefulness. Thus, the next proposition is made as follows:

Proposition7. Perceived usefulness partially mediates the relationship between perceived threat to professional autonomy and intention to use new HIT and e-health.

3.5 Perceived Ease of Use and Perceived Usefulness

The TAM indicates the significance of perceived ease of use as another key construct explaining user's IT adoption behavior. This construct deals with the amount of effort users should make to use new systems. Chau et al. [24] have indicated that in medical context, healthcare professionals are more concerned with the usefulness of new HIT rather than how much it is easy to use. Therefore, we propose that perceived ease of use doesn't affect healthcare professionals' intention to use new HIT and e-health and can only have a direct impact on perceived usefulness. As a result, the proposition is presented as follows:

Proposition8. There is a positive relationship between healthcare professionals' perceived ease of use and their perception about the usefulness of new HIT and e-health.

Based on the propositions mentioned above, the conceptual framework of this study is depicted in the following figure (Fig.1):

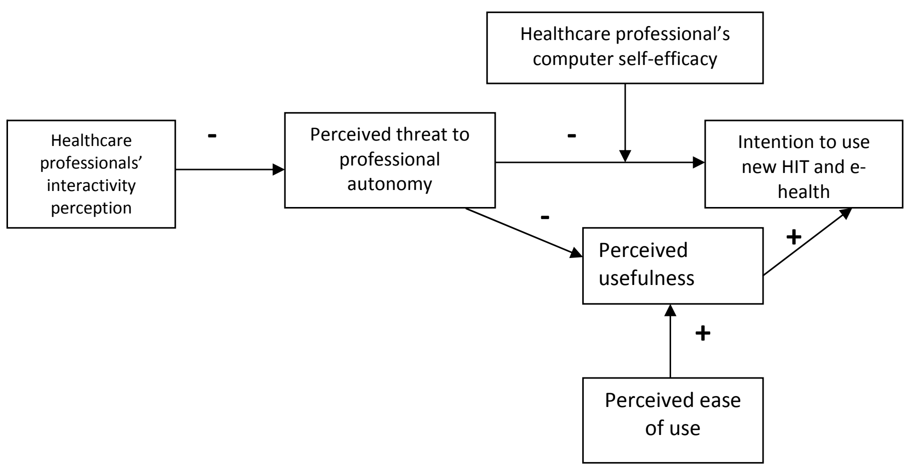


Fig. 1. Proposed Conceptual framework

4 Implications of the Study

4.1 Theoretical

The design of health industry practices has been transformed by new HIT and e-technologies. However, challenges have still arisen over the introduction of these systems. The evidence shows gaps in the fulfillment of HIT and e-health full potentials due to issues in the adoption of these systems by healthcare professionals. The theoretical contribution of this study is to cover this unmet need by integrating the TAM into new concepts and theories that have roots in healthcare professionals' characteristics and interactivity with the new systems. Therefore, the research adds to the body of IT adoption knowledge and extends the existing understanding in the field of IT adoption behavior in the health care context.

4.2 Practical

This study would also propose some practical implications for IT practitioners and hospital management. The study helps practitioner design future HIT and e-health systems to be acceptable for health care professionals in order to improve overall acceptance rate. Also, this study proposes future strategic policies for hospital management to decrease negative impact of perceived threat to professional autonomy. As a result, this study proposes how to develop future interactive HIS and e-health and how management practices should be design to improve healthcare professional's computer self-efficacy by offering training programs.

References

1. Abbott, A.: The System of Professions: An Essay on the Division of Expert Labor. University of Chicago Press, Chicago (1998)
2. Nielsen, A.C.: Attitudes Towards IT in Australian General Practice: Qualitative Research Report, AcNielsen PTY General Practice Branch, North Sydney (1998)
3. Adams, D.A., Nelson, R.R., Todd, P.A.: Perceived usefulness, ease of use, and usage of information technology: a replication. MIS Quarterly 16, 227–247 (1992)
4. Aggelidis, V.P., Chatzoglou, P.D.: Using a modified technology acceptance model in hospitals. International Journal of Medical Informatics 78, 5–126 (2009)
5. Ajzen, I.: Attitudes, Personality and Behavior. Open University Press, Milton Keynes (1988)
6. Anderson, J.: Increasing acceptance of clinical information systems. MD Computing 16, 62–65 (1999)
7. Anderson, J.: Clearing the way for physicians' use of clinical information systems. Communications of the ACM 40, 83–90 (1997)
8. Anderson, J.G., Aydin, C.E.: Evaluating the impact of health care information systems. International Journal of Technical Assessment in Health Care 13, 380–393 (1997)
9. Andrews, M.: The Hospital of the Future. U.S. News & World Report, 00415537, vol.146 (2009)
10. Baldwin, G.: Automating patient records. Technology in Practice, http://www.technologyinpractice.com
11. Bandura, A.: Self-Efficacy: The Exercise of Control. Freeman, New York (1997)
12. Bandura, A.: Social Foundations of Thought and Action: A Social Cognitive Theory. Prentice Hall, Englewood Cliffs (1986)
13. Basselier, G., Benbasat, I., Reich, B.: The influence of business managers' IT competence. Information Systems Research 14, 317–336 (2003)

14. Bates, D.W., Cohen, M., Leape, L.L., Overhage, J.M., Shabot, M.M., Sheridan, T.: Reducing the frequency of errors in medicine using information technology. Journal of the American Medical Informatics Association 8, 299–308 (2001)
15. Berner, E.S.: Clinical decision support systems: State of the Art. AHRQ Publication No.09-0069-EF. Agency for Healthcare Research and Quality, Rockville (2009)
16. Bonora, E.A., Revang, O.: A strategic framework for analyzing professional service firm _ developing strategies for sustained performance. In: Strategic Management Society InterOrganizational Conference, Toronto, Canada (1991)
17. Borkowski, N.M., Allen, W.R.: Does attribution theory explain physicians' nonacceptance of clinical practice guidelines? Hospital Topics: Research and Perspectives on Healthcare 81, 9–21 (2003)
18. Brennan, M., Coles, C.: Developing professional skills. The Lancet 362(9394), 362–1506 (2003)
19. Brown, S.A., Chervany, N.L., Reinicke, B.A.: What matters when introducing new information technology. Communications of the ACM 50, 91–96 (2007)
20. Burt, C.W., Hing, E.: Use of computerized clinical support systems in medical settings: United States. Advance Data 353, 1–8 (2001-2003)
21. Chang, J.Y., Chen, L.K., Chang, C.C.: Perspectives and expectations for telemedicine opportunities from families of nursing home residents and caregivers in nursing homes. International Journal of Medical Informatics 78, 494–502 (2009)
22. Chau, P.Y.K., Hu, P.J.: Examining a model of information technology acceptance by individual professionals: an exploratory study. Journal of Management Information Systems 18, 191–229 (2002)
23. Chau, P.Y.K., Hu, P.J.: Investigating healthcare professionals' decision to accept telemedicine technology: an empirical test of competing theories. Information and Management 39, 297–311 (2002)
24. Chismar, W.G., Wiley-Patton, S.: Does the extended technology acceptance model apply to physicians. In: Proceedings of the 36th Hawaii International Conference on System Sciences (2003)
25. Colliver, V.: Digital diagnosis: study could determine if online consultations will fly, San Francisco Chronicle, `http://www.mylifepath.com/bsc/newsroom/pub/`
26. Compeau, D.R., Higgins, C.A.: Computer Self-Efficacy: Development of a Measure and Initial Test. MIS Quarterly 19, 189–211 (1995b)
27. Compeau, D.R., Higgins, C.A., Huff, S.: Social Cognitive Theory and Individual Reactions to Computing Technology: A Longitudinal Study. MIS Quarterly 23, 145–158 (1999)
28. Conner, M., Armitage, C.J.: Extending the theory of planned behavior: A review and avenues for further research. Journal of Applied Social Psychology 28, 1429–1464 (1998)
29. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. Journal of Applied Social Psychology 22, 1111–1132 (1992)
30. Davis, F.D.: User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. International Journal of Man-Machine Studies 38, 475–487 (1993)
31. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User acceptance of computer technology: A comparison of two theoretical models. Management Science 35, 982–1003 (1989)
32. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13(3), 319–340 (1989)
33. Dearne, K.: Health's tech bypass, in The Australian, Sydney, pp. 1–4 (2003)

34. Dewet, T., Jones, G.R.: The role of information technology in the organization: a review, model, and assessment. Journal of Management 27, 313–346 (2001)
35. Dowswell, G., Harrison, S., Wright, J.: Clinical guidelines: attitudes, information processes and culture in English primary care. International Journal of Health Planning and Management 16, 107–124 (2001)
36. Freidson, E.: Professional Dominance: The Social Structure of Medicine. Atherton Press, New York (1970)
37. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading (1975)
38. Flower, J.: Digital technology essential and subversive. The Physician Executive 30, 42–45 (2004)
39. Gagnon, M.P., Godinb, G., Gagnéb, C., Fortina, J.P., Lamothec, L., Reinharza, D., Cloutierd, A.: An adaptation of the theory of interpersonal behavior to the study of telemedicine adoption by physicians. International Journal of Medical Informatics 71, 103–115 (2003)
40. Gefen, D., Straub, D.W., Boudreau, M.: Structural equation modeling and regression: Guidelines for research practice. Communications of the AIS (2000)
41. Guriting, P., Ndubisi, O.N.: Borneo online banking: Evaluating customer perceptions and behavioral intention. Management Research News 29, 6–15 (2006)
42. Harrison, S., Dowswell, G., Wright, J.: Practice nurses and clinical guidelines in a changing primary care context: an empirical study. Journal of Advanced Nursing 39, 299–307 (2002)
43. Holsinger, J., Beaton, B.: Physician professionalism for a new century. Clinical Autonomy 19, 473–479 (2006)
44. Hsu, M.H., Chiu, C.M.: Internet self-efficacy and electronic service acceptance. Decision Support Systems 38, 369–381 (2004)
45. Hu, P.J., Chau, P.Y.K., Sheng, O.R., Tam, K.Y.: Examining the technology acceptance model using physician acceptance of telemedicine technology. Journal of Management Information Systems 16, 91–112 (1999)
46. Hu, P.J., Chau, P.Y.K., Sheng, O.R.: Adoption of telemedicine technology by health care organizations: an exploratory study. Journal of Organizational Computing and Electronic Commerce 12, 197–221 (2002)
47. Igbaria, M.: User acceptance of microcomputer technology: an empirical test. Omega 21, 73–90 (1993)
48. Igbaria, M., Guimaraes, T., Davis, G.: Testing the determinants of microcomputer usage via a structural equation model. Journal of Management Information Systems 11, 87–114 (1995)
49. Istepanian, R.S.H.: Telemedicine in the United Kingdom: current status and future prospects. IEEE Transactions on Information Technology in Biomedicine 3, 158–159 (1999)
50. King, W.R., He, J.: A meta-analysis of the technology acceptance model. Information and Management 43, 740–755 (2006)
51. Lowenhaupt, M.: Removing barriers to technology. The Physician Executive 30, 12–14 (2004)
52. Marakas, G.M., Yi, M.Y., Johnson, R.D.: The Multilevel and Multifaceted Character of Computer Selfefficacy: Toward Clarification of the Construct and an Integrative Framework for Research. Information Systems Research 9, 126–163 (1998)
53. Mathieson, K.: Predicting user intentions: comparing the technology acceptance model with the theory of planned behaviour. Information Systems Research 2, 173–191 (1991)

54. Montague, E.N.H., Kleiner, B.M., Winchester, W.W.: Empirically understanding trust in medical technology. International Journal of Industrial Ergonomics 39, 628–634 (2009)
55. Murray, D.: Healthcare challenge. Australian Information Week 3, 10–18 (2002)
56. Nonaka, I., Takeuchi, H.: The Knowledge –Creating Company. Oxford University Press, New York (1995)
57. Pain, D., Fielden, K., Shibl, R.A.: Options on the use of clinical decision support systems for paediatric prescribing in a New Zealand hospital. Logistics Information Management 16, 201–206 (2003)
58. Paul, D.L., McDaniel Jr., R.R.: A field study of the effect of interpersonal trust on virtual collaborative relationship performance. MIS Quarterly 28, 183–227 (2004)
59. Purao, S., Storey, V.C.: Evaluating the adoption potential of design science efforts: The case of APSARA. Decision Support Systems 44, 369–381 (2008)
60. Scott, R.E.: E-Records in health—preserving our future. International Journal of Medical Informatics 76, 427–431 (2007)
61. Ratnasingam, P., Gefen, D., Pavlou, P.A.: The role of facilitating conditions and institutional trust in electronic marketplaces. Journal of Electronic Commerce 48, 69–82 (2005)
62. Reid, M., Levy, Y.: Integrating Trust and Computer Self-Efficacy with TAM: An Empirical Assessment of Customers' Acceptance of Banking Information Systems (BIS) in Jamaica. Journal of Internet Banking & Commerce 13 (2008)
63. Rogers, E.: Diffusion of Innovations, 4th edn. The Free Press, New York (1995)
64. Shortliffe, E.H.: The evolution of electronic medical records. Academic Medicine 74, 414–419 (1999)
65. Sittig, D., Krall, M., Dykstra, R., et al.: A survey of factors affecting clinician acceptance of clinical decision support. BMC Med. Inform. Decis. Mak. 6 (2006)
66. Sood, S., Mbarika, V., Jugoo, S., Dookhy, R., Doarn, C.R., Prakash, N., Merrell, R.C.: What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed. e-Health 13, 573–590 (2007)
67. Srinivasan, S.: Role of trust in e-business success. Information Management & Computer Security 12, 66–72 (2004)
68. Szajna, B.: Empirical evaluation of the revised technology acceptance model. Management Science 42, 85–92 (1996)
69. Thompson, T.J., Brailer, D.G.: The decade of health information technology: delivering consumer-centric and information-rich health care. In: Framework for Strategic Action, Office of the National Coordinator for Health Information Technology. Department of Health and Human Services, Washington (2004)
70. Tierney, W.M., Overhage, J.M., McDonald, C.J.: Demonstrating the effects of an IAIMS on health care quality and cost. Journal of the American Medical Informatics Association 4, 41–46 (1997)
71. Triandis, H.C.: Interpersonal Behavior. Brooke/Cole, Monterey (1977)
72. Tung, F.C., Chang, S.C., Chou, C.M.: An extension of trust and TAM model with IDT in the adoption of the electronic logistics information system in HIS in the medical industry. Int. J. Med. Inform. 77, 324–335 (2008)
73. Walter, Z., Lopez, M.S.: Physicians acceptance of information technology: Role of perceived threat to professional autonomy. Decision Support Systems 64, 206–215 (2008)
74. Watts, C.: Erosion of healthcare professional autonomy and public respect for the profession. Surgical Neurology 71, 269–273 (2008)

75. Varonen, H., Kortteisto, T., Kaila, M.: What help or hinder the implementation of computerized decision support systems (CDSSs): a focus group study with physicians. Fam. Pract. 25, 162–167 (2008)
76. Venkatesh, V., Davis, F.D.: A model of the perceived ease of use: Development and test. Decision Sciences 27, 451–481 (1996)
77. Venkatesh, V., Davis, F.: A theoretical extension of the technology acceptance model: Four longitudinal field studies. Management Science 46, 186–204 (2000)
78. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.: User acceptance of information technology: toward a unified view. MIS Quarterly 7, 425–478 (2003)
79. Wenn, A., Tantnall, A., Sellitto, C., Darbyshire, P., Burgess, S.: A sociotechnical investigation of factors affecting IT adoption by rural GPs'. In: Information Technology in Regional Areas, Using IT: Make IT happen, Online, Rockhampton, Queensland Australia (2002)
80. Western, M., Dwan, K., Makkai, T., del Mar, C., Western, J.: Measuring IT use in Australian General Practice', University of Queensland, Australia (2001)
81. Yi, M.Y., Jackson, J.D., Park, J.S., Probst, J.C.: Understanding information technology acceptance by individual professionals: Toward an integrative view. Information and Management 43, 350–363 (2006)
82. Wang, Y.S., Wang, Y.M., Lin, H.H., Tang, T.I.: Determinants of user acceptance of Internet banking: an empirical study. International Journal of Service Industry Management 14, 501–519 (2003)
83. Zuger, A.: Dissatisfaction with medical practice. N. Engl. J. Med. 350, 69–75 (2004)
84. Kleiner, B.M.: Sociotechnical system design in health care. In: Carayon, P. (ed.) Handbook of Human Factors and Ergonomics in Health Care and Patient Safety, pp. 79–94. Lawrence Erbaum Associates Inc., Mahwah (2006)
85. McMillan, S.J.: Exploring models of interactivity from multiple research traditions: Users, documents, and systems. In: Lievrouw, L., Livingston, S. (eds.) Handbook of New Media, pp. 163–182. Sage, London (2002)
86. Zack, M.H.: Interactivity and communication mode choices in ongoing management groups. Information System Research 4, 207–239 (1993)

The Study of Address Tree Coding Based on the Maximum Matching Algorithm in Courier Business*

Shumin Zhou[1], Bin Tang[1], and Wen Li[2]

[1] Engineering Research Center of Nuclear Technology Application (East China Institute of Technology), Ministry of Education, 330013, Nanchang, Jiangxi, China
[2] School of Information Engineering, East China Institute of Technology, 344000, Fuzhou, Jiangxi, China
smzhou@ecit.edu.cn

Abstract. As an important component of EMS monitoring system, address is different from user name with great uncertainty because there are many ways to represent it. Therefore, address standardization is a difficult task. Address tree coding has been trying to resolve that issue for many years. Zip code, as its most widely used algorithm, can only subdivide the address down to a designated post office, not the recipients' address. This problem needs artificial identification method to be accurately delivered. This paper puts forward a new encoding algorithm of the address tree - the maximum matching algorithm to solve the problem. This algorithm combines the characteristics of the address tree and the best matching theory, and brings in the associated layers of tree nodes to improve the matching efficiency. Taking the variability of address into account, the thesaurus of address tree should be updated timely by increasing new nodes automatically through intelligent tools.

Keywords: Address Tree, Maximum Matching, Courier Business, Address Node.

1 Introduction

The address tree coding method has been uses for three decades in EMS mail management in mainland China. Its most widely used algorithm is zip code that consists of six Arabic numerals [1]. The numerals are divided into four grades which separately represent the provinces (autonomous regions or municipalities), postal delivery zones, cities and post offices. Zip code can only subdivide the address down to a designated post office, so that it is impossible to deliver to the correct address without artificial identification [2].

Therefore, the primary question is how to encode address tree to facilitate automatic delivery to replace artificial operating. In this paper, we first set up a reasonable model of address tree, and construct the address parser based on the maximum matching algorithm to analyze and standardize the address to be delivered according to the address tree model, at last establish an appropriate mechanism to adjust and expand the model continuously.

* This research is partially supported by National High-tech R&D Program of China (863 Program: 2007AA06Z111), JingGang Star Talents Training Plan of Jiangxi Province, China.

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 38–45, 2010.

2 The Address Tree Modeling

2.1 The Rules of Address Tree Modeling

The address that can be read and delivered should include the following three elements: the correct address name, explicit subordination relationship, and accurate address spot. Tree model meets these requirements on the foundation of rigorous validation in detail [3].

The address tree model is defined as: using a multi-tree mathematical model to describe the subordination relations of address, which is described by tree nodes and their relations. The former is used to indicate a specific address or a region spot, and the latter defines the relationship between the addresses.

The model describes a specified address by linking the name of nodes from the root down to the leaf layer by lay. Tree trunk is taken as delivery route, and leaf is considered as delivery spot. Every spot is arranged a 16-bit digital code as the unique identity.

Address tree is composed by the following types of nodes:

(1) Node: a specific physical address name;
(2) Root node: the widest area which the address tree model can revolve;
(3) Parent node: address name of the upper layer from the local one;
(4) Child node: address name of the lower layer from the local one;
(5) The node with homonymous name: one of them is taken as official name, the other are showed in the subsequent brackets.

The example of address tree model is displayed in Fig.1.

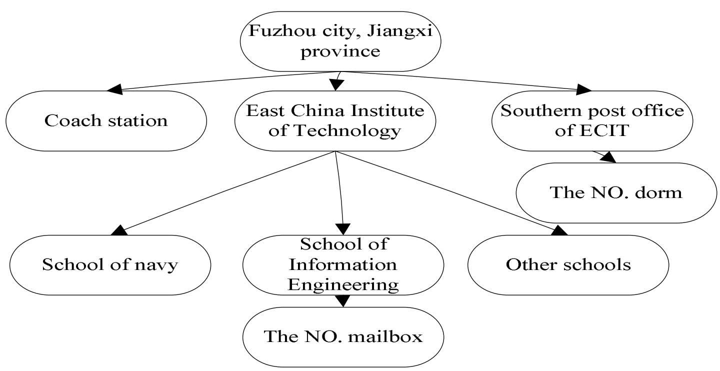


Fig. 1. The example of the address tree model which is made up of four layers and nine nodes describes some sites of Fuzhou City in Jiangxi Provence

2.2 The Establishment of Address-Based Tree

The leaf nodes or the branches close to those will take up a significant portion of the address actually [4]. The main trunk of address tree such as regions, roads, and streets

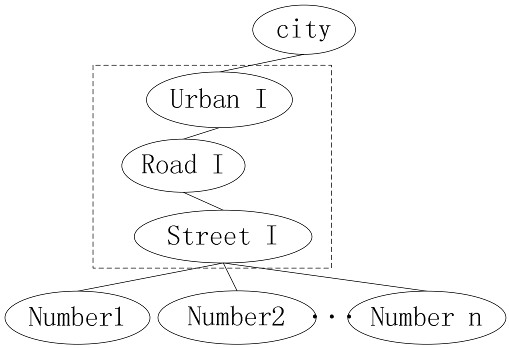


Fig. 2. The main trunk 'Urban I→ Road I→ Street I' is extracted to be an address-based tree

will only account for little [5-6]. The main trunk is thus extracted to be an address-based tree which is shown in figure 2. This approach can save the storage space and speed up the searching for address.

2.3 The Foundation and Maintenance of Address Tree

There are two ways to found and maintain the address tree:

(1) It is better to design address model questionnaires according to the features of local address, then hand out them to the delivery sectors that can answer the assigned questions, and collect and dispose them to ensure that they can enter into the system, finally set up the address tree model to improve thesaurus information of address.

(2)The other way is to expand new address in database on the basis of the approval made by municipal government which can deal with the variability of new road, living area, and region.

3 The Basic Analysis Algorithm of Address Tree

3.1 The Standardization of Address Name

The basic analysis algorithm of address tree is a tool which is used to help computers read the address and complete delivering. There are a variety of ways for writing address, especially in different languages. For example, uppercase and lowercase letters or the combination of the two, many synonyms, and different types of numbers and so on. It is necessary to emphasize that most of address name is made up of Chinese characters in our system. There are widely divergent formulations to express the same meaning, especially for address. For instance, 'International Hotel, road 1, Zhongshan City', 'NO.18, road 1, Zhongshan City', and 'International Hotel, Zhongshan City' represent the same address. If the three items of address name are dealt with computer, we can only get incorrect conclusion in figure 3.

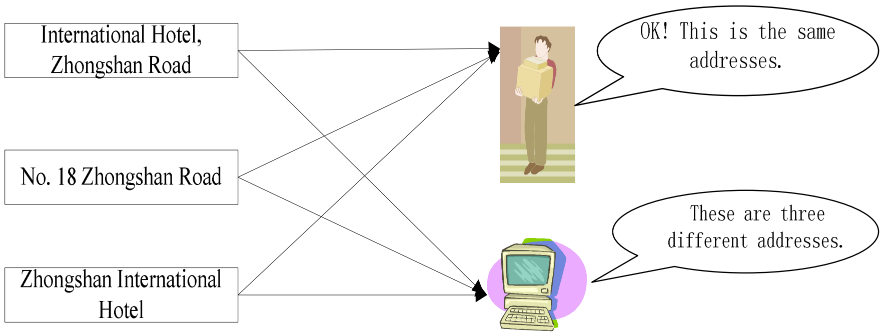


Fig. 3. The quite different conclusions between computer and artificial identifying the three addresses show the defects of computer

So the first step of analysis is that the address name should be standardized to meet the format input of the algorithm. There are five principles to achieve the objective:

(1) Change all of lowercase letters into uppercase ones;
(2) Transform Chinese numerals before keywords such as floor, building, and room into lowercase ones;
(3) Unify synonymous words to one, like building converted into block;
(4) Deal with portmanteau words in Chinese, as ‘木志’->’梽’,’子子’->’孖’;
(5) Consolidate quantifiers, like 6# converted to 6hao.

3.2 The Basic Analysis Algorithm of Address Tree

The address name should be analyzed and broken down by the basic analysis algorithm of address tree that adopts recursive thought after standardization. The algorithm variables are shown as follows: MAXL stands for the largest permitted matching number, IADD is on behalf of the initial address to be analyzed, P is the father node ID, RADD represents the address which is returned by matching, L[i] represents each node which is successfully matched, M[L[i]] means information description of the itch node, RLEN[i] is the length of the matching string , RLEV[i] is defined as the layers from the root node to the current matching one. Algorithm flow is shown in Figure 4.

The illustration of the algorithm:

(1) Introduce the method that can split the address from left to right to match the address tree nodes. For example, the address ‘Fuzhou city, linchuan street, xuefu road, 56 hao’ is split as ‘fuzhou city’, ‘linchuan street’, ‘xuefu road’, and ’56 hao’ separately to match the nodes from top to bottom;

(2) Homonymy problem: when analyzing the address, it is permitted that an address always has many homonymy which are displayed in brackets. If the keywords in the address are matched with the words in brackets, the former words should be transformed to the relevant ones without brackets. But, sometimes homonymy has another manifestation, as homonymy deficient in regional keywords. Like the city of Beijing is the same as Beijing, and each province in China has a short name;

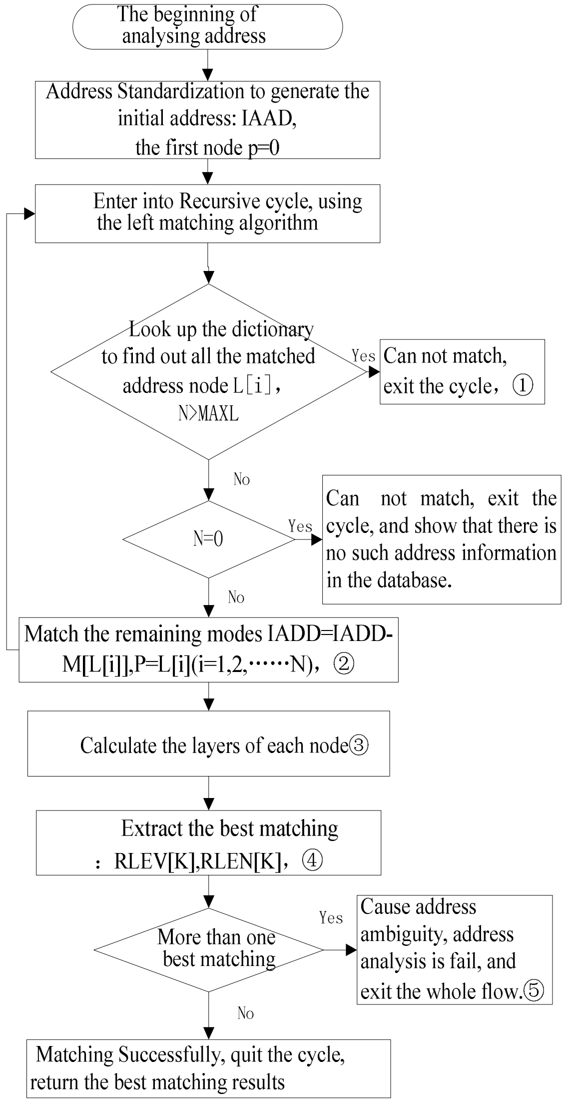


Fig. 4. The procedure of the maximum matching algorithm to analyze the address tree

(3) The best matching: a lot of items of reasonable matching may be found, and then the best matching exists. In the algorithm, the definition of the best matching principles is: Firstly the number of matched sub/Sun-rise nodes layers is the least; secondly, the length of matching is the largest;

(4) Skip the invalid characters or the non-existent nodes to go on analyzing;

(5) Ambiguity address: There are more than one matched words from different nodes in the lower layer, such as ‘city block-songyuan road-longjing street’ and ‘city block-qiwan country-longjing street’. When the address ‘city block, longjing street,

10 hao' is analyzed, the ambiguity is generated. So we cannot achieve the valid consequence.

The specification of the flow chart:

①when the number of matched address exceeds the designated count, the original address must have some problems, such as some street-some region- Beijing city, will be found dozens of coincident address;

②Export: L[1], L[2], …, L[N];

③Calculate the layer disparity between the sub node and Sun-rise node, if it is the sub-node, the Level = 1; if it is the sub-sub-node node , Level = 2, and so on;

④(RLEV[K],RLEN[K])=(Min(RLEV[i]) ,Max(RLEN[i])), i=1…N. Extract the best matching using the formulation.

⑤Ambiguity address.

4 The Expansion and Optimization of Address Tree

4.1 The Intelligent Expansion of the Address Tree

If there is no matched node in the address tree, the system will add the new node in need automatically to improve the address database. Like figure 5.

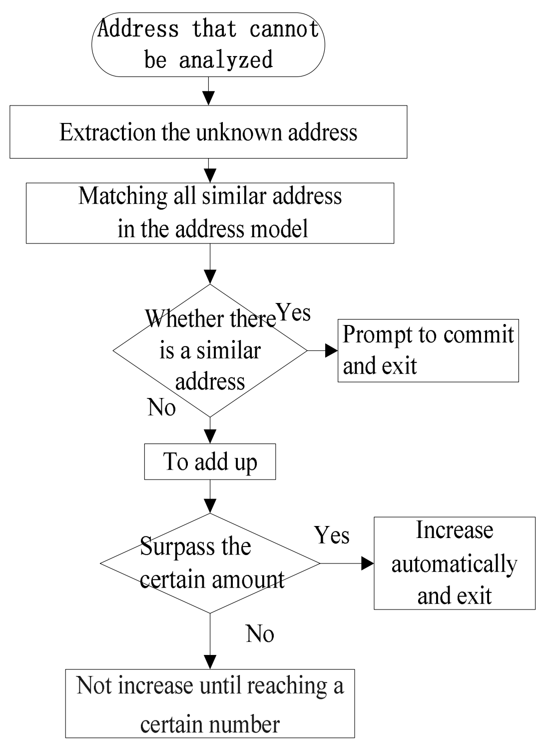


Fig. 5. The processing about whether add nodes and how to do it

To match similar address is mainly according to homophony, similar shape, and similar meaning and so on. The system automatically adds new address nodes, which need to be accumulated to a certain quantity. When some address is partially made up of the node name, it is proved that the name has high credibility. If the quantity is not enough, it needs to continue accruing.

4.2 The Layers Which the Node Can Be Associated with

Some nodes to be mentioned in the address tree algorithm are not allowed to skip the multi-layer or even any layer to match. The introduction of the concept 'the layers which the node can be associated with' to constrain the running conditions to increase the accuracy of analysis algorithm, at the same time, it speeds up the analysis process because of early prohibition to invalid address. Park I node can be added with the amount of associated layers 3 as shown in Figure 6.

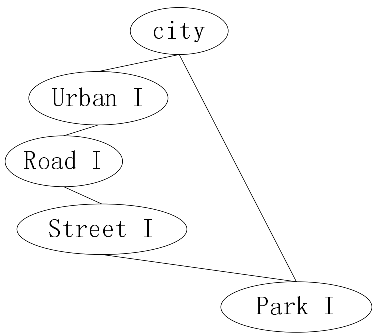


Fig. 6. Three layers includes Urban I, Road I and Street I nodes which Park I node is associated with

4.3 The Buffer of Searching

An access rate statistics is added to each node according to the principles of statistical prediction in order to accelerate the analysis speed. When the address is successfully matched, its visit frequency that is taken as the reference of weight in the following matching analysis or divided address tree is automatically added to 1. The larger value of weight is used as the preferred path; similarly, the more visit frequency of sub tree will possess greater weights.

5 Conclusions

This paper has elaborated the standardization of address name, the modeling of the address tree, put forward a corresponding analysis algorithm to match address, and then defined the layers that the node can be associated with to restrict invalid address, meanwhile, applied caching technology to accelerate searching speed. The research on the address tree coding not only considers the response speed of the server but also the accuracy of address matching, which has important referenced significance for China to realize city digitalization and cyberize.

References

1. Jie, C.Y., Liu, Y., Liang, N.Y.: On the Chinese Word Segmentation Method. Chinese Information Processing, 36–40 (1989)
2. Cheng, W., Liu, Y.J., Lu, Z.J.: Comparing the Best String-matching Algorithm Research and Application. J. Computer Engineering and Design 25(1), 1430–1432 (2004)
3. Guo, H.: An Improved Segmentation Algorithm. J. MM Micro-Computer Applications 18(2), 13–15 (2002)
4. Taylor, J., Tisseyre, B., Charnomordic, B., Guillaume, S.: A Segmentation Algorithm for the Delineation of Agricultural Management Zones. J. Computer and Electronics in Agriculture 70(3), 199–208 (2010)
5. Kim, J., Jung, H., Lee, J.: Route Optimization for Network Mobility with Tree-based Care of Address. In: 15th IEEE Asia-Pacific Conference on Communications, pp. 652–655 (2009)
6. Jiang, N., Ding, X.Q., Ren, Z.: A Suffix Tree based Handwritten Chinese Address Recognition System. In: 9th International Conference on Document Analysis and Recognition, pp. 292–296 (2007)

Customer Credit Scoring Method Based on the SVDD Classification Model with Imbalanced Dataset

Bo Tian[1], Lin Nan[2], Qin Zheng[1], and Lei Yang[3]

[1] School of Information Management & Engineering, Shanghai University of Finance and Economics, Shanghai, 200433 China
[2] School of International Business Administration, Shanghai University of Finance and Economics, Shanghai, 200433 China
[3] School of Economics and Management, Tongji University, Shanghai 200092, China
youngtb@gmail.com, yangyoungya@sina.com

Abstract. Customer credit scoring is a typical class of pattern classification problem with imbalanced dataset. A new customer credit scoring method based on the support vector domain description (SVDD) classification model was proposed in this paper. Main techniques of customer credit scoring were reviewed. The SVDD model with imbalanced dataset was analyzed and the predication method of customer credit scoring based on the SVDD model was proposed. Our experimental results confirm that our approach is effective in ranking and classifying customer credit.

Keywords: Customer Credit Scoring, Pattern Classification, Support Vector Domain Description Model, Imbalanced Dataset.

1 Introduction

Pattern classification and predication is one of main problems in statistical decision, pattern recognition and artificial intelligence, signal detection and estimation. Classical statistical methods of pattern classification mainly include Bayes statistical discriminate method, Fisher discriminate method, log-linear regression model and so on [1-3]. Numbers of samples in classical statistical methods are usually assumed to be sufficiently large. But samples usually are finite even deficient in practice. So in recent years, artificial intelligence methods such as neural-networks, clustering method, support vector machine (SVM) model which are based on finite samples become more and more popular in pattern classification field [4-6].

Customer credit scoring of banks is also called default risk prediction, which is predicting the possibility of losing that banks are suffered because the customers of banks are reluctant or unable to fulfill the credit contract [7]. Evaluation and predication of customer credit of banks is to predict the probability that customers repay loans on schedule based on all kinds of information that customers offered, and decide whether or not to approve the loan applications of the customers. Customer credit scoring is a powerful tool for the management of credit risk of banks which is a

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 46–60, 2010.

kind of pattern classification problem. In early time, the main objects of credit scoring are small and medium enterprises such as shopkeepers, mail corporations, financial firms. And the numeric graded credit scoring decision systems and statistical classification techniques are used. Individual credit card service appeared in 1960s. The numbers of individual customers and the total consumption amounts exceeded the small and medium enterprises gradually. The credit scoring decision-making processes should be performed automatically. And the development of computer and Internet techniques provided technical guarantee for the automatization of decision-making. Following, statistical methods such as Bayes discriminate method, Fisher discriminate method, and log-linear regression model were applied in the processes of individual customer credit scoring of banks [1, 3, 7]. The customers credit scoring methods based on classical statistical theories are usually under the hypothesis of asymptotic theory that the numbers of the samples were prone to infinite. And prior knowledge such as probability distributions of the samples and properties of the estimators about the samples are used. In 1990s, data-driven artificial intelligence classification methods such as neural-networks and SVM models were introduced in the analysis processes of customer credit scoring of banks [7-9]. The data-driven classification methods are based on the statistical learning theory. The disadvantages of statistical asymptotic theory can be tided over using the data-driven classification methods. The principle of the data-driven classification methods is that the decision-making function is achieved by the learning process using small or finite samples about the objects. And prior knowledge about the samples needed not to be known. Minimization of experimental risk is used in neural- networks predication models, which make the total output error be minimized. Individual credit scoring method based on the neural-networks model was investigated in reference [8]. But the practical applications of the neural-networks are limited because of several shortcomings such as over-fitting phenomenon in learning processes, lack of generalization ability, and local extremum values. Cortes and Vapnik proposed the SVM model in which the decision super-plane is constructed by minimizing of structural risk, and the complexity of models and experimental risk are balanced effectively [10-11]. The SVM models have strong generalization ability. And problems such as small number of samples, non-linear map, high dimension description, and local extremum values can also be solved. So the SVM models are very suitable to be used in pattern classification with small samples, approximation of functions and so on. Some improved SVM models were then proposed by other researchers. Least squares SVM model was proposed by Suykens [12]. Wavelet SVM model was proposed by Zhang [13]. Support vector domain description (SVDD) model was proposed by Tax [14]. Applications of the SVM models were also investigated. Individual credit scoring method based on the SVM model was investigated in reference [15]. Different costs of misclassification were considered in the cost sensitive SVM model proposed by Zadrozny [16]. Classification algorithm of SVM model with imbalanced dataset was proposed in [17].

Evaluation and predication of customer credit of banks are becoming more and more important with the development of individual credit card service in commercial banks. The object of the management of credit risk turns to maximizing the profit of commercial banks from minimizing the probability of breach of contracts [7]. Customer credit scoring of banks is a kind of pattern classification problem with imbalanced dataset because there exists obvious discrimination of sample numbers between two classes named well-record and bad-record [18]. It is subjective to determine the weights of positive and negative samples manually in the weighted SVM model in reference [17]. An improved SVM model, the SVDD model which based on dataset description method was proposed by Tax [14]. The SVDD model initially deals with the problem of one-class classification [19]. The principle of the SVDD model is to construct the hyper-sphere with minimizing the radius which contains the most of positive examples, and others samples named outliers are located outside of the hyper-sphere. The computing tasks of the SVDD model are to calculate the radius and center of the hyper-sphere by using the given samples. And the SVDD model can be used in describing the dataset and detecting outliers. The dataset are described by using the samples located at boundary of hyper-sphere in the SVDD model, which makes the SVDD model to be high computing efficient. As we analyzed, customer credit scoring of banks is a pattern classification problem with imbalanced dataset. New predication method of customer credit scoring of banks based on the SVDD model with the imbalanced dataset was proposed in this paper. And residual of the paper was organized as follows. Main techniques of customer credit scoring were reviewed in section one. The SVDD classification model with imbalanced dataset was analyzed in section two. And the multiplicative updating principle to compute the parameters of the model was also discussed in this section. Then predication method of customer credit scoring of banks based on the SVDD model was proposed in section three. Section four reported the experimental comparing results of artificial dataset and benchmark credit dataset of banks using the proposed method and the SVM- based method. Conclusions were drawn in the last section.

2 The SVDD Model with Imbalanced Dataset

An optimal closed high-dimensional hyper-sphere is established in the SVDD model as for classification problems. Positive examples are included in the hyper-sphere, and outliers are located outside of the hyper-sphere. And the SVDD model can be used in describing dataset and detecting outliers. Following the primary SVDD model with one class of samples is reviewed. And the multiplicative updating principle to compute the parameters of the model is analyzed. Then the SVDD model which containing two classes of samples was discussed, and the decision-making function based on the hyper-sphere of the SVDD model is shown. Kernel transformation in feature space is also mentioned in this section.

2.1 The SVDD Model to Describe One Class of Samples

Let dataset $\{x_i, i = 1, 2, ...N\}$ be the known samples, where N is the number of samples. The SVDD model is used to describe the dataset. The parameters of the

SVDD model are the radius R and the center $\mathbf{a}$ of the hyper-sphere. The object is to describe the dataset by using a hyper-sphere with minimized radius in feature space. In other words, all the samples should be located in the hypersphere. The mathematical form of the model is minimizing the function $F(R,\mathbf{a}) = R^2$ with the constraint condition $\|x_i - \mathbf{a}\|^2 \le R^2 (\forall i = 1,2,...,N)$. If outliers appear in the dataset, the distances from the samples $\{x_i, i = 1,2,...,N\}$ to the center of the hyper-sphere were not strictly smaller than R . But large distance should be penalized. Thinking the influence of outliers, slack variable $\xi_i \ge 0,\ (i = 1,2,...,N)$ are introduced in the objective function. So the problem of minimizing the radius of the hyper-sphere can be described by the following quadratic programming with inequality constraints

$$\begin{cases} \min R^2 + C\sum_{i=1}^{N} \xi_i \\ sub: \|x_i - \mathbf{a}\|^2 \le R^2 + \xi_i, \xi_i \ge 0, i = 1,2,...,N. \end{cases} \tag{1}$$

where the positive constant parameter C is called penalty factor. The parameter C controls the tradeoff between the radius of the hyper-sphere and the error.

Using the Lagrange multiplier algorithm for Eq.(1), the corresponding Lagrange function is

$$\begin{aligned} &L(R, \mathbf{a}, \alpha_i, \beta_i, \xi_i) = \\ &\quad R^2 + C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i (R^2 + \xi_i - \|x_i - \mathbf{a}\|^2) - \sum_{i=1}^{N} \beta_i \xi_i \end{aligned} \tag{2}$$

where $\alpha_i \ge 0, \beta_i \ge 0$ are Lagrange multipliers. Lagrange function L should be minimized with respect to R , $\mathbf{a}_i$, ξ , and maximized with respect to α_i and β_i . The extremum conditions of Lagrange function L are

$$\frac{\partial L}{\partial R} = 0, \frac{\partial L}{\partial \mathbf{a}} = 0, \frac{\partial L}{\partial \xi_i} = 0 \tag{3}$$

such that

$$\sum_{i=1}^{N} \alpha_i = 1 \tag{4}$$

$$\mathbf{a} = \sum_{i=1}^{N} \alpha_i x_i \tag{5}$$

$$C - \alpha_i - \beta_i = 0 \tag{6}$$

We can get $0 \le \alpha_i \le C$ from Eq.(6) because $\alpha_i \ge 0, \beta_i \ge 0$. When Eq.(4-6) are substituted into Lagrange function Eq.(2), the dual form of the Lagrange optimization problem turns into

$$\begin{cases} \max \sum_{i=1}^{N} \alpha_i (x_i \cdot x_i) - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j (x_i \cdot x_j) \\ sub: \sum_{i=1}^{N} \alpha_i = 1, \quad 0 \le \alpha_i \le C, \;\; i, j = 1, 2, ..., N. \end{cases} \tag{7}$$

where $x_i \cdot x_j$ is the inner product of x_i and x_j. Usually the dataset is not distributed in the hyper-sphere ideally. So the inner product can be substituted by some kernel function in high-dimensional feature space. After solving the quadratic programming problem containing inequality constraints denoted by Eq.(7), the parameters of the SVDD model $\{\alpha_i, i = 1, 2, ...N,\}$ is achieved. The parameters satisfy the following conditions

$$\begin{cases} \|x_i - \mathbf{a}\|^2 < R^2 \to \alpha_i = 0 \\ \|x_i - \mathbf{a}\|^2 = R^2 \to 0 < \alpha_i < C \\ \|x_i - \mathbf{a}\|^2 > R^2 \to \alpha_i = C \end{cases} \tag{8}$$

The multiplicative updating algorithm to solve the quadratic programming problem containing inequality constraints denoted by Eq.(7) of the SVDD model will be analyzed in next subsection.

2.2 Multiplicative Updating Algorithm to Solve Parameters of SVDD Model

The general formulation of the nonnegative quadratic programming is analyzed firstly. Consider the following minimization problem of quadratic function which contains inequality constraints

$$\min F(X) = \frac{1}{2} X^T \mathbf{A} X + \mathbf{b} X \tag{9}$$

In Eq.(9), the i th component of X is denoted as X_i, and the constrain conditions are $X_i \ge 0,\; i = 1, 2, ..., N.$ And $\mathbf{A}$ is $N \times N$ symmetric nonnegative matrix. Iterative algorithm can be constructed to solve the minimum value of Eq.(9) with the nonnegative constrain condition of X shown as reference[19]. Matrix $\mathbf{A}$ is expressed by the subtraction of two nonnegative matrix $\mathbf{A}^+$ and $\mathbf{A}^-$ as

$$\mathbf{A} = \mathbf{A}^+ - \mathbf{A}^- \tag{10}$$

where

$$(\mathbf{A}^+)_{i,j} = \begin{cases} (\mathbf{A})_{i,j} & \text{if } (\mathbf{A})_{i,j} > 0 \\ 0 & \text{others} \end{cases}, \quad (\mathbf{A}^-)_{i,j} = \begin{cases} -(\mathbf{A})_{i,j} & \text{if } (\mathbf{A})_{i,j} < 0 \\ 0 & \text{others} \end{cases} \tag{11}$$

and $(\mathbf{A})_{i,j}$ is the (i, j) component of matrix $\mathbf{A}$ in Eq.(11). The iterative formulation of the multiplicative updating algorithm is as [19]

$$X_i^{(k+1)} = X_i^{(k)}[\frac{-b_i + \sqrt{b_i^2 + 4(\mathbf{A}^+X^{(k)})_i(\mathbf{A}^-X^{(k)})_i}}{2(\mathbf{A}^+X^{(k)})_i}] \quad (12)$$

where $k \in \mathbf{Z}^+$ is iterative times, b_i, $(\mathbf{A}^+X^{(k)})_i$ and $(\mathbf{A}^-X^{(k)})_i$ are the i th component of vector $\mathbf{b}$, $\mathbf{A}^+X^{(k)}$ and $\mathbf{A}^-X^{(k)}$.

When the iterative multiplicative updating algorithm is used in the Lagrange optimization problem shown as Eq.(7), and let $(\mathbf{A})_{i,j} = x_i \cdot x_j$, $b_i = x_i \cdot x_i$, the iterative computing formulation for the parameters of the SVDD model is shown as

$$\alpha_i^{(k+1)} = \alpha_i^{(k)}[\frac{-b_i + \sqrt{b_i^2 + 4(\mathbf{A}^+\alpha^{(k)})_i(\mathbf{A}^-\alpha^{(k)})_i}}{2(\mathbf{A}^+\alpha^{(k)})_i}] \ . \quad (13)$$

2.3 The SVDD Model Containing Two Classes of Samples

The SVDD model with one-class samples can be extended to the case of samples with two classes. Consider dataset $\{(x_1, y_1)\,(x_2, y_2)\,,..\,(x_N, y_N)\}$ come from two different classes of samples, where N is the number of samples. And x_i is the feature vector of the i th sample, $y_i = 1$ or -1 , $i = 1, 2, ..., N$. Not losing generality, for the samples $x_i, i = 1, 2, ..., l$, let $y_i = 1$, and for the samples x_i, $i = l+1, l+2, ..., N$, let $y_i = -1$. In other words, $\{x_i, i = 1, 2, ..., l\}$ are the positive samples, and $\{x_i, i = l+1, l+2, ..., N\}$ are negative samples or outliers. The positive samples $\{x_1, x_2 ..., x_l\}$ are in the hyper-sphere, and the negative samples $\{x_{l+1}, x_{l+2} ..., x_N\}$ are outside the hyper-sphere in the SVDD model. Slack variable $\xi_i^+ \geq 0$, $(i = 1, 2, ..., l)$ and $\xi_i^- \geq 0$, $(i = l+1, \quad l+2, ..., N)$ are introduced in the objective function for each sample in the dataset similar with in one-class case. The problem of minimizing the radius of the hyper-sphere can be formulated by the following quadratic programming with inequality constraints

$$\begin{cases} \min R^2 + C_1 \sum_{i=1}^{l} \xi_i^+ + C_2 \sum_{i=l+1}^{N} \xi_i^- \\ sub: \|x_i - \mathbf{a}\|^2 \leq R^2 + \xi_i^+, \xi_i^+ \geq 0, i = 1, 2, ..., l; \\ \qquad \|x_i - \mathbf{a}\|^2 \geq R^2 - \xi_i^-, \xi_i^- \geq 0, i = l+1, l+2, ..., N. \end{cases} \quad (14)$$

where the positive constant parameters C_1 and C_2 are penalty factors. Using Lagrange multiplier algorithm for Eq.(14), we can draw the corresponding Lagrange function as

$$
\begin{aligned}
&L(R,\mathbf{a},\alpha,\beta,\xi_i^+,\xi_i^-)= \\
&R^2+C_1\sum_{i=1}^{l}\xi_i^+ + C_2\sum_{i=l+1}^{N}\xi_i^- - \sum_{i=1}^{l}\beta_i\,\xi_i^+ - \sum_{i=l+1}^{N}\beta_i\,\xi_i^- \\
&-\sum_{i=1}^{l}\alpha_i\,(R^2+\xi_i^+ - \|x_i-\mathbf{a}\|^2) - \sum_{i=l+1}^{N}\alpha_i\,(\|x_i-\mathbf{a}\|^2+\xi_i^- - R^2)
\end{aligned} \tag{15}
$$

where $\alpha_i \ge 0, \beta_i \ge 0$ are Lagrange multipliers. Similar with Eq.(3), Lagrange function L should be minimized with respect to $R,\mathbf{a},\xi_i^+,\xi_i^-$ and maximized with respect to α_i and β_i. After computing the extremum conditions of Lagrange function L, the dual form of the Lagrange optimization problem Eq.(15) are shown as following quadratic programming problem containing inequality constraints

$$
\begin{cases}
\max \sum_{i=1}^{l}\alpha_i(x_i\cdot x_i)\text{-}\sum_{i=l+1}^{N}\alpha_i(x_i\cdot x_i)\text{-} \\
\quad \sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j(x_i\cdot x_j)+2\sum_{i=1}^{l}\sum_{j=l+1}^{N}\alpha_i\alpha_j(x_i\cdot x_j)-\sum_{i=l+1}^{N}\sum_{j=l+1}^{N}\alpha_i\alpha_j(x_i\cdot x_j)\,. \\
sub:\sum_{i=1}^{l}\alpha_i-\sum_{i=l+1}^{N}\alpha_i=1,\ \ 0\le\alpha_i\le C_1, i=1,2,...,l; \\
\quad 0\le\alpha_i\le C_2,\ \ i=l+1,l+2,...,N.
\end{cases} \tag{16}
$$

Let $\alpha_i' = y_i\alpha_i$, then we have $\sum_{i=1}^{N}\alpha_i' = 1$ and $\mathbf{a}=\sum_{i=1}^{N}\alpha_i' x_i$. Then Eq.(16) can be simplified as

$$
\begin{cases}
\max \sum_{i=1}^{N}\alpha_i'(x_i\cdot x_i)\text{-}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i'\alpha_j'(x_i\cdot x_j) \\
sub:\sum_{i=1}^{N}\alpha_i'=1,\ 0\le\alpha_i\le C_1, i=1,2,...,l; \\
\qquad\qquad 0\le\alpha_i\le C_2, i=l+1,l+2,...,N.
\end{cases} \tag{17}
$$

Similar with Eq.(7), the quadratic programming problem containing inequality constraints denoted as Eq.(17) can be solved using the iterative multiplicative updating algorithm. Then the parameters of the SVDD model Eq.(16) is

$$
\alpha_i = \frac{\alpha_i'}{y_i}, i=1,2,...,N.. \tag{18}
$$

Then the radius R and the center $\mathbf{a}$ of the hyper-sphere of the SVDD model are achieved. So the dataset containing two classes of samples are separated by the the hyper-sphere. Decision-making function to classify a new sample using the hyper-sphere of the SVDD model will be discussed.

2.4 Decision-Making Function and Kernel Transformation in Feature Space

The SVDD model is achieved from the known samples. If a new sample is in the hyper-sphere of the SVDD model, it belongs to the positive class. Otherwise it is negative one or outlier. So the following decision-making function can be constructed

$$y(x) = \mathrm{sgn}(R^2 - ((x \cdot x) - 2\sum_{i=1}^{N} \alpha_i (x_i \cdot x) + \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j (x_i \cdot x_j))) . \tag{19}$$

For a new sample denoted by x, if the computing result of Eq.(19) is $y(x) \geq 0$, it belongs to the positive class. And if the result is $y(x) < 0$, it is outlier or negative sample.

In order to determine the decision-making function, the radius and the center of the hyper-sphere should be computed. From Eq.(8), we can see that most of parameters α_i are zero. Only part of parameters are non-zero. The samples corresponding the non-zero α_i values are called support vectors. They determine the radius and the center of the hyper-sphere. The center $\mathbf{a}$ is calculated by Eq.(5). We assume that $\alpha_k \neq 0$ for some support vector x_k. Then the radius of the hyper-sphere can be calculated as

$$R = ((x_k \cdot x_k) - 2\sum_{i=1}^{N} \alpha_i (x_k \cdot x_k) + \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j (x_i \cdot x_j)))^{\frac{1}{2}} . \tag{20}$$

If the inner product of x_i and x_j is substituted by kernel function $K(x_i, x_j)$, the decision-making function can be shown as

$$y(x) = \mathrm{sgn}(R^2 - (K(x,x) - 2\sum_{i=1}^{N} \alpha_i K(x_i, x) + \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j K(x_i, x_j))) . \tag{21}$$

Kernel functions are usually constructed by mapping functions from the sample space to the high dimensional feature space. Examples of the kernel functions are linear kernel, polynomial kernel and the Gaussian kernel and so on.

3 Predication Method of Customer Credit Scoring Based on the SVDD Model

General predication methods driven by known samples will be discussed briefly. Then new predication method of customer credit scoring of banks based on the SVDD classification model is proposed. Merits of the proposed method over the SVM model based method are analyzed.

There are mainly two parts called as learning stage and predicating stage in pattern predication methods driven by dataset of samples. The processes of predication method are as following. Firstly, the original dataset are collected, cleared, completed and normalized in preprocessing stage. The normalized dataset $\{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$ are established, where $x_i \in R^p$ is the input p -dimension feature vector which represents known sample, and N is the number of samples. $y_i \in \{+1, -1\}$ is the decision value corresponding the input vector x_i , $i = 1, 2, ..., N$. In the learning stage, suitable mathematical models are selected to establish the classification decision- making function $y = f(x, P)$, where P is parameter set of the models. Different models such as artificial neural networks[8], SVM model[6], and SVDD model can be used to simulate the decision-making function. After establishing the models, the optimized decision-making function $y = f(x, P^*)$ will be achieved by some learning algorithm using the dataset of the known samples, where P^* denoted the optimized parameter set. At last, the predication result for a new sample x^* can be gained by inputting the sample x^* into the optimized decision-making function and computing the value of the function $y = f(x^*, P^*)$ in the predicating stage.

When the SVDD classification model is used in predication of customer credit scoring of banks, positive samples (well-recorded customers) are assumed to be in the hyper-sphere, and negative samples (bad-recorded customers) be out of the hyper-sphere. The predication method of customer credit scoring of banks based on the SVDD model can be summed as following. The dataset of learning samples $\{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$ are achieved after the preprocessing stage from existing original data records. Some special kernel function is selected, and the SVDD classification model is determined. Parameters $\{\alpha_i, i = 1, 2, ..., N\}$ of the SVDD model denoted by the quadratic programming problem Eq.(16) are solved by using the iterative multiplicative updating algorithm Eq.(13). Then the center $\mathbf{a}$ and the radius R of the hyper-sphere are calculated by Eq.(5) and Eq. (20) separately. Such, the SVDD classification model is established by the known samples. In the predication stage, when a new sample x^* to be predicated is substituted into the decision-making function Eq.(21), comes out the predicating value. We can explain which class the sample belongs to according to the value meaning.

The difference of the SVDD classification model and the SVM model will be discussed briefly. The SVM model mainly classifies samples of two or more classes, the principle of which is to maximize the margin hyper-plane that gives the maximum separation between two classes of samples. Both positive and negative classes of samples are needed in the processes of computing the parameters of the model. But in the SVDD model, optimal hyper-sphere is established that contain the most of positive samples and exclude the most of outliers. So the model can be solved by one class of samples or two classes of samples. The SVDD classification model has following characteristic cs and advantages over the SVM model:

1) The support vectors in the SVM model are the learning samples that determine the margin of hyper-plane. The support vectors in the SVDD model are the samples that are located at the boundary of the hyper-sphere. Both experiments and theoretical analysis showed that the number of support vectors of the latter is smaller than the former. So the computational costs of the SVDD model are smaller than that of the SVM model usually [21].

2) The numbers of two classes of samples influence the decision-making function of the SVM model greatly. So some scholars determine the weights of positive and negative samples in weighted SVM model by using the ratio of the number of positive and negative samples[17]. But in practical applications, the ratio of the number of positive and negative samples may be unknown. So the ratio is determined subjectively. On the other hand, the decision-making function is established by the hyper-sphere which is solved by support vectors located at the boundary of the hyper-sphere in the SVDD model. The ratio of the number of positive and negative samples has less influence in the decision-making function because the hyper-sphere can be determined by only one class of samples in some extreme situations. Simulation results in next section illustrate this advantage.

3) There are many matrix operations in solving the parameters denoted by quadratic programming problems in the SVM model, which take up much time [22]. The iterative multiplicative updating algorithm is used to solve the parameters of the SVDD model in our work. Not only the solving processes are simplified but also the computing efficiency is improved. Following experiments on artificial synthesized dataset and benchmark credit dataset of banks show the improvement of the proposed credit scoring predication method.

4 Experimental Results

Experiments on artificial dataset and benchmark individual credit dataset of banks are performed in this section, and effectiveness of the proposed customer credit scoring method based on the SVDD classification model with imbalanced dataset is illustrated. The weighted LS-SVM model is a new powerful classification model for imbalanced datasets proposed recently in machine learning [12]. So the proposed predication method is compared with the weighted LS-SVM model based method using same samples come from imbalanced datasets. We mainly want to indicate the improvement of the learning and predicating processes of the proposed method for the imbalanced dataset which containing different numbers of positive and negative samples. The learning and predicating accuracies are compared using the proposed method and the weighted LS-SVM model.

4.1 Experiments on Artificial Dataset

Two-spiral classification is one of classical problems in pattern recognition[12]. The distributions of the samples of two-spiral function are shown as figure one.

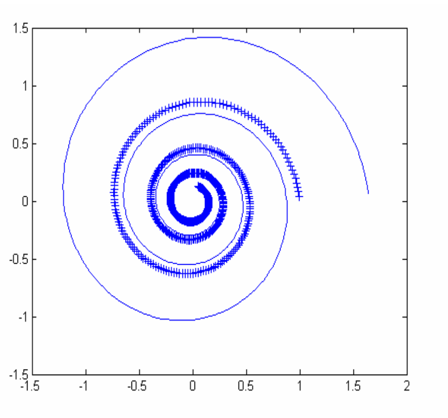


Fig. 1. The distribution of the samples of two-spiral function

Samples of the first class in figure 1 are produced by Eq.(22).

$$\begin{cases} x_1(t) = \exp(-0.2t)\cos(2t) \\ y_1(t) = \exp(-0.2t) sin(2t) \end{cases}. \tag{22}$$

And samples of the second class are produced by Eq.(23).

$$\begin{cases} x_2(t) = \exp(-0.2t + 0.5)\cos(2t) \\ y_2(t) = \exp(-0.2t + 0.5) sin(2t) \end{cases} \tag{23}$$

The values of parameter t are from 0.02 to 10, and the step is 0.02. There are five hundreds samples in each class. The first class is denoted as positive sample set, and the second is negative. The objective of experiments is to illustrate the improvement of accuracy of learning and predicating of SVDD model over the SVM model for classification problems with imbalanced dataset. The learning and predicating experiments using different number of samples are performed. And Gauss function Eq.(24) is used as kernel function.

$$K(x, y) = \exp(-\frac{\|x - y\|^2}{2\delta^2}) \tag{24}$$

Table one shows a group of learning and predicating results of two-spiral function using both models with several different numbers of positive and negative samples. This group of experiments is designed as follows. The parameter δ of the kernel function Eq.(24) is 2, and penalty factors C_1 and C_2 are both 4 in Eq.(17).

Table 1. Learning and predicating results of two-spiral classification problem

Model	WLS-SVM		SVDD	
(n^+, n^-)	ARL(%)	ARP(%)	ARL(%)	ARP(%)
(200,200)	99.25	96.33	98.25	95.83
(200,180)	95.26	94.52	96.32	94.19
(200,160)	92.78	91.72	94.72	93.28
(200,140)	92.65	91.67	96.47	92.88
(200,120)	90.00	88.82	94.06	92.35
(200,100)	85.67	87.43	93.33	90.86
(200, 80)	87.86	84.86	93.57	90.56
(200, 60)	85.39	84.46	91.92	91.62

In table one, n^+ and n^- denote the numbers of positive and negative samples. In learning process, n^+ positive and n^- negative samples are selected randomly from artificial two-spiral dataset show in table one. And same learning samples are used in both models in the each group of comparing experiment. In predicating process, residual parts of the artificial dataset are as samples to be predicated. In table one, (n^+, n^-) denotes numbers of learning samples, ARL denotes accuracy rate of learning, and ARP denotes accuracy rate of predicting. Accuracy rate of learning is the ratio of correctly classified samples with the total learning samples. Accuracy rate of predicting is the ratio of correctly classified samples with the total predicated samples. All experiments are performed under following conditions: hardware CPU Pentium4 2.4 GHZ, RAM 512MB; software Windows XP and Matlab7.0.

From the experimental results shown in table one, we can see the WLS-SVM model has higher accuracy of learning and predicating than the SVDD model slightly when the number of positive samples equals to the number of negative. But when the numbers of positive samples are different from the number of negative greatly, the SVDD model keeps higher accuracy of learning and predicating compared with the WLS-SVM model apparently. So the SVDD model is more effective to deal with classification problems with imbalanced artificial dataset.

4.2 Experiments on Benchmark Dataset

In this subsection, the experiment samples are selected from an opening database of computer institute of UCI university [23]. There are three sets of individual credit scoring of banks about Australia, Germen, and Japan in the database. Experimental results using the database about Australia are listed following. There are total 690 samples in the database. The number of positive sample (good credit) is 307, and others are negative sample (bad credit). There are fourteen index of evaluation and one credit value to compose a sample.

The database is preprocessed firstly. All the records in the database have been numerically disposed. So the index and credit value are expressed by numerical values accordingly. We can see that the input vector which denotes each sample is a fourteen-dimension vector from the index of credit of each sample. The inner product of two vectors is relative with each component of the both vectors. But there exists magnificent discrimination in the value ranges of each index in original database. In order to balance the effect of each component of the input vector (each credit index), all the values of index are normalized using as

$$\hat{x}_{i,j} = \frac{x_{i,j} - \min x_j}{\max x_j - \min x_j} \tag{25}$$

where $\max x_j$ and $\min x_j$ denote the maximum and minimum values of the j th index of all the samples in the database, and $j = 1, 2, ..., 14$, $i = 1, 2, ..., 690$. Such, the normalized data are achieved.

Table 2. Learning and predicating results of individual credit scoring of banks

Model	WLS-SVM		SVDD	
(n^+, n^-)	ARL(%)	ARP(%)	ARL(%)	ARP(%)
(200,150)	73.71	62.35	80.29	70.29
(200,160)	75.83	63.33	81.94	68.49
(200,170)	79.46	68.13	83.24	71.88
(200,180)	81.84	66.67	85.26	72.58
(200,190)	84.10	69.67	83.33	73.67
(200,200)	85.75	71. 38	85.00	74.48
(190,200)	82.82	70.33	84.87	72.33
(180, 200)	82.89	67.74	84.47	73.55
(170, 200)	78.11	66.56	83.51	73.44
(160, 200)	76.67	64.85	80.56	70.30
(150, 200)	75.72	64.41	78.86	68.82

Table two shows a group of learning and predicating results of individual credit scoring of banks using the dataset of Australia using the weighted LS-SVM model and the SVDD model with different numbers of positive and negative samples. The kernel function and parameters are same as experiments on the artificial dataset. In table two, n^+ and n^- denote the numbers of positive and negative samples. In learning process, n^+ positive and n^- negative learning samples are selected randomly from the database about Australia. And same learning samples are used in both models in the each group of comparing experiment. In predicating process, residual parts of the dataset are as samples to be predicated. Experiment environments are same as those of the artificial dataset.

From table two, we can see that the SVDD model and the WLS-SVM model have similar accuracy of learning when the number of positive samples equals to the number of negative ones. The SVDD model has higher accuracy of learning and predicating compared with the WLS-SVM models when the number of positive samples are different from the number of negative ones greatly similar with experimental results on the artificial dataset. We see that the accuracies of learning and predicating using the dataset of the individual credit scoring of Australia are not as well as those of the artificial dataset of two-spiral. One reason is that the dataset of the individual credit scoring is not separable strictly itself [24].

5 Conclusions

New customer credit scoring predication method based on the SVDD classification model with imbalanced dataset was proposed in this paper. Main predication methods were reviewed firstly. Then the SVDD classification model for imbalanced dataset was analyzed. And the multiplicative updating principle to solve the parameters of the model was discussed. Following, new learning and predicating method of customer credit scoring of banks based on the SVDD model was proposed. At last, Experiments on the synthesized two-spiral dataset and the benchmark dataset of individual credit of banks using the proposed method and the WLS-SVM-based method were performed. Experimental results illustrated that the proposed method is more effective than the WLS-SVM-based method for classification problems with imbalanced dataset such as predication of customer credit scoring of banks. The learning and predicating accuracies of SVDD model are higher than the LS-SVM model under same experiment conditions.

Acknowledgment

The work was partly supported by the National Natural Science Foundation of China (70971083), Leading Academic Discipline Program, 211 Project for Shanghai University of Finance and Economics(the 3rd phase) and the National Post-Doctoral Foundation of China(20080440644).

References

1. Nanda, S., Pendharkar, P.C.: Development and Comparison of Analytical Techniques for Predicting Insolvency Risk. International Journal of Intelligent Systems in Accounting Finance and Management 10(3), 155–168 (2001)
2. Kay, S.M.: Fundamentals of Statistical Signal Processing. Estimation Theory, Detection Theory, vol. I, II. Prentice-Hall, New Jersey (1998)
3. Altman, E.L.: Financial Ratios Discriminate Analysis and the Prediction of Corporate Bankruptcy. Journal of Finance 23(3), 589–609 (1968)

4. Pendharkar, P.C.: A Computational Study on the Performance of ANNs under Changing Structural Design and Data Distributions. European Journal of Operational Research 138(1), 155–177 (2002)
5. Gomez Skarmeta, A.F., Delgado, M., Vila, M.A.: About the Use of Fuzzy Clustering Techniques for Fuzzy Model Identification. Fuzzy Sets and Systems 106(2), 179–188 (1999)
6. Bernhard, S., Sung, K.K.: Comparing Support Vector Machines with Gaussian Kernels to Radical Basis Function Classifiers. IEEE Transaction on Signal Processing 45(11), 2758–2765 (1997)
7. Thomas, L.C., Edelman, D.B., Crook, J.N.: Credit Scoring and Its Applications. Cambridge University Press, Cambridge (2002)
8. Desai, V.S., Crook, J.N., Overstreet, J.G.A.: A Comparison of Neural networks and Linear Scoring Models in the Credit Union Environment. European Journal of Operational Research 95(1), 24–37 (1996)
9. David, R.H., Edelman, D.B., Gammerman, A.J.: Machine Learning Algorithm for Credit-card Applications. Journal of Mathematics Applied in Business and Industry 4(1), 43–51 (1992)
10. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1998)
11. Cortes, C., Vapnik, V.: Support Vector Networks. Machine Learning 20(3), 273–297 (1995)
12. Suykens, J.A.K., Vandewalle, J.: Least Squares Support Vector Machine Classifiers. Neural Proc. Lett. 9(3), 293–300 (1999)
13. Zhang, L., Zhou, W., Jiao, L.: Wavelet Support Vector Machine. IEEE Trans. on Systems, Man, and Cybernetics-part B: Cybernetics 34(1), 34–39 (2004)
14. Tax, D.M.J., Duin, R.P.W.: Support Vector Data Description. Machine Learning 54(1), 45–66 (2004)
15. Wang, Y., Wang, S., Lai, K.K.: A New Fuzzy Support Vector Machine to Evaluate Credit Risk. IEEE Transaction on Fuzzy Systems 13(6), 820–831 (2005)
16. Zadrozny, B., Langford, J., Abe, N.: Cost Sensitive Learning by Cost Proportionate Example Weighting. In: The 3rd IEEE International Conference on Data Mining, pp. 435–442 (2003)
17. Lin, Y., Lee, Y., Wahba, G.: Support Vector Machines for Classification in Nonstandard Situations. Machine Learning 46(2), 191–202 (2002)
18. Thomas, L.C.: A Survey of Credit and Behavioral Scoring: Forecasting Financial Risk of Lending to Customers. International Journal of Forecasting 16(2), 149–172 (2002)
19. Tax, D.M.J.: One-class Classification. PhD thesis, Delft University of Technology, `http://www.ph.tn.tudelft.nl/~davidt/thesis.pdf` (last accessed, March 2010)
20. Sha, F., Lin, Y., Saul, L.K., Lee, D.D.: Multiplicative Updates for Nonnegative Quadratic Programming. Neural Computation 19(8), 2004–2031 (2007)
21. Tax, D.M.J., Duin, R.P.W.: Support Vector Domain Description. Pattern Recognition Letters 20(11-13), 1191–1199 (1999)
22. Sha, F., Saul, L.K., Lee, D.D.: Multiplicative Updates for Nonnegative Quadratic Programming in Support Vector Machines. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems, vol. 15, pp. 897–904 (2003)
23. Asuncion, A., Newman, D.J.: UCI Machine Learning Repository, `http://www.ics.uci.edu/~mlearn/MLRepository.html` (last accessed, March 2010)
24. Schölkopf, B., Smola, A.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge (2002)

K-BRiR: An Efficient and Self-organized Routing Mechanism for Structured Peer-to-Peer Service Networks*

Zhenhua Tan[1], Guangming Yang[1], Zhiliang Zhu[1], Wei Cheng[1], and Guiran Chang[1,2]

[1] Software College, Northeastern University, Shenyang City, Liaoning, China
[2] Computing Center, Northeastern University, Shenyang City, Liaoning, China
{tanzh,yanggm,zzl,chengw}@mail.neu.edu.cn, chang@neu.edu.cn

Abstract. Peer-to-Peer service network takes an important role in the internet technology. The routing mechanism is the core technology for P2P service network. In this paper, an efficient and self-organized routing mechanism named K-BRiR for structured P2P service network is presented. K-BRiR includes a basic routing protocol BRiR which is based on ring topology. The definitions and algorithms are discussed. Our experimental results confirm that K-BRiR is stable and efficient in P2P service network.

Keywords: Peer-to-peer, Service Network, Routing Mechanism, Self-organized, BRiR (Basic Routing in Ring topology), K-BRiR.

1 Introduction

Peer-to-peer (commonly abbreviated to P2P) service network, which is self-organized and distributed, is composed of participants that make a portion of their service resources directly available to other network participants, without the need for central coordination instances [1-7], and developed very fast recently.

The P2P service network can be classified as unstructured and structured based on the topology. Each peer in the structured P2P records some other peers' fingers via a certain algorithm, and the route (or lookup) can be finished in a certain hops. Typical algorithms are Chord [8], CAN [9], Tapestry [10], Pastry [11], and so on. Compared to the unstructured overlay network, structured P2P has advantages in system cost and route efficiency. However, it is slow 4 to 9 times than the one-hop network [12].

As the network scale increasing continuously, the structured P2P service network require more efficient routing performance, while the traditional structured P2P routing protocols using O (logn) routing fingers to get O (logn) routing path and spending O (logn) network churning. Table 1 summarizes several P2P service systems' performance.

* This paper is supported by Supported by Major Program of National Natural Science Foundation of China (no. 70931001).

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 61–74, 2010.

In order to improve the routing algorithm efficiency of the structured peer-to-peer overlay networks, using much less routing fingers to get longer routing path and reducing the network churning, the routing algorithms are studied by this paper.

Table 1. Performance comparison of several P2P systems

Name	Route Hops	Churning when nodes' joining
Chord	$O(1/2*\log_2 N)$	$O(\log^2 N)$
CAN	$O(d/2*N^{1/d})$	$O(d/2N^{1/d})$
Pastry	$O(1/b*\log_2 N)$	$O(\log_{2b} N)$
Tapestry	$O(\log_B N)$	$O(\log^2 N)$

In order to get higher routing performances, this paper presents a basic routing algorithm of structured P2P service networks, called BRiR (Basic Routing in Ring-topology), in ring topology with cache nodes, successor/previous nodes and a super node. The BRiR can get O (N) routing path only use O (L+2) routing fingers, and the routing operations can be finished in O (2) hops in average. Cache policy, routing maintenance algorithms for nodes' joining and leaving and election algorithms for super node were presented to ensure the self-organization in BRiR. To tackle the churning problem during super-electing in BRiR, this paper presents an improved routing algorithm K-BRiR of structured P2P service networks by clustering the nodes' communication history based on BRiR. K-BRiR computes the nodes' stability via storing the nodes' communication history distributed, using the K-means algorithm to cluster the BRiR nodes by the stabilities into K regions which inter-linked each other by steady center node, and it can get O (N) routing path only use O (L+K) routing fingers, and the routing operations can be finished in O (K) hops in average. A distributed cluster model, routing maintenance algorithms for nodes' joining and leaving and cluster-center node's updating algorithm are presented to ensure the self-organization in K-BRiR.

In the remainder of the paper, section 2 provides routing protocol BRiR and K-BRiR. Section 3 describes the maintenance algorithms of K-BRiR. Section 4 describes how to optimize the routing performances by K-means. The simulations and results follow in section 5, with conclusions afterwards in section 6.

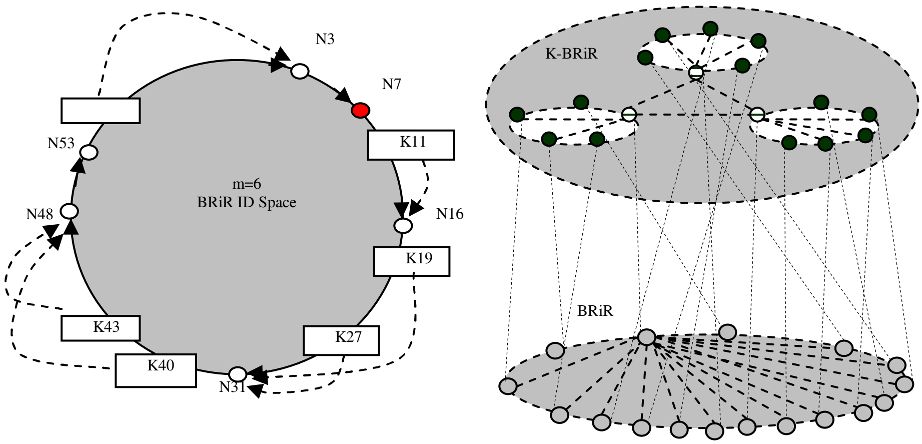
2 K-BRiR Routing Protocol

2.1 Topology

BRiR (Basic Routing in Ring topology) topology is the base of K-BRiR topology. K-BRiR is an overlay of BRiR. Figure 1 shows the topology. BRiR constructs its ring topology based on DHT. Every node or key word has a unique identifier with m-bits via DHT algorithm. All of the nodes identifiers are ranged between 0 and 2^m, and constructs as a ring in the identifier space.

In BRiR, given a key, the node responsible for storing the key's value can be determined using a hash function that assigns an identifier to each node and to each key (by hashing the node's IP address and the key). This paper uses SHA-1 hash algorithm to get identifier.

Use NodeID (or ID for short) and KeyID to express the identifier of node and key word. BRiR topology is composed of all of these identifiers as a ring by identifiers clockwise ranked.



(a) BRiR (b) K-BRiR

Fig. 1. K-BRiR topology is an overlay of the BRiR. BRiR is the basic topology, and K-BRiR is the clustered result by k-means according to some specified data in BRiR.

K-BRiR stored the node's communication history based on BRiR to calculate node's stability, then, the K-BRiR clustered the node's stability based on K-means. As a result, the BRiR overlay network could be divided into K sub-BRiR, just as shown in figure 1(b).

2.2 Basic Definition for K-BRiR

In order to demonstrate more clearly, this paper do some simple definitions first.

Definition 1. *Pre(*NodeID*)*. Given a peer NodeID, we call its closest predecessor node as *Pre(NodeID)* in the BRiR identifier space. E.g. in figure 1(a), Pre(N16) = N7, Pre(N3) = N53.

Definition 2. *Successor(*NodeID*)*. Given a peer NodeID, we call its closest successor node as *Successor(NodeID)* in the BRiR identifier space. E.g. in the figure 1(a), Successor (N16) = N31, Successor (N53) = N3.

Definition 3. *Successor(KeyID)*. Given a key word KeyID of resource, we call the first node whose identifier NodeID is equal or follows KeyID in the BRiR identifier space as *Successor(KeyID)*. E.g. Successor(K11) = N16, Successor(K55) = N3.

Definition 4. *Previous(NodeID)*. Given a peer NodeID, we call its closest previous node as *Previous(NodeID)* in the BRiR identifier space. E.g. in the figure 1(a), *Previous* (N16) = N7, *Previous* (N3) = N53.

Each KeyID is stored on the node *Successor(KeyID)*. E.g. the resource K11 is stored on the N16. According to DHT, we should store the <keyID, value> pair in the system. This paper simplifies the <KeyID, Value> as KeyID, and uses *KeyList* with structure <*Key, value*> to store the <k, v> pair information. So, K19 and K27 are stored in the *KeyList* of N31.

Three kinds of node are defined in the BRiR. (1) Normal node. This node provides or gets service from other nodes in BRiR. All nodes are normal node first. (2) Super node. This node records all of the other nodes' indexes (e.g. node IDs) as a directory node. It is unique in the BRiR, and is elected or built dynamically. Different from super node in hybrid P2P system, e.g. KazzaA, super node in BRiR is not fixed but dynamical. (3) Cluster-Center node. This kind of node is the core of cluster result in K-BRiR which is clustered based on BRiR by k-means.

2.3 Basic Communication Routing Mechanism

2.3.1 Routing Table

Every peer node in BRiR has a route table with same structure which includes five parts: Cache-Nodes, Super-Node, Successor-Node, Previous-Node and Cluster-Center-Nodes. In order to express neatly, these five parts are named shortly as CN, SpN, ScN, PreN and CCN in this paper, and figure 2 shows the route table. The five parts of the route table take different roles during routing.

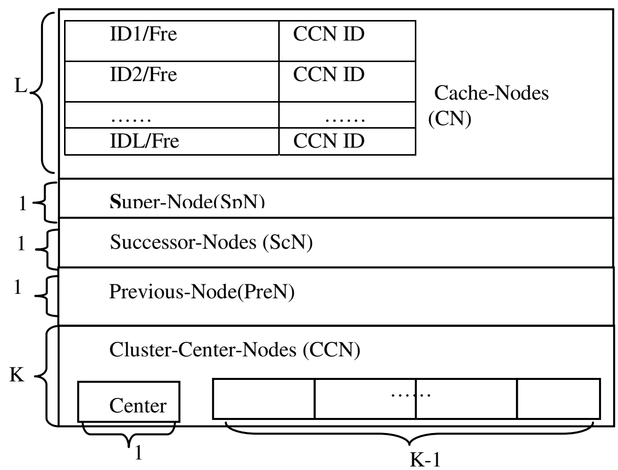


Fig. 2. Route table in BRiR is composed of five parts: CN, SpN, ScN, PreN and CCN

(1) Cache-Nodes (CN for short).

The Cache-Nodes item is used to store the identifiers of nodes who communicated with the current node recently (and we can call these nodes old nodes). The recent L nodes would be kept in the cache list. And, if the current node is a super node, the length of Cache-Nodes list can be adjusted to 2^m at most.

Cache-nodes list (or called CN list) has double lists, one is to record the CN identifiers and frequency of the related communication, marked as "ID/Fre", and one is to store CCN ID.

As we can see from figure 2, the node IDi has communicated with the current node f times. Obviously, we should make algorithm to maintain this cache list to replace old ones when more new nodes coming. We replace the least frequency node when the new node coming if the cache list is full and the algorithm can be written in pseudo-code below:

```
If(CN is not full){
  CN.add(new-node.ID,fre=1);//id=new-node.ID, fre=1
}Else {
  i = indexof(min(CN.Fre));
  CN.ID[i] = new-node.ID;
  CN.Fre[i] = 1;
}
```

(2) Super-Node (SpN for short).

Super-Node item is used to record the identifier of super node which is selected by algorithm and unique in BRiR. The super node takes major roles in routing, just like a route catalog.

(3) Successor-Node (ScN for short).

Successor-Node item points to the nearest subsequence peer, that is: Node.ScN = Successor(Node.ID).

(4) Previous-Node (PreN for short).

The Previous-Node item points to the nearest previous peer, that is: Node.PreN = Previous(Node.ID).

(5) Cluster-Center Node(CCN for short).

After cluster by k-means, the BRiR could be divided into k logical sub-net, and every sub-net has a central node, we call this central node 'Cluster-Center-Node', or 'CCN' in short. A CCN node stores the related sub-net nodes' ID list. Of course, the cluster dependencies are defined in the later sections. The CCN node can be only exists in K-BRiR, never in BRiR.

2.3.2 Routing Algorithm

In P2P system, a routing means look up resource URL in P2P service network. In BRiR, given a KeyID, the destination node is the successor (KeyID) in a routing; we can call this destination node as *pEnd* for short. Assume the node that starts the routing is pStart. In K-BRiR, we take five steps to search the destination node pEnd. Figure 3 shows the routing processes.

When the node pStart starts the routing, it firstly lookup keyID in its cache list; and it continues to lookup in its cluster center (pStart.center) when its cache doesn't feedback pEnd. Also, the routing can be sent to the other k-1 cluster centers (CCN) to get pEnd while the pStart.center doesn't get pEnd. If no correct destination found, the pStart would start the lookup in super node or its successor in the whole BRiR. The super node records all of the nodes' identifiers and could feed back a concrete result (*pEnd* or null) to *pStart* when the super node is alive. Or else, the *pStart* sends the routing message to its successor (*pStart.ScN*) and the successor will search the destination node one by one until the routing gets the *pEnd* or returns to *pStart*.

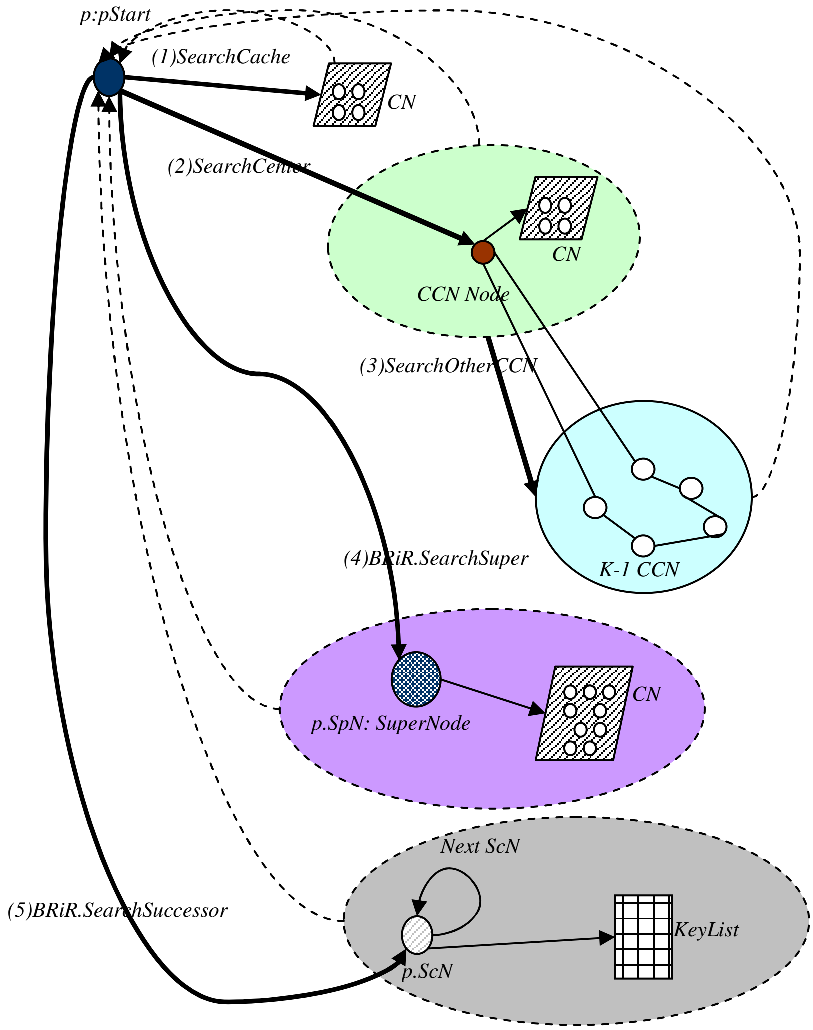


Fig. 3. There are five steps in K-BRiR's routing: *SearchCache*, *SearchCenter*, *SearchOtherCCN*, *SearchSuper* and *SearchSuccessor*

We write the pseudo code below according to the algorithm.

```
Step 1. SearchCache
  If(exists pEnd=successor(keyID) in p.CN [1..L]){
    Stop route;
    Return pEnd;
  }else{
    if p.center is alive{
      pStart.Send(Routing Message) to pStart.center;
      Notify Step 2.SearchCenter;
    }Else{
```

```
      pStart.Send(Routing Message) to pStart.SpN;
      Notify Step 4.BRiR.SearchSuper;
    }
Step 2. SearchCenter
  //we call the pStart.center as center here
  If(exists pEnd=successor(keyID) in center.CN [1..L]){
    Stop route;
    Return pEnd;
    }else{
      notify Step 3.SearchOtherCCN;
  }
Step 3. SearchOtherCCN
//we call the pStart.center as center here
  Flag = false;
  For(i=2 to K){
    If(exists pEnd=successor(keyID) in pStart.CCN[i].CN
    [1..L] }
      Flag = true;
      Stop route;
      Return pEnd;
    }
   }
  If(flag = false){
    Notify pStart.SpN: Step 4.BRiR.SearchSuper;
    End if;
   }
Step 4.BRiR.SearchSuper();
  //Rank the super.CN as an ascending list order by
nodeIDs.
  For(i=1 to L){
   If (keyID between super.CN[i] and super.CN[i+1]){
     If(super.CN[i+1].searchKey(KeyID) != null){
       super.send(super.CN[i+1]) to pStart;
      Stop Route;
     }
   }
  }
  Super.Send(Routing Message) to pStart.ScN;
  Notify Step5.BRiR.SearchSuccessor;
Step5.BRiR.SearchSuccessor();
  //We call the pStart.ScN as p here:
 If(p.ID = pStart.ID){
  p.send("No such keyID resource, finished!") to pStart;
 }Else{
  If(p.SearchKey(KeyID) != null){
    p.Send (p.ID) to pStart ;
    Stop Route;
  }Else{
    p.send(Routing message) to p.ScN;
    p.ScN goto Step 5.BRiR.SearchSuccessor again.
  }
}
```

3 Maintenance Algorithms

Nodes may join or leave the P2P system at any possible time. Some nodes may be invalid during the communication for some reason. By the way, the P2P service networks would be churned when a certain node joins or leaves the system.

Definition 5. *ChurningCount.* The P2P service network "*ChurningCount*" is the amount of nodes which are involved in when a node joins or leaves the P2P communication network.

3.1 When Peer Joins in BRiR

When a certain peer node wants to join the P2P system, it must know the location of another node in the P2P system. In fact, service providers always provide or configure some well-known peer nodes for public.

(1) When the service provider or sponsor user creates the first node in the P2P system, it will be the super node, and its successor and previous node is itself.

(2) When users want to join the P2P system, and they have known some well-known peer location. Assume the well-know peer is *wNode* and the new peer is *newNode*. The *wNode* can help *newNode* to find the previous and successor of *newNode*, so that the *newNode* can join into the appropriate position. Any new node must register in super node, so, the *newNode* should send its related information to the super node. Of course, super node updates its cache-Nodes items with the identifier of the new node.

3.2 When Peer Leaves BRiR

In P2P system, we need to probe the nodes' leaving because there isn't any notification when some nodes left the system. Thus, any peer may leave the system actively or at random passively. The node *p* notifies the super node (*p.SpN*) and cluster center node (*p.CCN*) that its successor has left by sending message and the super node then remove related item in cache-nodes of its route table. Then node *p* gets a new successor node in the identifier space via DHT.

3.3 Electing Super Peer Node

Generally, the first node is configured by service provider for public, it is robust relatively. However, we need a kind of algorithm to elect a new super node while the super node left the system or has been expiration (for the network reason or else). This paper presents an election algorithm by improving the selection algorithm for distributed processes made by Chang and Roberts [13]. Any node would sponsor the election operation in identifier space as soon as it finds the unique super node invalid in BRiR. At last, we elect the node with smallest identifier as super node, and all of the other nodes send their information to this node for registering. The following is the description for electing algorithm.

Step 1: Declare message *node-list* used to store identifier of nodes who took part in the selection. And we use a variable *selected-node-id* to store the identifier of the current selected super node in the message. At the very first, the node *p* that sponsors the selection operation will be the selected node.

Selected-node-id = *p.id*;
Node-list. add (p.id);

The node *p* sends the message *node-list* and *Selected-node-id* to its successor node.

Step 2: As soon as node p_i receives the message, it compares the *Selected-node-id* with itself identifier, and the smaller identifier will replace the current *Selected-node-id*. Then, the node p_i updates its route table by replace the SpN with *Selected-node-id*.

Selected-node-id = min (p_i*.id, Selected-node-id);*
Node-list.add(p_i*.id);*
p_i*.SpN = Selected-node-id;*

Step 3: The selection process completes when the *Selected-node-id* equals to the identifier of p_i. It means that the node p_i is the final selected super node. Then the p_i updates its cache-nodes of its route table with the received message *node-list*. So, this super node p_i has all of the nodes info now. Then stop electing.

Of course, if the *Selected-node-id* doesn't equal to the identifier of p_i, then node p_i sends message *node-list* and the *Selected-node-id* to its successor node. Go to the step 2.

4 Optimized by K-means

We compared the performance of BRiR among chord, pastry and tapestry in algorithm complexity of route table size, complexity of route hops, fault tolerance and network churning count in table 2.

Table 2. Performance of BRiR

	Route table size	Route hops	Fault tolerance	Churning
BRiR	O(L+2) (usually L=8)	Usually O(1) (O(N) when super node invalid)	All the time	2 or 3 (O(Nlog2N) when selection super node)
Chord	O(M)	O($\log_2$N)	Sometimes	O(M)
Pastry	O($\log_{2b}$N)	O($\log_{2b}$N)	Sometimes	O($\log_{2b}$N) (usually: b=4)
Tapestry	O($\log_2$N)	O(log2N)	All the time (With great band-width cost)	O(blog$_b$N) (b is a basic number)

As we can see, with the nodes amount increasing, there will be great high network churning count and bandwidth cost when electing super node, but inefficient routing performance at the same time; and the super node in BRiR will have great pressure when more nodes join into the system.

To solve these problems, we improved the BRiR routing algorithm via clustering the nodes communication history by K-means in K-BRiR. The K-BRiR stored the node's communication history based on BRiR to calculate node's stability, then, the K-BRiR clustered the node's stability based on K-means. As a result, the BRiR overlay network could be divided into K sub-BRiR.

4.1 Distributed Nodes' Communication History

To store the node's communication history, a set of data structure definitions were made in this paper. The identity i or j stands for the node's ID.

Definition 6. S_{ij} stands for the times of successful transactions between node i and j; F_{ij} stands for the times of failed transactions between node i and j.

Definition 7. (A) I_{all} Stands for all of the BRiR nodes.(B) $history(i, j) = \begin{cases} true \\ false \end{cases}$ Stands for whether communication occurs between i and j, the "true" means i have communicated with j. (C) $I_i = \{ j \mid j \in I_{all}, history(i, j) = true \}$.(D) $Count_i = \sum_{j \in I_i} (S_{ij} + F_{ij})$.

Definition 8. $\lambda_i = \sum_{j \in I_i} S_{ij} \Big/ Count_i$, stands for the successful rate.

Definition 9. α-list stores the communication history. $\alpha - list(i) = \{ < j, S_{ij}, F_{ij} > \mid j \in I_i \}$.

α-list (i) is a ternary vector, and each element stores the successful and failed communication times between i and j. In our past work, we designed a similar list in article [14].

4.2 Distance Definition

The purpose of nodes cluster is to get more stable super node. From the perspective of social networks, the person who has more relations must have more stability in the social networks. So, the paper define the node's stability with $Count_i$ and λ_i .

Definition 10. $\beta(i)$ stands for the nodes' stability by accessed times and successful communicated rate: $\beta(i) = < Count_i, \lambda_i >$.

Definition 11. $\sigma(i, j)$ is the distance between $\beta(i)$ and $\beta(j)$:

$$\sigma(i, j) = \sqrt{| Count_i^2 - Count_j^2 |} + \sqrt{| \lambda_i^2 - \lambda_j^2 |} \tag{1}$$

4.3 Cluster Algorithm for α-list Based on K-means

According to K-means, the cluster result will have K set. In K-BRiR, these K set can be called K logical sub-net, and each sub-net should have one available node at least and also should have a kernel node who is the most stable one. Before cluster, we define a standard to finish the K-means cluster activity.

Definition 12. ε is the finish standard for K-means cluster in K-BRiR. We calculate the average stability $\overline{\beta}_c$ by $\overline{\beta}_c = < \overline{Count}, \overline{\lambda} >$ during the sub-net whose cluster center is 'c', and $\overline{Count}$ is the average value of $Count_i$ of the current sub-net, $\overline{\lambda}$ is the average of λ_i in the sub-net. When values can meet ε during twice cluster in a row, the cluster can be finished.

$$fin(\varepsilon) = \left(\left| \sum_{c1=1}^{K} \left| (|\beta_{c1}| - |\overline{\beta_{c1}}|) \right| - \sum_{c2=1}^{K} \left| (|\beta_{c2}| - |\overline{\beta_{c2}}|) \right| \right| < \varepsilon \right) \tag{2}$$

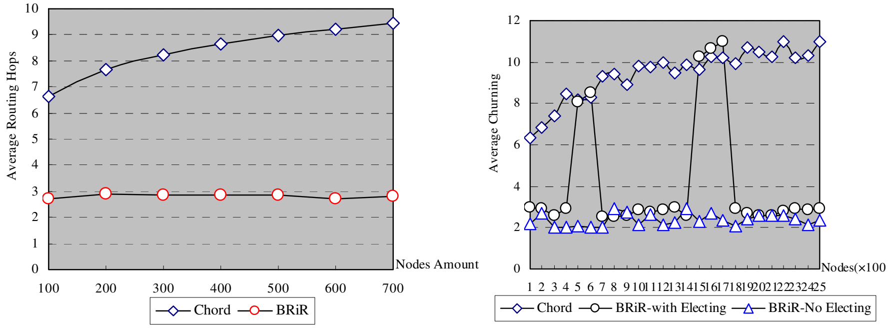
Definition 13. (a) $X = \{x_1, x_2, ..., x_n\}$ is a set of node, and , $\bigcup_{i=1}^{n} x_i = I_{all}$, $\forall x_i \neq x_j, x_i \cap x_j = \phi$. (b) $C = \{c_1, c_2, ..., c_k\}$, $\forall c_i \neq \phi$, $\forall c_i \cap c_j = \phi \Rightarrow X = \bigcup_{i=1}^{k} C_i$, is a K division for *X*. (c) $Kn = \{n_1, n_2, ..., n_k\}, \forall n_i \in Kn \Rightarrow n_i \in c_i$ is a set of K central nodes.

We describe the distributed cluster algorithm below.

Step 1. Super node selects K available nodes' at random as the cluster center *Kn*.
Step 2. These k central nodes send their local *α-list* to super node.
Step 3. The super node sends all of the k *α-list* to all of the other nodes in BRiR.
Step 4. Every node p calculate the distance via formula 1
Step 5. The super node re-calculates the average stability, and adjusting the cluster center for the given center C[i].
Step 6. The super node judge the *fin (ε)*, and the cluster would be stopped when *fin (ε)* is true. Or else go to step 2.
Step 7. Every nodes register their ID information in its related central node: p send ID info to p.center;

5 Simulations and Results

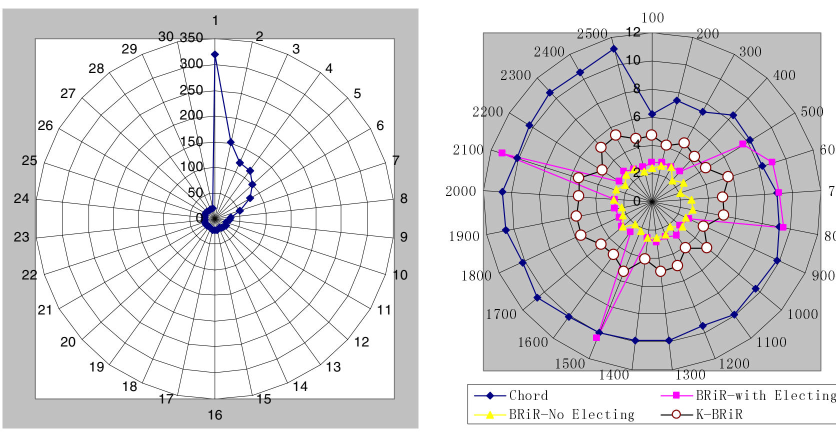
According to the principles of BRiR, we make a simulation system for BRiR with M.S. Dot Net IDE in C# programming language. Via comparing with Chord, we simulate the routing and maintenance performances for BRiR during random communication test. We test the average routing hops between chord and BRiR; the results are shown in figure 4(a), and average churning count in figure 4(b).



(a) Average routing hops (b) Average churning count

Fig. 4. Performance of average routing hops and churning count. BRiR has low churning count and the results is very smooth when no electing happened; But the churning will be increased suddenly when electing existed in BRiR because the electing was looped in all the topology.

We make the simulations for K-BRiR later. Figure 5 is the result.



(a) Cluster convergence (b) Comparison for Routing hops

Fig. 5. Simulation results of K-BRiR. We generated 2000 nodes in BRiR, and then simulated the cluster by k-means with k=6; (a) shows the cluster convergence and (b) shows the Routing hops comparison results.

Figure 5 (a) shows that the K-BRiR can be clustered smoothly, and has good cluster convergence. Figure 5 (b) shows that the K-BRiR has more stable routing performance than BRiR, and less routing hops than Chord. So, K-BRiR is an efficient and stable routing algorithm.

At last, we compare performances between K-BRiR and BRiR in Table 3 below. The results shows that K-BRiR is more stable and also efficient with O(K) routing hops.

Table 3. Comparison between K-BRiR and BRiR

	BRiR	K-BRiR
Routing Hops	Usually $O(1)$ and O(N/2) when electing happens	O(K)
Route Table Scale	O(L+2)	O(L+K)
Error Tolerance	High	High
Churning	O(2)and O(NlogN) when electing	O(K+2) and (K+N/K) when electing
Stability for Super Node	no	yes

6 Conclusions

This paper presents an efficient and self-organized routing mechanism for structured P2P service networks. BRiR is the basic routing mechanism based on ring topology with good routing performance but not stable. To improve it, K-BRiR clusters the nodes communication history by K-means; as a result, it has more stability and better routing performance than BRiR and traditional chord. Algorithms are designed in the paper as well as grouped definitions are created. Simulations showed the rightness and efficiency of BRiR and K-BRiR.

Of course, there also remain some problems in K-BRiR to be studied further, for example, how to design algorithm to select the initiating K value, how to define more comprehensive stability and etc. In our future work, this paper will do more to solve these problems and developed more useful simulation system and P2P service prototype systems.

References

1. Joung, Y.J., Yang, L.W., Fang, C.T.: Keyword Search in DHT based Peer-to-Peer Networks. IEEE Journal on Selected Areas in Communications 25(1), 339–348 (2007)
2. Farag, A., Ahmad, R.: Recommendation Retrieval in Reputation Assessment for Peer-to-Peer Systems. J. Electronic Notes in Theoretical Computer Science 244(1), 13–25 (2009)
3. Do, T.T., Hua, K.A., Jiang, N., Liu, F.Y.: Patch Peer: A Scalable Video-on-demand Streaming System in Hybrid Wireless Mobile Peer-to-peer Networks. J. Peer-to-Peer Networking and Applications 2(3), 182–201 (2009)
4. Ranka, S., Aluru, S., Buyya, R.: Security Issues in Cross-Organizational Peer-to-Peer Applications and Some Solutions. Communications in Computer and Information Science 40, 422–433 (2009)
5. Nakashima, T., Ono, A., Sueyoshi, T.: Performance Framework for P2P Overlay Network. In: 2008 Fourth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 101–104 (2008)
6. Do, T.T., Hua, K.A., Lin, C.S.: Ext Range: Continuous Moving Range Queries in Mobile Peer-to-Peer Networks. In: Proceedings of the Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, pp. 317–322 (2009)
7. Steinmetz, R., Wehrle, K.: Peer-to-Peer Systems and Applications. Springer, Berlin (2005)

8. Stoica, I., Morris, R., Karger, D.: Chord: A Scalable Peer-To-Peer Lookup Service for Internet Applications. In: Proceedings of the 2001 ACM Sigcomm Conference, pp. 149–160 (2001)
9. Ratnasamy, S., Francis, P., Handley, M.: A Scalable Content-Addressable Network. ACM Press, USA (2001)
10. Zhao, B., Kubiatowicz, J., Joseph, A.: Tapestry: An Infrastructure for Fault-tolerant Wide-area Location and Routing. University of California, Berkeley (2001)
11. Rowstron, A., Druschel, P.: Pastry: Scalable, Distributed Object Location and Routing for Large-scale Peer-to-peer Systems. In: IFIP/ACM International Conference on Distributed Systems Platforms (Middleware), pp. 329–350 (2001)
12. Jain, S., Mahajan, R., Wetherall, D.: A Study of the Performance Potential of DHT-based Overlays. In: Proceedings of the 4th Conference on USENIX Symposium on Internet Technologies and Systems, pp. 11–12 (2003)
13. Chang, E.G., Roberts, R.: An Improved Algorithm for Decentralized Extrema-finding in Circular Configurations of Processors. Communications of the ACM 22(5), 281–282 (1979)
14. Tan, Z.H., Cheng, W., Chang, G.R.: A Personalized Information Retrieval Algorithm Based on Word-Correlativity Model. Journal of Northeastern University (Natural Science) 29(4), 504–507 (2008)

Study on Amortization Time and Rationality in Real Estate Investment

Yancang Li*, Shujing Zhou, and Juanjuan Suo

College of Civil Engineering, Hebei University of Engineering, Handan 056038, China
liyancang@163.com

Abstract. Amortization time and rationality has been discussed a lot in real estate investment research. As the price of real estate is driven by Geometric Brown Motion (GBM), whether the mortgagors should amortize in advance has become a key issue in amortization time research. This paper presents a new method to solve the problem by using the optimal stopping time theory and option pricing theory models. We discuss the option value in amortizing decision based on this model. A simulation method is used to test this method.

Keywords: Real Estate, Amortization Time, GBM, Optimal Stopping Time Theory, Option Pricing Theory.

1 Introduction

In the real estate investment field, with the unveiling of the new loan policy, the problem whether the mortgagors should amortize in advance will directly affect the real benefits of the vast common house purchasers. And the existing literatures of real estate investment under real options' framework have been drawn attention to the modeling of the developers [1－8], but few focused on the individual real estate investment [9]. So the research of the individual mortgagors'optimal time to amortize has significance in theory and practice.

2 Description of the Problem

Generally, in the real estate investment, the mortgagors have to choose prepayments or deferring the amortizations after the interest rate is increased. The choice of the two strategies is the result the the balance of the two incomes. So, how to make a choice between the two? What is the border of their optimal time to amortize? What will affect the border? How to reach the border and what is its reachable probability and time? These are the major problems which should be solved.

In order to solve the questions proposed above, the following assumptaion were made.

(1) The price of real estate is driven by geometric Brown motion (GBM).

(2) In order to essure that the repayment decision will be immediately realized under the current market price, the repayment process of the mortgagors is instantaneous.

* Corresponding author.

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 75–85, 2010.

(3) To avoid the impact of the demand, amortizing activity of the mortgagors was defined not affect the price of real estate. It means that the mortgagors are recipients of the price.

(4) The mortgagors can be made at any moment to continue (deferring the amortizations) or stop (prepayments).

(5) The benefit function of the mortgagors is to maximize the expected present value of options. And the repayment process is irreversible.

3 Proposed Model

Suppose that the price of real estate (P) obeys the following GBM:

$$d_P = \mu P d_t + \sigma P d_W \tag{1}$$

where μ is the drift parameter of P; σ is the standard deviation of μ ; *W* is the standard Wiener process; d_W is the increment of *W*, subjected to the normal distribution which is a zero mean and a variance of d_t ($d_W = \varepsilon\sqrt{d_t}$, $\varepsilon \sim$N (0,1)).

Noting K is executive-cost, that is, the capital needs to be repaid, θ is the minimum saving repayment of the mortgagers which is a constant in a certain period of time, and then $K+\theta$ means the executive principal of repaying option. At the moment *t,* if the mortgagers are made prepayments(cessation) decision, they obtain profit: $P_t - K - \theta$ ($P_t > K + \theta$) , where P_t is the current real estate price at the moment *t*. By using the discount factor of the continuous compound interests, the proceeds that discount to zero moment is:

$$V(P,t) = e^{-\rho t}(P_t - K - \theta) \tag{2}$$

Where $V(P,t)$ is the payoff function, which is the present value of real estate possessed by the mortgagers at the moment *t*. ρ is a discount rate. Our goal is to maximize the present value of this expectation. It is clear that if the mortgagers chose prepayments recorded τ as the optimal time, the timing of the optimal stopping time is as follows:

$$V(P,t) = e^{-\rho t}(P_t - K - \theta) \tag{3}$$

Where the random variable τ is the stopping time, and $\tau = \{t \geq 0, P \geq P^*\}$. P^* is a question for the critical value: we take the stopping decision when $P < P^*$, or else we take the continuative decision. Moreover, $E_t[e^{-\rho\tau}(P_\tau - K - \theta)]$ is $E_t V(P,\tau)$ which means the payoff function of no prepayment at the moment *t*. The reason for the expected present value is that the moment τ of future payment is uncertain, which corresponds with prices P_τ not sure, and the relative benefits with P_τ can not be identified.

In Eq.(3), when the solution τ^* was obtained, then,

$$V^*(P,t) = \sup_\tau\{E_t[e^{-\rho\tau}(P_\tau - K - \theta)]\} = E_t[e^{-\rho\tau^*}(P_{\tau^*} - K - \theta)] \tag{4}$$

Where, the random variable $\tau^* = \inf\{t \geq 0, P \geq P^*\}$ is the optimal stopping time.

In fact, Eq.(2) and Eq.(4) indicate the stopping profit and the optimal continuing profit of the mortgagers. If the stopping profit is less than the optimal continuing profit, namely $V(P,t) < V^*(P,t)$, the mortgagers are sure to wait, in order to obtain greater benefits. To meet all the condition set by (P,t), we note it CR: $CR = \{(P,t)|V(P,t) < V^*(P,t), t \geq 0\}$; Once the optimal continuing profit is equal to the stopping profit, the mortgagers would make a decision to stop. To meet the condition, we note it SR: $SR = \{(P,t)|V(P,t) = V^*(P,t), t \geq 0\}$.

SR is no more than the meeting condition because $V(P,t)$ is a continuous function. In fact, before $V(P,t) > V^*(P,t)$, the mortgagers have made the stopping decision.

Now, we will deal with P^*. As long as the waiting time is long enough, there are countless moments when the price of real estate P hits P^*, but the mortgagers has made the stopping decision in the first observation to the P hitting the P^*. So $\tau^* = \inf\{t \geq 0, P = P^*\}$, this is actually the first reachable time. Therefore, if we want to determine τ^*, we only need to calculate P^*.

In fact, the solution of the optimization problem Eq.(3) τ^* is the solution of Snell envelope [10] $V(P,t) = V^*(P,t)$. Now we inspect the changes of the payoff function from t to τ. By Ito formula with considering Eq.(1) and $(d_W)^2 = 0$, $d_t d_W = 0$, $(d_t)^2 = 0$, it is not difficult to get Eq.(5):

$$d_{V(P,t)} = [\frac{\partial_{V(P,t)}}{\partial_P} + \frac{\partial_{V(P,t)}}{\partial_t} + \frac{1}{2}\sigma^2\frac{\partial^2_{V(P,t)}}{\partial_{P^2}}]d_t + \sigma P\frac{\partial_{V(P,t)}}{\partial_P}d_W \tag{5}$$

In fact, as d_W is non-existent, d_P also does not exist, their differential expression does not show the usual meaning. Such expression is only used as a convenience in the random calculous; the true meaning is limited to Ito integral.

$$V(P,\tau)-V(P,t)=\int_t^\tau[\frac{\partial_{V(P,s)}}{\partial_P}+\frac{\partial_{V(P,s)}}{\partial_s}+\frac{1}{2}\sigma^2\frac{\partial^2_{V(P,s)}}{\partial_{P^2}}]d_s+\int_t^\tau \sigma P\frac{\partial_{V(P,s)}}{\partial_P}d_W \quad (6)$$

Considering $V(P,t)=V^*(P,t)$, we can obtain

$$\begin{aligned}&\sup E_t V(P,\tau)-V(P,t)=V^*(P,\tau)-V(P,t)\\&=\sup E_t\{\int_t^\tau[\frac{\partial_{V(P,s)}}{\partial_P}+\frac{\partial_{V(P,s)}}{\partial_s}+\frac{1}{2}\sigma^2\frac{\partial^2_{V(P,s)}}{\partial_{P^2}}]d_s\}=0\end{aligned} \quad (7)$$

Where, the reason why the left item of the final equation $\sup E_t\{\int_t^\tau \sigma P\frac{\partial_{V(P,s)}}{\partial_P}d_W\}$ disappears is that non-random variable Ito integral correspond with the Gaussian nature [11].

According to Eq.(7), it is not difficult to get the following differential equation:

$$\frac{\partial_{V(P,t)}}{\partial_P}+\frac{\partial_{V(P,t)}}{\partial_t}+\frac{1}{2}\sigma^2\frac{\partial^2_{V(P,t)}}{\partial_{P^2}}=0,P<P^* \quad (8)$$

Due to the solution of Eq.(8) is $V(P^*,t)=e^{-\rho t}(P^*-K-\theta)$, the solution is just as: $V(P,t)=e^{-\rho t}F(P)$, and

$$\frac{1}{2}\sigma^2P^2\frac{\partial^2_{F(P)}}{\partial_{P^2}}+\mu P\frac{\partial_{F(P)}}{\partial_P}-\rho F(P)=0 \quad (9)$$

Where,

(1) If P = 0 , the option has no value. Namely: $V(0)=0$, therefore $F(0)=0$;

(2) The mortgagers take the stopping decision when P hits P^*, and their getting profit is $V(P,\tau^*)=e^{-\rho\tau^*}(P^*-K-\theta)$, this condition is called value-matching condition. It is equivalent to $F(P^*)=P^*-K-\theta$;

(3) $F(P)$ is continuous and smooth at P^*, it is equal to the smooth-pasting condition: $F'(P^*)=1$.

Due to the existing condition about the optimal solution in the boundary condition of Eq.(9), the second order differential equation judging by the form needs three boundary conditions. The solution is clearly described as follows:

$$\begin{cases} \frac{1}{2}\sigma^2 P^2 F_{PP} + \mu P F_P - \rho F = 0 \\ S.T.Zero - initial : F(0) = 0 \\ Value - matching : F(P^*) = P^* - K - \theta \\ Smooth - pasting : F^{'}(P^*) = 1 \end{cases} \tag{10}$$

The option is unlimited time level, so Eq.(9) does not contain time variable. Solving this equation is normally based on the zero-initial condition and assuming it has the solution as $F(P) = AP^{\beta}$, substitute into Eq.(10) and get the following quadric equation:

$$Q(\beta) = \frac{1}{2}\sigma^2 \beta(\beta - 1) + \mu\beta - \rho = 0 \tag{11}$$

As the quadratic equation's discriminant is greater than zero, so it must have two different solutions. We can easily obtain:

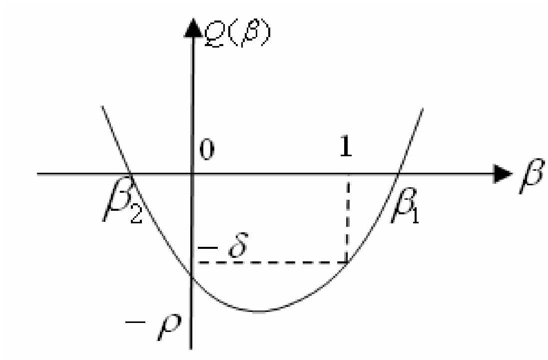


Fig. 1. The Situation of The Solution

$$\beta_1 = \frac{1}{2} - \frac{\beta}{\sigma^2} + \sqrt{(\frac{\mu}{\sigma^2} - \frac{1}{2})^2 + \frac{2\rho}{\sigma^2}}, \beta_2 = \frac{1}{2} - \frac{\beta}{\sigma^2} - \sqrt{(\frac{\mu}{\sigma^2} - \frac{1}{2})^2 + \frac{2\rho}{\sigma^2}} \tag{12}$$

The general solution Eq.(10) can be shown as linear combination of two independent solutions:

$$F(P) = A_1 P^{\beta_1} + A_2 P^{\beta_2} \tag{13}$$

Considering: $Q(1) = \mu - \rho = -\delta < 0, Q(0) = -\rho < 0$, thus $\beta_1 > 1$, $\beta_2 < 0$ (Fig. 1) and zero is absorption wall ($F(0) = 0$) to get $A_2 = 0$, thereby Eq.(13) become:

$$F(P) = A_1 P^{\beta_1} = AP^{\beta_1} \tag{14}$$

Substitute it into value-matching and smooth-pasting condition to get the following equations:

$$\begin{cases} AP^{\beta_1} = P - K - \theta \\ A\beta_1 P^{\beta_1 - 1} = 1 \end{cases} \tag{15}$$

And then:

$$P^* = \frac{\beta_1}{\beta_1 - 1}(K + \theta) \tag{16}$$

$$A = \frac{(\beta_1 - 1)^{\beta_1 - 1}}{(\beta_1)^{\beta_1}} \left(\frac{1}{K + \theta}\right)^{\beta_1 - 1}$$

After the analysis, we can draw the following conclusions:

(1) Time-choosing: The continuous region of the mortgagers is: $CR = \{P : P < P^*, P* > K + \theta\}$ and the stopping region is $SR = \{P : P \geq P^*, P^* > K + \theta\}$. It means that the mortgagers will choose prepayment if $P \geq P^*$, otherwise, they will continue to wait. The variable is presented as:

$$P^* = \frac{\beta_1}{\beta_1 - 1}(K + \theta) \qquad \beta_1 = \frac{1}{2} - \frac{\beta}{\sigma^2} + \sqrt{\left(\frac{\mu}{\sigma^2} - \frac{1}{2}\right)^2 + \frac{2\rho}{\sigma^2}} \tag{17}$$

(2) Decision Rules: according to $P^* = F(P^*) + K + \theta$, the mortgagers can get the optimal expectation PV:

$$V^*(P,t) = \frac{(\beta_1 - 1)^{\beta_1 - 1}}{(\beta_1)^{\beta_1}} \times \frac{(P_0)^{\beta_1}}{(K + \theta)^{\beta_1 - 1}} \tag{18}$$

P_0 is the price at the zero moment.

Conclusion 2 indicates that the decision threshold under the real option is higher than that normally got by the simple cost, this part is just the option value brought by the continuous decision.

4 Model Analysis

4.1 Reachability Analysis

We have calculated the timing threshold of the prepayment P^*. However, there are still many questions to the random process Eq.(1). A natural question is: Will the price P reach the threshold P^*? If so, how long can we go through the first achieved? Otherwise, P^* is no guidable meaning. It is regrettable that there is no consideration about this important timing topic in the traditional ROs framework.

For the random process Eq.(1), the density function [12] of the first reachable time τ^* is presented as:

$$g(t, P_0, P^*) = \frac{\log(P^*/P_0)}{\sigma\sqrt{2\pi t^3}} e^{\frac{[\log(P^*/P_0)-(\mu-\sigma^2/2)t]^2}{2\sigma^2 t}} \tag{19}$$

and P^* ($P^* > P$) is the threshold of P. Order $s = \mu - \frac{\sigma^2}{2}$:

If s>0, P^* must be achieved by probability 1 and the reachable E(T) &Var(T) are:

$$E(T) = \log(\frac{P^*}{P_0}) / s \qquad Var(T) = \sigma^2 \log(\frac{P^*}{P_0}) / s^3 \tag{20}$$

If s<0, P^* may be achieved, and the reachable probability $P_r = (\frac{P^*}{P_0})^{\frac{2\mu}{\sigma^2}-1}$, but its first order matrix does not exist;

If s=0, P^* must be achieved by probability 1 but its E(T) is infinity.

Similarly, if $P^* < P$ and s<0, P^* may be achieved by probability 1 in the limited time. Otherwise, its first and second order matrices don't exist. But textual economic implication means P^*>P.

4.2 Comparative Static Analysis

Because the increase of σ would reduce β_1 , so that the threshold $P^* = \frac{\beta_1}{\beta_1 - 1}(K + \theta)$ increases.

$$\frac{\partial_Q}{\partial_{\beta_1}}\frac{\partial_{\beta_1}}{\partial_\sigma}+\frac{\partial_Q}{\partial_\sigma}=0$$

$$\frac{\partial_Q}{\partial_\sigma}=\sigma\beta_1(\beta_1-1)>0 \tag{21}$$

$$\frac{\partial_Q}{\partial_{\beta_1}}=\sigma^2\beta_1-\frac{1}{2}\sigma^2+\mu=\sqrt{(\mu-\frac{1}{2}\sigma^2)^2+2\rho\sigma^2}>0$$

$$\frac{\partial_{\beta_1}}{\partial_\sigma}<0$$

And $\frac{\partial_Q}{\partial_{\beta_1}}\frac{\partial_{\beta_1}}{\partial_\rho}+\frac{\partial_Q}{\partial_\rho}=0$, it is not difficult to get: $\frac{\partial_{\beta_1}}{\partial_\rho}<0$.

$\frac{\partial_Q}{\partial_{\beta_1}}\frac{\partial_{\beta_1}}{\partial_\delta}+\frac{\partial_Q}{\partial_\delta}=0$, $\frac{\partial_{\beta_1}}{\partial_\delta}>0$.

As $\rho-\mu=\delta$, the increase of μ would raise P^* when ρ is a constant. With the increase of δ (with ρ fixed) , ρ or σ would raise P^*

When $\mu\geq 0$, $\frac{\partial_{E(\tau^*)}}{\partial_{\beta_1}}<0$

Greater β_1 means smaller E(T), and the flexibility between $E(\tau^*)$ and P^* is unit flexibility: $\frac{\partial_{E(\tau^*)}}{\partial_{P^*}}\frac{P^*}{E(\tau^*)}=1$.

This conclusion is also very anastomotic with the intuition because the smaller threshold can be achieved faster on average. The unit flexibility conclusion indicates that the changing amplitude of E(T) is the same.

When $\mu<0$, the first P_r of P^* is $e^{-2(P^*-P_0)\mu/\sigma^2}$.

The smaller P^* has the greater probability of being hit.

This conclusion means that P^* near the initial position has a greater P_r, it is consistent with the fact.

4.3 Economic and Policy Implication Analysis

(1) It is the same as the additional conclusions of the traditional kind that σ raises the threshold level. This can vividly explain the numerous real estate speculations in the city. In order to be fit for the long-term development of the housing market, the government should carry out the effective macro-control in stabilizing house-price, to

cause price fluctuation returning rationality but blindly indulge the abnormal development of the housing market.

(2) Higher μ means that τ^* has greater probability to be hit as quickly as possible, and this leads to the increase of the optimal critical level. Therefore, the current housing market implicates a destabilizing factor. It is an effective way of stimulating real housing demand, prospering and stabilizing housing market that the government inhibit excessive prices rise.

(3) Higher ρ means the greater opportunity cost of waiting, making the mortgagers demand higher threshold. Therefore, the central bank raising interest rate can effectively curb overheating housing needs. From the partial and short-term point of view, the mortgagers are facing with monthly payment increase or purchasing power decline. But only in this way, the new policy can advise us buying house rationally or deferring, thus easing the supply and demand contradiction, and stabilizing price, which will benefit both our country and individuals. However, it is beneficial to the people from the overall and long-term point of view. As the prices continue to excessive rise, more people can not afford; what is more, once the bursting of bubble impacts on economy, resulting in deterioration of the financial system by the government to pay for(such as Japan), including those who have been buying houses, and everyone has to bear negative consequences.

6 Conclusion

This paper studied the amortized problem when the price of real estate is driven by GBM, and through the analysis of the model, we can draw the following conclsions.

(1) The continuous region of the mortgagers is: $CR = \{P : P < P^*, P^* > K + \theta\}$ and the stopping region is $SR = \{P : P \geq P^*, P^* > K + \theta\}$, which mean that the mortgagers will choose prepayment if $P \geq P^*$, otherwise, they will continue to wait. The variable is presented as:

$$P^* = \frac{\beta_1}{\beta_1 - 1}(K + \theta) \tag{22}$$

$$\beta_1 = \frac{1}{2} - \frac{\beta}{\sigma^2} + \sqrt{(\frac{\mu}{\sigma^2} - \frac{1}{2})^2 + \frac{2\rho}{\sigma^2}}$$

(2) Based on $P^* = F(P^*) + K + \theta$, the mortgagers make the prepayment decision and can get the optimal expectation PV: $V^*(P,t) = \frac{(\beta_1 - 1)^{\beta_1 - 1}}{(\beta_1)^{\beta_1}} \times \frac{(P_0)^{\beta_1}}{(K+\theta)^{\beta_1 - 1}}$, where P_0 is the price at the zero moment.

(3) P^* is not always achieved in the limited time and its reachability depends on the sign of s. If s>0, P^* can be achieved in the limited time; when s<0 (s=0) , P^* will be achieved by the probability $P_r = (\frac{P^*}{P_0})^{\frac{2\mu}{\sigma^2}-1}$ ($P_r = 1$) , but its E(T) is infinity.

(4) The increase of δ (with ρ fixed) 、ρ or σ would raise P^*.

(5) Greater β_1 means smaller E(T), and the flexibility between $E(\tau^*)$ and P^* is unit flexibility.

(6) The smaller P^* has the greater probability of being hit.

The study provides a novel method for the real estate investment decision and it has guiding significance to other related fields. The results of the empirical research will be studied further.

Acknowledgment

The project was supposed by the Fund of Teaching Research project of Hebei University of Engineering, Fund of Doctoral Degree of Hebei University of Engineering and Fund of Youth's Academic Backbone of Hebei University of Engineering.

References

1. Titman, S.: Urban Land Prices under Uncertainty. American Economic Review 75(3), 505–514 (1985)
2. Hendershott, P.H., Turner, B.: Estimating Constant-quality Capitalization Rates and Capitalization Effects of below Market Financing. Journal of Property Research 16(2), 109–122 (1999)
3. Capozza, R., Li, Y.: The Intensity and Timing of Investment: the Case of Land. The American Economic Review 84(4), 889–903 (1994)
4. Grenadier, S.: The Strategic Exercise of Options: Development Cascades and Overbuilding in Real Estate Markets. Journal of Finance 51(5), 1653–1679 (1996)
5. Leishman, C., Jones, C., Fraser, W.: The Influence of Uncertainty on House Builder Behavior and Residential Land Values. Journal of Property Research 17(2), 147–168 (2000)
6. Brandão, L.E., Dyer, J.S.: Decision Analysis and Real Options: A Discrete Time Approach to Real Option Valuation. Annals of Operations Research 135(1), 21–39 (2005)
7. Wu, F., Li, H.Z., Chu, L.K., Sculli, D.: An Approach to the Valuation and Decision of ERP Investment Projects based on Real Options. Annals of Operations Research 168(1), 181–203 (2008)
8. Downing, C., Wallace, N.: A Real Options Approach to Housing Investment. Journal of Real Estate Finance and Economics 40(4), 220–241 (2001)

9. Hendershott, P.H., MacGregor, B.D.: Investor Rationality: Evidence from U.K. Property Capitalization Rates. Real Estate Economics 27(4), 299–322 (2005)
10. Klebaner, F.C.: Introduction to Stochastic Calculus with Applications. Imperial College Press, UK (1998)
11. Song, J., Rhys, H., Jindrichovska, I.: The Timing of Real Option Exercise: Some Recent Developments. The Engineering Economist 47(4), 436–450 (2002)
12. Hans, L.: Price Bubbles in Housing Markets: Concept, Theory and Indicators. International Journal of Housing Markets and Analysis 2(1), 78–90 (2009)

Risk Evaluation of Business Continuity Management by Using Green Technology

Chen Gang

School of Economic and Management, Tongji University Shanghai P.R. China
Shanghai University of Finance and Economic
gladychen@126.com

Abstract. IT disasters can be seen as the test of the ability in communities and firms to effectively protect their information and infrastructure, to reduce both human and property loss, and to rapidly recover. In this paper, we use a literature meta-analysis method to identify potential research directions in Green Business Continuity Management (GBCM). The concept and characteristics of GBCM are discussed. We analysis the connotation and the sources of green technology risk. An assessment index system is established from the perspectives of GBCM. A fuzzy comprehensive assessment method is introduced to assess the risks of green technology in Business Continuity Management.

Keywords: Green Technology, Business Continuity Management, Fuzzy Comprehensive Assessment

1 Introduction

Contemporary environmental problems call for ecology-friendly technology and green technological innovation. Fundamentally, green technological innovation is the core approach to realize the sustainable development strategy [1]. Green technological innovation gradually becomes one of research emphases of the technology innovation area due to its characteristics of accordance with ecological rules and economic rules, and it has gained much attention in the academic, the enterprise and the government literatures. Because of the existence of external positive effect and market failure, green technological innovation has the characteristics of high investment and high risks. Therefore, strengthening the research on the risk evaluation of green technological innovation is of great significance to promote the development of enterprise's green technology innovation.

Business continuity management (BCM) has become a hot topic in recent years and more significantly. More and more green technologies have been used in Business continuity management and there is becoming a new concept -Green Business Continuity Management (GBCM).

The studies on green technology emerge in endlessly. The researchers make research mainly based on the connotation, necessity, dynamic mechanism and system factors of green technological innovation. Vieki pointed out that green technological innovation

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 86–92, 2010.

played an important role [2]. Adam B J and Carolyn F analyzed the interaction between environmental policy and technological innovation [3]. Chen et al, put forward an audit indexes with measuring function of green technological innovation [4]. But all of the research has not involved the risk evaluation of green technological innovation. Therefore, in order to evaluate technological innovation activities of Green Business Continuity Management, and effectively improve the management level of enterprise's green technological innovation [5]. Green technology in Business Continuity Management is the process during which the advanced technology with some characteristics is spread and transferred among Business Continuity Management in a certain way when it is affected by some green technological environment or a special environmental policy.

For these reason, the problem of green technology risk in Business Continuity Management is examined [6]. At first, the connotation and the sources of green technology risk are analyzed. Then, an assessment index system is established from the perspectives of the subject of green technology in Business Continuity Management. At last, Fuzzy Comprehensive Assessment method is introduced to assess the risks of green technology in Business Continuity Management.

2 Sources of Green Technology Risk

An advanced technology which is spread and used widely among the latent user of Business Continuity Management will improve the technical level of the participants and will be helpful to enhance the technical in Business Continuity Management. However, not all the advance technology can be spread and transferred in Business Continuity Management.

In Business Continuity Management, the subjects of technology demand have more information about their resource, their absorbency and the cooperative degree than the subjects of technology supplier. At the sane economy and interest, subjects of technology demand will have immoral conduct, which will bring about moral risk, benefits. In addition, the know-how, the environmental consciousness, the staffs of subject of technology demand for innovative green technology are also risk source which have some influences on Business Continuity Management.

Owing to business continuity is under the menace of being broken off frequently and the severe results. Risk Evaluation of Green Business Continuity Management is becoming an important research consideration of Business Continuity Management. How to evaluate green technology risk in Business Continuity Management has become a critical and thorny problem, which has attracted increasing attention of many researchers and practitioner. Today, it is vital that board members and senior executives understand the nature and scope of BCM. They need to be in a position to evaluate and enhance the status of the activity within their organizations.

Green technology transformation depends on government's drive in environmental management, which is different from the technology transformation of traditional Business Continuity Management. The running cost of Business Continuity Management will rise for adopting effective environmental measures. When the government's environmental policy becomes less strict and when the tax and fine, which is levied on the enterprises for environment pollution is lower than the cost which was increased for adopting green technology, the enterprises will give up adopting green technology and even hinder from transferring and diffusing green technology. At present, environmental policies in our country lacks of effective incentive mechanism, especially effective economic measures to promote enterprises to adopt green technology. Sewage charge, as a primary economic incentive measure, makes enterprises pay the charge for disposing pollutants rather than invest in the activities of green technology innovation and transformation due to the low tax and fines, and charge only for some single factors.

3 Assessment Indexes of Green Technology Risks in Green Business Continuity Management

However, the theoretical system of Green Business Continuity Management (GBCM) has not come into being; there are even less quantitative researches on GBCM owing to the concept of GBCM being put forward too late. In practice, there are only a few famous companies implementing Green Business Continuity Management (GBCM). As for the risk of green technology in Green Business Continuity Management, there are even less researches about it. Therefore, it is a realistic and theoretic task, which needs urgent solution, how to carry out researches on assessment of green technology risk in Green Business Continuity Management.

3.1 Concept of Green Business Continuity Management

It is referred to the consciousness of the importance and feasibility of green technology and pollution prevention. Scarcity of environmental consciousness shows in the following. Firstly, the notion that environmental cost is inevitable in order to develop economy. Secondly, be accustomed to the status quo and be instinctively paradoxical to new technology. Thirdly, lack of cognition for increased environmental revenue and overall revenue of Green Business Continuity Management, which is brought about by green technology. Fourthly, ambiguous environmental criteria and cockamamie report requisition among Business Continuity Management participants lead to communicative conflict. Therefore, lack of environmental consciousness is the factor which will lead to some latent risks.

It is referred to the degree and the direction of technical progress in a society, including the degree and progress in new products, process, raw material and the basic discipline [7-9]. This is an exterior decisive factor that will decide the prospects of a new technique. Furthermore, its influence will filter into every stage of technical diffusion. Thereby, not only the technical environments in society but also the technical level should be considered during the course of technical transformation. Green Business Continuity Management requires the knowledge, know-how and the practical experience about environmental control be quickly transferred in Green Business

Continuity Management. However, some factors, like recognition deviation in technical environment, scare of consciousness of technical innovation protection and variance of Green Business Continuity Management participants in knowledge and technology, will beget technical information risks in a Green Business Continuity Management.

3.2 Indexes of Green Technology Risks in GBCM

The index weight is the quantitative performance which reveals the relative importance off each index in the whole index system. Whether the index weight is reasonable or not will affect the comprehensive evaluation results. In this paper, the combining method of AHP and information entropy is used to combining the weighted evaluation indexes with Weighted Summation. When computing the weight, AHP method emphases on expert's subjective experience and judgments, while information entropy method emphases on objective information of the system in the data [10]. Therefore, combining the two kind of method to computing the weight of evaluation index can reflect the actual situation of the problems more objectively and comprehensively.

Table 1. Risk assessment indexes on the green technology in Green Business Continuity Management.

Risk Assessment Indexes on the Green Technology in GBCM	Political risk	Moral risk (S1) Risk of law(S2)
	Technical risk	Green Technical environment(S3) Green technological advancement (S4) Green technological life cycle(S5)
	Financial risk	Green technological investment(S6) Tax and fine risk(S7)
	Management risk	Rationality of organization(S8) Administrator quality and decision-making ability(S9) Worker training and re-education(S10) Environmental consciousness(S11)

The comprehensive weight is obtained through the combining method of information entropy and AHP,

$W=(w_1,w_2, w_3,w_4,w_5,w_6,w_7,w_8,w_9,w_{10},w_{11})$

$=(0.084,0.052,0.116,0.088,0.092.0.044,0.056,0.125,0.095,0.143,0.105)$

Each evaluation index represents the characteristic attribute of different projects and different evaluation index has different dimension.

4 Fuzzy Comprehensive GBCM

The assessment of green technology transformation risk in Green Business Continuity Management is a multi-level and multi-factor system. Fuzzy Comprehensive Assessment method evaluates the green technology using in Green Business Continuity Management.

The data of Green Business Continuity Management is from questionnaire investigated. Using triangle fuzzy function, every index has degrees such as P= (P_1, P_2, P_3, P_4 ,…P_m). For example, if index has sever degree= (1, 2, 3, 4, 5, 6, 7,). 1means very unimportant, 2 means less important, 3 means unimportant, 4 mean average, 5 means important, 6 means more important, 7 means very important. The triangle fuzzy function of index is $\left(\frac{j-1}{m}, \frac{i}{m}, \frac{j+1}{m}\right)$, j is the counts of index aggregate, m is the count of last element of index aggregate. Here M is 7. If the value of index is less important, then J=2, and the triangle fuzzy function is (1, 2, 3).

Table 2. Triangle fuzzy function of indexes value

Indexes Value	1	2	3	4	5	6	7
Triangle Fuzzy Function Fi=(a,b,c)	(1,1,2)	(1,2,3)	(2,3,4)	(3,4,5)	(4,5,6)	(5,6,7)	(6,7,7)

Using median method, we can get value of triangle fuzzy function,

$$W_i=\frac{W(Fi)}{\sum_{i=0}^{n} W(Fi)} \tag{1}$$

$$W（Fi）=(b + c- a) /3 \tag{2}$$

Using vertex method we can get value index Y as following

P= [P_{ij}] is value aggregate, W= [W_i] is right of every value. So the value of every index is $P^w_{ij0}=w_{j0}p_{ij0}$, $P^w_{ij1}=w_{j1}p_{ij1}$, $P^w_{ij2}=w_{j2}p_{ij2}$, ……..

X=(x_1,x_2,x_3) and Y=(y_1, y_2, y_3) is triangle fuzzy function of two index, so interval of these tow indexes is $d(x, y))=\{[(x_1-y_1)^2+(x_2-y_2)^2+(x_3-y_3)^2]/3\}^{1/2.}$ For example, The best value of index is P*j=(1, 1, 1), the worst value of index is P^{-}j=(0, 0, 0). Interval of two indexes is:

$$d^{*}_{i}=\sum_{j=1}^{n}d(P^{w}_{ij},P^{*}_{j})\quad \text{i=1,2,....m, j=1,2,...n} \tag{3}$$

$$d^{-}_{i}=\sum_{j=1}^{n}d(P^{w}_{ij},P^{-}_{j})\quad \text{i=1,2,....m, j=1,2,...n} \tag{4}$$

Validity of index is

$$s_i=d^{-}_{i}+n-\frac{d^{*}_{i}}{2n},\quad \text{i=1,...m,} \quad \text{n is the number of index.} \tag{5}$$

We can get the value index Y as following

$$Y=\sum_{i}w_i s_i \tag{6}$$

From questionnaire investigate we get the data and deal them with the method above, we get

F_1= (0.1876, 0.2047, 0.2168)
F_2= (0.1834, 0.1992, 0.2067)
F_3= (0.1217, 0.1305, 0.1374)
F_4= (0.1399, 0.1501, 0.1592)
F_5= (0.1867, 0.1987, 0.2047)
F_6= (0.1053, 0.1168, 0.1201)

S_0=7.9645 ; S_1=7.2176 ; S_2=5.8152 ; S_3=6.1452 ; S_4=7.3622 ; S_5=7.1645 ; S_6=5.5218

Y=6.9823

5 Conclusions and Recommendations

After analyzing the indexes of risk of green technology which Green Business Continuity Management using, this paper construct a multiple factors assessment index system of green technology risk in terms of the fuzziness of green technology risk in Green Business Continuity Management. Furthermore, the validity of this method is examined by analyzing an instantiation. It is helpful for risk management to Green Business Continuity Management and also has an important referenced value for the risk evaluation of other areas.

Acknowledgment

The work is supported by Leading Academic Discipline Program, 211 Project for Shanghai University of Finance and Economics (the 3rd phase).

References

1. Singh, V.P.: The Entropy Theory as a Tool for Modeling and Decision-making in Environment and Water Resources. Water SA 1(26), 1–11 (2000)
2. Tong, H., Zhang, S.: Multi-attribute Group Decision Making Algorithm for Web Service Selection based on QOS. Journal of Southeast University 22(3), 499–508 (2006)
3. Fenton, N., Wang, W.: Risk and Confidence Analysis for Fuzzy Multi-criteria Decision Making. Knowledge-Based Systems 19(6), 430–437 (2006)
4. Li, K., Wei, X., Cai, X.: Diffusion Mechanism of Green Technology in Enterprise Cluser. Finance and Economics Science 7, 81–87 (2007)
5. Guo, F., Dan, B., Zhang, X.: Reach on the Technology Service Center. R&D Management 5, 28–33 (2007)
6. Lv, Y., Sheng, M.: Case Study for Diffusion of Green Technology. Research of Environmental Science 7, 25–27 (1997)
7. Song, G., Yang, X., Zhang, Y.: Study on the Cooperation Risks of Virtual Enterprises. Soft Science 3, 83–86 (2004)
8. Xu, H., Sun, L., Mo, J.: Knowledge Diffusion and Technology Innovation in Virtual Enterprises. Industrial Engineering and Management 1, 26–29 (2003)
9. Huang, W.: Evaluation of Investment Risk in Hydraulic Engineering based on BP Network. Human Light Industry College 2(1), 34–38 (2004)
10. Zhang, J.: Risk Evaluation of Projection using Support Vector Machine. West-China Exploration Engineering 6, 233–237 (2008)

Automated Methodologies for the Design of Flow Diagrams for Development and Maintenance Activities

Handigund Shivanand M. and Bhat Shweta

Department of Computer Science & Engineering, Bangalore Institute of Technology, Bangalore – 560 004, Karnataka, India
{Handigund.Shivanand.M,Bhat.Shweta}@Springer.com

Abstract. The Software Requirements Specification (SRS) of the organization is a text document prepared by strategic management incorporating the requirements of the organization. These requirements of ongoing business/ project development process involve the software tools, the hardware devices, the manual procedures, the application programs and the communication commands. These components are appropriately ordered for achieving the mission of the concerned process both in the project development and the ongoing business processes, in different flow diagrams viz. activity chart, workflow diagram, activity diagram, component diagram and deployment diagram. This paper proposes two generic, automatic methodologies for the design of various flow diagrams of (i) project development activities, (ii) ongoing business process. The methodologies also resolve the ensuing deadlocks in the flow diagrams and determine the critical paths for the activity chart. Though both methodologies are independent, each complements other in authenticating its correctness and completeness.

Keywords: Activity Chart; Activity Diagram; Component Diagram; Critical Path; Dead Lock; Defined Attribute; Deployment Diagram; Organization Workflow.

1 Introduction

The software developer's task begins with the procurement of project charter [11] of client organization. This charter is a legal document containing the details regarding the cost of the project, the schedule and the Software Requirements Specification (SRS) [16] etc. The SRS describes the business procedures, the actors and the resources of the client organization, in addition to the achievement of project mission viz. software tools, hardware devices, manual procedures, application programs and communication commands and their flow dependencies in the realization of the business process.

In the project development process, the activities viz. the software tools utilization, hardware devices utilization, development of application programs, framing guidelines set for manual procedures and gluing through communication commands that bridges the gap amongst the activities. Similar types of activities are present in the ongoing business process except the application programs development activity. The

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 93–104, 2010.

project components are utilized to develop the ongoing business components. In both cases, the components are clustered with different topologies.

An activity of project development process considers the SRS, the project duration, the available resources, the project components & the project cost as input, and prepares an activity chart for the entire project development. Activity chart is a directed graph comprising a start node leading to several paths of intermediate nodes. Each intermediate node comprises a group of any combination of project components with implicit dependencies to be realized within a period to achieve a noticeable milestone in the project development. Each such component group node forms a milestone or deliverable. These nodes are connected by communication command glues as directed edges with arrows directing to nodes to be realized at later stages. The project bridges the gap between the available input and the expected output with the use of human resource and the project components. The critical paths in the activity chart from start to end node are to be identified to facilitate the computation of the least long time of the project.

The project outcome is a software tool or combination of tools that comprise the application programs developed in the project along with the embedded software tools. The outcome of the project development process are used as components in the ongoing business process. The business components are distributed to different work processes with interrelationships framed to realize the business norms. The actors' usecases are realized with dependencies amongst the components spanning one or more work processes. These are to be designed in the form of a workflow diagram in which each business component is represented by a node and the dependency between two components (communication command glue) as a directed edge with dependent node at the head of the arrow. The dependencies involved may be the sequence, the in-hierarchy or the period. The workflow diagram (graph) depicts the complete working of the business process.

A perspective view of the business process comprises number of works. A work may be formed by a sequence of purposeful activities derived from the intersection of work processes with usecases. Thus, it comprises activities of some work processes (physical objects) and involving part of actors & work processes (information objects). Each work is represented in Unified Modeling Language (UML) by an activity diagram in which the activities of each objects are grouped in a swimlane and are connected through transitions (dependencies) spanning one or two swimlanes. Similarly in UML, the interdependencies of the software tools of business process are represented by a component diagram. The combined interdependencies between the software and hardware devices are represented by a deployment diagram.

1.1 Motivation

1.1.1 Project Development

The activity chart, apart from helping in the realization of project activities, provides the following facilities.

1. Estimation of budget and time frame for individual activities groups and entire project.
2. Allocation of hardware, software and other resources for activities of project.
3. Effective utilization of environmental resources like library routines, prototypes and patterns of historical projects.

Other flow diagrams help in the project development for realizing,

1. the distribution of project work to different teams.
2. the interleaving of task allocation to activities for optimal resource utilization,
3. the procurement schedule of different software tools.

1.1.2 Ongoing Business

The activity chart, apart from helping in the realization of ongoing business activity, provides the following facilities.

1. Serves as an aid in the development of ongoing business process vision, mission and objectives.
2. Allocation of hardware, software and other resources for activities of business process.
3. Effective utilization of environmental resources like library routines, prototypes and patterns of historical projects.

Other flow diagrams help in providing additional facilities.

1. Automation of business & work processes.
2. Resolution of deadlocks amongst the activities dependencies.
3. Optimization of the work processes, time and outcomes of the business process.
4. Inter-utilization of facilities between the environment and the system.
5. Isolation of work processes or their parts for different products, services or results.
6. Formation of coherent clusters of activities which minimizes the repetitive creation of information chunks in the flow.

The flow diagram topology design complexity will increase with the growth of component nodes. Manually flow diagrams are designable, but the design process is time consuming and error prone.

1.2 Literature Survey

Arguably, there is a practical interest in both project development and ongoing business processes [9]. The advancement of software project management transformed the project development into a technology. This is evident from the constant up-gradation of project management techniques by Project Management Institute (PMI) [11]. To reap the benefits of this technology, there is a necessity of automating the ordering of project processes. Analogously for ongoing business process, the scientific interest is raised because, managing the processes is notoriously difficult. Researchers have opined that the flow design depends on the complexities of resource sharing & business policies [8] which results in forming a best design a paranoia.

Kwan & Balasubramanian [5] studied the project development & business management factors and identified different flows as dominant factors, whose small violation in derived constraints will mow down the good workflow design [12]. All these researchers have concentrated mainly on modeling, verification & architectural issues. They kept the design of the workflow outside the scope [8]. Sun et al [14] have proposed a dataflow dependent workflow design method, ignoring the references and definitions of data jeopardizing its correctness & completeness. Nanshan Du et al [15] proposes a data-flow skeleton filled with activities driven workflow design method, that uses dependencies to deduce a data flow skeleton and then fills with activities to form workflow process. This enhances the complication.

Vladimir A. Shepelev et al [4] have attempted to develop task schema approach for the design of workflow in which the data required for the tasks are abstracted through a complex process. The details of which are not available in the published paper. This task schema approach is analogous to our proposed first cut graph, but in the absence of clear cut methodology, the method may have limited use.

1.3 Taxonomy

Activity Chart. In project development, the activity chart is a directed graph in which each activities group is depicted as milestone in paths from start to end. Each milestone is labeled with time duration. The edges represent the communication commands between the milestones along the paths from start to end. Some paths may indicate the critical paths (the least longest time) for the project development. In ongoing business process, the activity chart depicts the flow order of chunks of business processes formed by the intersection of the use-cases and the work processes.

Activity Diagram. In Unified Modeling Language (UML) [6], the activity diagram depicts the flow of activities for a work. These activities are streamlined with swimlanes of physical or information objects. The semiotics of this diagram is designed with UML norms and is helpful in identifying the operations, the object classes, the work processes and the use-cases.

Component Diagram. In UML, the component diagram depicts the dependencies between the software tools and/or the application programs. It will help in identifying the inter stage resource allocation within the business process.

Critical Path. It is a path with least longest time in the activity chart. There may be one or more critical path/s that take the same time. It helps in identifying the project duration or the product duration.

Dead Lock. In sequencing the various components, the directed graph may contain cycle/s indicating the occurrence of a deadlock/s, which is/are undesirable.

Defined Attribute. The defined attribute is the data item whose value is modified in the realization of the statement.

Deployment Diagram. In UML, this diagram contains the dependencies amongst the software tools & the interleaved hardware devices used in the project development and business process.

Organization Workflow. The information system of the organization utilizes the software tools, hardware devices, manual procedures, application programs and communication commands. These have to be appropriately sequenced in project development as well as business process. This facilitates the optimal allocation of resources.

2 Proposed Methodologies

In this paper, we propose two methodologies for the design of flow diagrams viz. activity chart, workflow diagram, activity diagram, component diagram and deployment diagram [10], which can be applied to both project development process and ongoing business process. These methodologies take the activities & their dependencies as input and is represented as entries in the table. The proposed methodologies are generic in nature to facilitate the mapping of a dependency table to another flow diagram by just ignoring certain class of components.

2.1 Sifting of Components

The *activity chart* comprises the activities and their connecting glues along with their time schedule. Here, the activity stage is the sequence of dependent components to be realized to complete a chunk of work, indicative of a noticeable stage (milestone) in the project. Further, if the noticeable stage delivers a result in the form of document, the milestone becomes deliverable. Here, each node indicates a group of components as a unit based on their dependency and realization in a certain time frame. The grouping of activities into a milestone node is application dependent. Each node may consume & produce more than one input/ output. The sample milestone unit activity comprises a sequence of activities as shown in the figure 1.

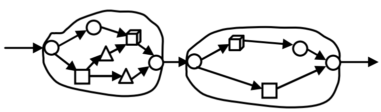


Fig. 1. Activities unit in the activity chart

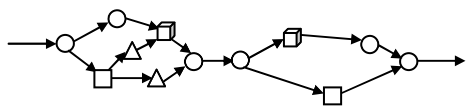


Fig. 2. Activities in the workflow diagram

In activity chart, based on the time required for realization of individual implicit components of the node, the time period required for activity node unit can be computed. Thus, the component units form sequences of flow containing one or more paths from start to end. The least time required for the completion of the project is the longest path from start to end called critical path in the activity chart.

The components of the *workflow* in the project development or in ongoing business process are as specified in the introduction. In the ongoing business process workflow, the components are distributed amongst different work processes and number of work processes may concurrently consume and/or provide the information from/to appropriate actors. Thus, there exists the number of sources & sinks (actors). Normally, in business organization there is a single line source and a line sink to liaise between the environment (actors) & the system (components) for consuming and/or providing information. The ensuing workflow diagram may contain a start node and an end node. Since these nodes are borrowed nodes that couple components with actors, they along with their explicitly connected edges can be discarded in the ensuing workflow diagram. The notations in the Fig.s 1 and 2 are symbolic to indicate the syntactic identity for components.

The software tool within a node of activity chart is used either as external facilitator or as consumer for the program development. Thus, the software tool in the workflow means the outcome application program of the project development. In addition, the flow order of components in workflow is dependent on business norms & the requirements of the environment, whereas in activity chart, it is dependent on project development stages. Except the workflow, other flow diagrams contain the explicit or implicit presence of the start and/or end node.

In *activity diagram,* the activities are the action states derived from combination of some software tools and manual procedures for the required outcome. The outcome is mapped to another activity through the communication command and is represented by a transition. The work activities are partitioned into swimlanes of physical or information objects. Each swimlane contains activities invoked by the appropriate object. The analogy between the activity diagram and the work flow enabled the use of workflow design methodology for the initial part design. In the later part, the nodes are realigned into swimlanes without altering the transitions amongst the nodes. The activities are partitioned into object swimlanes depending on the attributes they transit. If the defined attributes [3] are from an object class, the activity belongs to the swimlane reserved for that physical object. On the other hand, if the defined attributes are from two object classes, the activity commonly belongs to swimlanes reserved for these physical objects. In such case, it is desirable to allocate the adjacent swimlanes to object classes to cover their common activity in the appropriate swimlanes. If the activities are jointly from a subgroup of actor/s and/or object classes then, the activities partitioned to an information object swimlane. If the result of the activity is flows through two or more transitions, these are considered as concurrent transitions derived from a single transition. They are linked through a fork. If the numbers of concurrent transitions are incident to a single activity, the concurrent transitions are linked to a single transition bridged by a join. The communication command indicating the conditional flow of transition (if any) is considered as either alternating transition emanating from a single transition or alternations merging to a single transition. In both cases, the junction is represented by a diamond. The UML syntactic start and end symbols are augmented to the diagram based on invoking & revoking the transition/s.

The *component diagram*, is same as workflow except the following slight automated modifications. The component nodes of workflow representing hardware devices & manual procedures are depleted to a single point, so that depleted point cojoins the incident & outdance edge/s to form a continuous edge from predecessor node/s to successor node/s. In the resultant diagram, each node is replaced by syntactic UML notation.

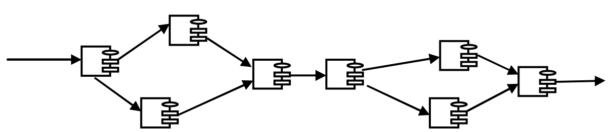


Fig. 3. Components in the component diagram

The *deployment diagram*, is analogous to workflow except the following slight automated modifications. The nodes presenting manual procedure are depleted to a single point, so that depleted point co-joins the incident & outdance edge/s to form a continuous edge from predecessor node/s to successor node/s. In the resultant diagram, enclose each node in a transparent rectangular box (UML notation for the deployment diagram components).

2.2 Preamble to the Methodologies

Each of the proposed methodologies works with input data mentioned in the dependency table [13] of entries, each of which contains the component type, component name and the determining components. The methodologies generate the desired flow diagrams as output. Here, the methodologies need not be used afresh for each flow diagram. Instead a flow diagram is generated and with slight modifications other diagrams can be generated. For a particular sequence of generating flow diagrams from workflow, the techniques are indicated in section 2.1.

Table 1. Dependency table

Type	Activity	Dependent On
Manual procedure	A1	-
Software tool	A2	A1
Manual procedure	A3	A12
Software tool	A4	A2, A8
Application program	A5	-
Software tool	A6	A11
Manual procedure	A7	A6
Application program	A8	A5
Application program	A9	A5
Application program	A10	-
Manual procedure	A11	A4, A9
Software tool	A12	A9, A10, A11
Application program	A13	A3

2.3 Methodology - 1

1. Compute out-degree of each component node.

$$\text{Out-degree} = \text{Number of occurrences of the component in the "Dependent On" column} \tag{1}$$

2. Compute in-degree of each component node.

$$\text{In-degree} = \text{Number of components present in the "Dependent On" column of the entry} \tag{2}$$

3. Compute effective degree of each component node.

$$\text{Effective degree} = \text{Out-degree} - \text{In-degree} \tag{3}$$

The results computed using equations 1, 2 & 3 are entered in the appropriate entries of Table 2.

Table 2. Dependency table with computed out-degree, in-degree & effective degree

Type	Activity	Dependent On	Out-degree	In-degree	Effective degree
Manual procedure	A1	-	1	0	1
Software tool	A2	A1	1	1	0
Manual procedure	A3	A12	1	1	0
Software tool	A4	A2, A8	1	2	-1
Application program	A5	-	2	0	2
Software tool	A6	A11	1	1	0
Manual procedure	A7	A6	0	1	-1
Application program	A8	A5	1	1	0
Application program	A9	A5	2	1	1
Application program	A10	-	1	0	1
Manual procedure	A11	A4, A9	2	2	0
Software tool	A12	A9, A10, A11	1	3	-2
Application program	A13	A3	0	1	-1

4. Sort the table entries in descending order of effective degree.

Table 3. Sorted dependency table

Type	Activity	Dependent On	Out-degree	In-degree	Effective degree
Application program	A5	-	2	0	2
Manual procedure	A1	-	1	0	1
Application program	A9	A5	2	1	1
Application program	A10	-	1	0	1
Software tool	A2	A1	1	1	0
Manual procedure	A3	A12	1	1	0
Software tool	A6	A11	1	1	0
Application program	A8	A5	1	1	0
Manual procedure	A11	A4, A9	2	2	0
Software tool	A4	A2, A8	1	2	-1
Manual procedure	A7	A6	0	1	-1
Application program	A13	A3	0	1	-1
Software tool	A12	A9, A10, A11	1	3	-2

5. Design flow diagram through following algorithmic steps. Repeat the step(6) through step(9) until all "Dependent On" column entries are marked as deleted.
6. Assign the top most unmarked component D(i, j) in the "Dependent On" column to a variable ComTobeDrawn where, i is the row index, j is the component position in i^{th} row.
7. Mark the component D(i, j) as deleted in the dependency table and set a flag to "Activity".
8. Repeat the step (i) through step(iv) until ComTobeDrawn is not null

i. Draw the component in ComTobeDrawn with appropriate syntactic notation connecting the previously drawn component node (if exists) through an edge such that the newly drawn component is at arrow head.
ii. If the flag is set to "Activity" then, assign ComTobeDrawn with A(i), where A(i) is the component in i^{th} row of “Activity” column and mark the component A(i) as deleted along with setting the flag to "Dependent On".
iii. Else if the flag is set to "Dependent On" and the top most unmarked matching component D(j, k) with ComTobeDrawn is found then, assign A(j) to-ComTobeDrawn and mark the D(j, k) & A(j) as deleted along with setting the flag to "Dependent On".
iv. Else if the flag is set to "Dependent On" and match the top most unmarked matching component D(j, k) with ComTobeDrawn is not found, assign null to ComTobeDrawn.

9. Draw remaining unmarked components in "Activity" column with appropriate syntactic notations.
10. Connect the "START" node to all component nodes with zero in-degree and connect all component nodes with zero out-degree to "END" node.

In Fig.4, each symbol denotes a specific type of component.

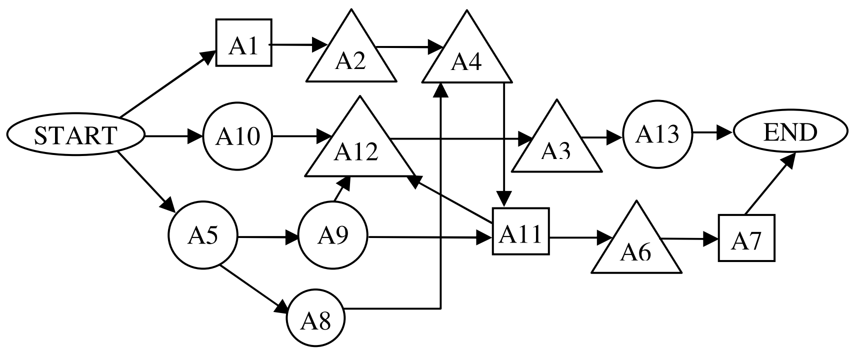


Fig. 4. Resultant flow diagram

2.4 Methodology - 2

Alternatively, the workflow can be designed for the dependency components using the following methodology also. The steps involved are as follows.

1. Represent each component by a node & draw a draft graph connecting each node to the node/s it determines, by a directed edge.
2. Identify the degree of each node by subtracting the in-degree from the out-degree of the component node.
3. Design a Depth First Search (DFS) [1], [2] table with three rows and the columns equal to number of component nodes of the dependency table. Store the component names in each column of the first row in the decreasing order of their calculated degrees. The second row & third row contains the numbers of the consecutive order of visiting nodes and processing nodes through edges respectively.

4. Start with the first column entry component node of the table, visit each node if there is a path in the draft graph, entering with consecutive higher number in the appropriate column of the visit row and marking the edge of the graph as visited.
5. Repeat the step (4) so long there exists a path in the draft graph.
6. Process the node entering higher consecutive number for the node in the appropriate column of process row and back track to the immediate preceding node discarding the edge in the draft graph.
7. Repeat the step (5) with the immediate preceding node as the start node.
8. Continue this procedure till all edges in the draft graph are discarded.

The DFS table for the illustrative example is as in Table 4.

Table 4. DFS table

A5	**A1**	**A9**	**A10**	**A2**	**A3**	**A6**	**A8**	**A11**	**A4**	**A7**	**A13**	**A12**
1	25	2	31	26	9	4	18	3 20	19 27	5	10	8 15 32
24	30	17	34	29	12	7	23	14 21	22 28	6	11	13 16 33

9. Draw an edge from the start node to the node containing highest value in the process row. Identify the next lower consecutive numbered node. If the number is in the column of process row, draw a forward edge from the last drawn node to the corresponding column node. If the number is in the column of visit row, back track to the immediate preceding drawn node.
10. Continue this process till all the consecutive numbers are navigated in the decreasing order.
11. Connect the nodes with zero out-degree to an additional end node.

The method-2 also designs the same flow diagram (Fig. 4.) as designed in method-1. The correctness and completeness of the two methodologies is authenticated by the uniqueness in the result of the two methodologies.

2.5 Technique for the Resolution of the Deadlock

In the flow diagrams, we have considered the issue of deadlock resolution for components which are sequence dependent and in-hierarchy dependent as suggested in figures 6 & 7. The deadlock resolution for period dependency (Fig.5) could not be considered as it leads to a three dimensional figure. Moreover, the component of higher frequency period takes the input as the aggregate function values of lower frequency period components. Here, the aggregate periods and the frequency layers are all information system specific. However attempt is made to illustrate the period dependency with a three dimensional figure considering the sample frequency periods (Fig. 5). When the execution of a component node need the result of the other and vice versa, the only solution [7] is to replicate the component node that creates the deadlock situation.

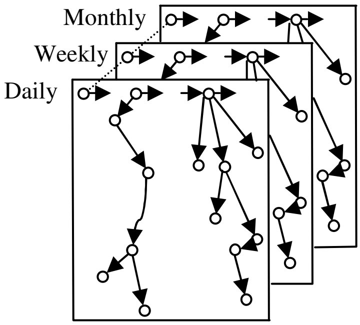


Fig. 5. The period dependency

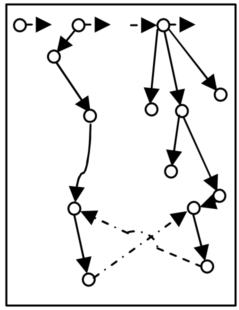


Fig. 6. Deadlock situation

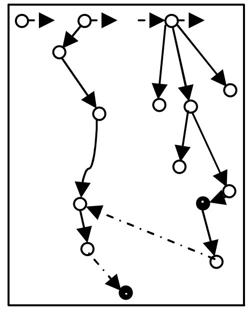


Fig. 7. A resolution way

3 Conclusion

An attempt is made to develop automatic, generic methodologies for the design of activity chart for project development process and ongoing business process. The other flow diagrams can be generated using the any of the proposed methodologies either independently or with slight modifications in the generated flow diagram. The proposal of two methodologies helps in authenticating the correctness and completeness of the each other in a natural way. The proposed methodologies are automatic and flexible to tailor to the needs of other diagrams. The critical path computation of the activity chart is not discussed in the proposed methodologies, as there exist number of algorithms to compute the longest paths from a single source to a destination in the directed flow graph. The proposed technique has not explicitly discussed the resolution of deadlock as the solution mentioned in section 2.4 is only a suggestive guideline.

Acknowledgments. The authors gratefully acknowledge the assistance of Dr. D. B. Phatak of IIT Bombay in the paper structure formulation and the All India Council of Technical Education (AICTE) for the grant under Research Promotion Scheme through its F.NO.: 8023/ BOR/ RID/ RPS-99/ 2007-08.

References

1. Tarjan, R.: Depth-First Search and Linear Graph Algorithms. SIAM Journal on Computing 1(2), 146–159 (1972)
2. Horowitz, E., Sahani, S.: Fundamentals of Data Structures. Galgotia Book Source P. Ltd., Gurgaon (1978)
3. Atkins, J., Jon, B.: A Survey on Program Slicing for Software Engineering. In: Technical Report, Deptt. Statistics and Computer Science, West Virginia University, West Virginia (April 1993)
4. Shepelev, V.A., Director, S.W.: Automatic Workflow Generation. In: Proceedings of the Conference on European Design Automation, Geneva, Switzerland, pp. 104–109 (1996)
5. Kwan, M.M., Balasubramanian, P.R.: Dynamic Workflow Management: A Framework for Modeling Workflows. In: Proceedings of the Thirtieth Hawaii International Conference on System Sciences, Maui, HI, USA, vol. 4, pp. 367–376. IEEE Computer Society Press, Los Alamitos (January 1997)

6. Eriksson, H.-E., Penker, M.: UML Tool Kit. Wiley Computer Publishing, John Wiley & Sons Inc. (1998)
7. Handigund, S.M., under the guidance Phatak, Dr. Deepak, B.: Reverse Engineering of Legacy COBOL Systems, Ph. D. Thesis, Indian Institute of Technology Bombay (2001)
8. Stohr, E.A., Zhao, J.L.: Workflow Automation: Overview and Research Issues. Information Systems Frontiers (3: 3), 281–296 (September 2001)
9. Hajo, A., Reijers: Design and Control of Workflow Processes: Business Process Management for the Service Industry: Technische Universiteit Eindhoven (2002)
10. Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., Lorensen, W.: Object-Oriented Modeling and Design. Prentice-Hall of India Private Limited, New Delhi (2003)
11. Duncan, W.R.: A Guide to the Project Management Body of Knowledge. In: PMI Standards Committee, Newtown Square, USA, pp. 50–51 (2004)
12. Sun, S.X., Zhao, J.L.: A Data Flow Approach to Workflow Design. In: Proceedings of the 14th Workshop on Information Technology and Systems, Washington D. C., USA, pp. 80–85 (December 2004)
13. Pressman, R.S.: Software Engineering a Practitioner's Approach, 6th edn. McGraw Hill Higher Education, New York (2005)
14. Sun, S.X., Zhao, J.L.: Activity Relations: A Dataflow Approach to Workflow Design. In: Proceedings of the 27th International Conference on Information Systems, Milwaukee Wis., USA, pp. 661–671 (December 2006)
15. Du, N., Liang, Y., Zhao, L.: Data-Flow Skeleton Filled with Activities Driven Workflow Design. In: Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication, Suwon, Korea, pp. 570–574 (2008)
16. Jalote, P.: An Integrated Approach to Software Engineering, 3rd edn. Narosa Publishing House, New Delhi (2008)

ACO-Based Projection Pursuit: A Novel Clustering Algorithm

Yancang Li*, Lina Zhao, and Shujing Zhou

College of Civil Engineering, Hebei University of Engineering, Handan056038, China
liyancang@163.com

Abstract. In order to find an effective method of solving the problem of subjectivity and difficulty in the high-dimension data clustering, a new method---an improved Projection Pursuit based on Ant Colony Optimization algorithm was introduced. The ant colony optimization algorithm was employed to optimize the function of the projected indexes in the PP. The ant colony optimization algorithm has the strong global optimization ability and the PP method is a powerful technique for extracting statistically significant features from high-dimension data for automatic target detection and classification. Application results show that the method can complete the selection more objectivity and rationality with objective weight, high resolving power, and stable result. The study provides a novel algorithm for the high-dimension data clustering.

Keywords: Clustering Algorithm, Projection Pursuit Algorithm, Ant Colony Optimization Algorithm, High-dimension Data, Improved.

1 Introduction

As a problem that needs urgent solution in clustering problems, it is more and more important to find an effective method to deal with the high-dimension data. This has signified the need for the related study. Up to date, by the attention and efforts of researchers in corresponding fields, many algorithms and models were proposed. Most of these algorithms and models are based on the knowledge of other new subjects, frontier subjects and cross subjects, such as fuzzy set method, grey set method, ANN (artificial neural network) selection, principal component analysis method, matter-element analysis, multivariate statistical analysis, k-median clustering and so on[1-7]. In some degree, all these studies give rise to the clustering. But, most of the methods employed have their own shortcomings. For example, the fuzzy set comprehensive selection cannot meet the nonnegative, additive principle and the convergent principle. And most of the models above have the shortcomings of subjectivity. There is much more to be done in this direction. The situation has signified the need for related research.

PP (Project Pursuit) is concerned with “interesting” projections of high dimensional data sets, with finding such projections by machine, and with using them

* Corresponding author.

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 105–113, 2010.

for nonparametric fitting and other data-analytic purposes. It emerges as the most powerful method to higher dimensions. The most exciting feature of PP is that it is one of the very few multivariate methods able to bypass the “curse of dimensionality” caused by the fact that high-dimensional space is mostly empty [8]. In addition, the more interesting PP methods are able to ignore irrelevant (i.e. noisy and information-poor) variables. This is a distinct advantage over methods based on interpoint distances like minimal spanning trees, multidimensional scaling and most clustering techniques[9]. PP methods have one serious drawback: diffict to optimize the function of the projected indexes.

To overcome this shortage, an Ant Colony Optimization (ACO) algorithm was introduced. ACO is a group of novel population-based algorithms that combine distributed computation, autocatalysis (positive feedback) and constructive greedy heuristic in finding optimal solutions for NP-hard combinatorial optimization problems. ACO has been successfully applied to most combinatorial optimization problems, e.g. TSP (Traveling Salesman Problem), JSP (Job-shop Scheduling Problem), QAP (Quadratic Assignment Problem), SOP (Sequential Ordering Problem) and so on [10]. Here, the information entropy based ACO proposed in [11] was employed to solve the problem of the projected indexes optimization.

The paper is organized as follows. In the following part, attention was paid to the generalities of the PP and ACO. Then, the clustering model based on the improved PP was set up. Finally, its application in practice was introduced and the advantages of the method we proposed were pointed out.

2 Basic Knowledge of PP and Information Entropy Based on ACO

2.1 Generalities of the PP

PP techniques were originally proposed and experimented with by Friedman, J. H., and Tukey, J. W, [12]. The original purpose of the PP was to machine-pick “interesting” low-dimensional projections of a high-dimensional point cloud by numerically maximizing a certain objective function or projection index. The most exciting feature of PP is that it is one of the very few multivariate methods able to bypass the “curse of dimensionality” caused by the fact that high-dimensional space is mostly empty. In addition, the more interesting PP methods are able to ignore irrelevant (i.e. noisy and information-poor) variables. This is a distinct advantage over methods based on interposing distances like minimal spanning trees, multidimensional scaling and most clustering techniques. PP emerges as the most powerful method to lift one-dimensional statistical techniques to higher dimensions. The PP methods have been successfully applied in many fields [13-17].

The definitions of PP are as follows [8]:

For letter X is an n-tuple of points $\left(x_1, x_2, \cdots, x_n\right)$ in $\Re^d$. By ave$\{X\}$we shall equally, indiscriminately denote either the sample mean $\frac{1}{n}\sum x_i$ or the expectation $E(X)$.

A linear projection from $\Re^d$ to $\Re^k$ is any linear map A, or $k \times d$ matrix of rank k:

$$Z = AX, X \in \Re^d, Z \in \Re^k \tag{1}$$

Define an orthogonal projection if the row vectors of A are orthogonal to each other and have length 1. If *X* is a *d*-dimensional random variable with distribution *F,* then *Z=AX* is a *k*- dimensional random variable with distribution F_A. If *k*=1, *A* reduces to a row vector a^T, and we use lower case letters F_a.

Note that any *d*-dimensional distribution is uniquely characterized by its one-dimensional projections F_a. This follows trivially from the fact that *F* is uniquely determined by its characteristic function Ψ and that the characteristic function Ψ_a of the one-dimensional projection F_a in direction *a* equals the section of Ψ along the same direction:

$$\Psi_a(t) = E\left(e^{ita^T X}\right) = \Psi_a(ta) \tag{2}$$

By definition, PP searchers for a projection A maximizing (or minimizing) a certain objective function or projection index $Q\ (F_A)$. While Q is a functional on the space of distributions on $\Re^k$, We find it more convenient also here to use random variable terminology.

2.2 Generalities of the Information Entropy Based on ACO

ACO is a family of meta-heuristics stochastic explorative algorithms inspired by the natural optimization mechanism conducted by real ants [10]. ACO algorithms can search the best solution by using the evolutionary procedure. As shown in [11], ACO is based on the following ideas. (1) From a starting point to an ending point, each path is associated with a candidate solution to a given problem. (2) The amount of pheromone deposited on each edge of the path followed by one ant is proportional to the quality of the corresponding candidate solution. (3) The edge with a larger amount of pheromone is chosen with higher probability. As a result, the ants eventually converge to a short path, hopefully the optimum or a near-optimum solution to the target problem.

The general framework of the ACO systems is:

```
Initialization
Repeat /*each iteration at this level is called a cycle*/
Each ant is positioned on an arbitrary starting node
Repeat/*each iteration at this level is called a step*/
Each ant moves to next node according to the state transition rule
Apply the local pheromone updating rule
Until all ants have completed their tours
Apply the global pheromone updating rule
Until end condition
```

Since the establishment of the first ACO system, called Ant System, several versions including the information entropy based ACO we mentioned have been proposed to solve the premature convergence problem of the basic Ant Colony Optimization algorithm. The main idea is to evaluate stability of the current space of represented solutions using information entropy, which is then applied to turning of the algorithm's parameters. The path selection and evolutional strategy are controlled by the information entropy self-adaptively. Simulation study and performance comparison with other Ant Colony Optimization algorithms and other meta-heuristics on Traveling Salesman Problem show that the improved algorithm, with high efficiency and robustness, appears self -adaptive and can converge at the global optimum with a high probability [11].

3 Improved PP Based on ACO

Optimizing the projected indexes function is one serious drawback of the PP methods. The ACO is an effective method for the global optimization. So, we can use the ACO to deal with the indexes function. Here, the ACO proposed in [11] was employed.

3.1 Procedure of Modification

The improved PP can be shown as follows.

(1) Normalization

In order to finish the elimination of all dimensions and uniform changes in the scope of the value of the indicators, the normalization is needed:

For the greater and better indicators:

$$x(i,j)=\frac{x^*(i,j)-x_{\min}(j)}{x_{\max}(j)-x_{\min}(j)} \tag{3}$$

For the smaller and better indicators:

$$x(i,j)=\frac{x_{\max}(j)-x^*(i,j)}{x_{\max}(j)-x_{\min}(j)} \tag{4}$$

Where $x_{\max}(j)$ and $x_{\min}(j)$ are the Jth maximum and minimum indicators values respectively. $x(i,j)$ is indicators normalized sequences.

(2) Construction of projection function

The key to the PP is to find the optimal projection which can fully show the features of the data. Suppose $a=a\{a(1),a(2),\cdots,a(p)\}$ is the $p-$dimensional unit vector, PP is to project $x(i,j)$ to $\overline{a}$ and obtain the value of one-dimensional projection $z(i)$.

$$z(i)=\sum_{j=1}^{p}a(j)x(i,j),\quad i=1,2,\cdots,n \tag{5}$$

Where $a(j)\ \ j=1,2,\cdots,p$ is the projection vector, and it is unit vector.

Projection indicators in the integrated value $z(i)$ require that the projector is characterized by the spread of local projection point intensive, and the projection target function can be expressed as:

$$Q(a)=S_Z.D_Z \tag{6}$$

Where S_z is the standard deviation of the projection value, and D_z is the local density of $z(i)$. They can be expressed as follows.

$$S_z=\left\{\sum_{i=1}^{n}\left[z(i)-\bar{z}\right]^2 \Big/ n-1\right\}^{1/2} \tag{7}$$

$$D_z=\sum_{i=1}^{n}\sum_{j=1}^{p}\left(R-r_{ij}\right)\cdot I\left(R-r_{ij}\right) \tag{8}$$

Where z is the mean value of $\{z(i),i=1\sim n\}$, R is the radius of windows and $r_{\max}+\frac{p}{2}\le R\le 2p$. r_{ij} is the distance of the samples and $r_{ij}=\left|z(i)-z(j)\right|$. $I\left(R-r_{ij}\right)$ is the unit step function, when $R\ge r_{ij}$,it equals to 1, otherwise, it is zero.

(3) Optimization of projection function

When the index value of the program is given, the projection function $Q(a)$ only changes with the projection direction a . Different projection direction reflects different construction of data characteristic. The best projection direction which is most greatly possible to expose the characteristic structure is the high-dimensional data. We can estimate the best projection direction through the solution projection target function maximization question.

$$Max: Q(a)=S_2O_2$$
$$s.t.\sum_{j=1}^{n}A^2(j)=1 \tag{9}$$

It is a complex misalignment optimization question. Here, we will employ the information entropy ACO in the step acceleration method style [11]. At the early stage of the evolution, the value of $\alpha'_{(t)}$ is small and with the process proceeding, the value becomes bigger in order to explore the solution space in the beginning and reinforce the local search ability at final stage to avoid stagnation. At the same time, $\beta'_{(t)}$ is biggest at early stage in order to make the algorithm find the optimal route and later it becomes smaller to reinforce the function of random operation, which can also avoid the stagnation. To eliminate the effect of sampling noise, a pheromone update rule with stronger robustness is introduced:

$$Max: Q(a) = S_2 O_2 \tag{10}$$

$$Max: Q(a) = S_2 O_2 \qquad s.t. \sum_{j=1}^{n} A^2(j) = 1 \tag{11}$$

Where $\tau^*(t) = (nL_{nn})^{-1}$ is the pheromone vector at the itineration of t. L_{nn} is the best route found so far. ρ is the evaporation factor.

For the terminal principle, the general algorithms usually use the maximum iteration time. Here we employ the information entropy as the termination constraint because, in some cases, it is very difficult to decide the maximum iterative time for the complicated problems. We define when the information entropy is smaller than a given value (such as 0.01) the algorithm terminates and the solution is obtained.

The steps of the model as shown in Figure 1:

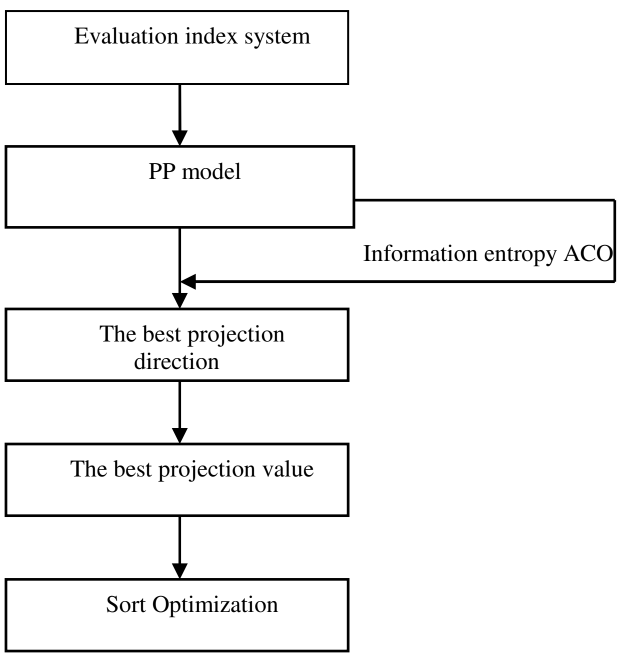


Fig. 1. Programming diagram of PPC model based on RAGA

4 Applications

To validate the efficiency of the modification proposed in the paper, we compared its performance with other PP algorithms. All algorithms were benchmarked on the same problem in [17].The data used are shown in Table 1.

Table 1. Comprehensive assessment of marine ecological indexes in a bay

Item		No.1 spot	No.2 spot	No.3 spot	No.4 spot	No.5 spot	No.6 spot
Statistic result of quality status of the sea Water	suspended matter (mg/L)	6	7	7.7	6.4	6.3	6.1
	Petroleum(mg/L)	0.0205	0.036	0.0205	0.018	0.0095	0.0105
	inorganic nitrogen(mg/L)	0.4795	0.465	0.4065	0.04685	0.195	0.06625
	activated phosphate(mg/L)	0.0119	0.2515	0.02515	0.01	0.0077	0.0021
	dissolved oxygen(mg/L)	10.2	9.85	9.76	9.81	9.78	9.92
	salinity of sea water(‰)	31.5	31.3	31.3	31.5	31.4	31.6
Quality of marine sediment	Cu (ug/L)	32.1	45.7	36.2	27.8	22.6	32.3
	Pb (ug/L)	39.5	42.1	37.1	39.8	42.5	47.8
	Zn (ug/L)	103.5	87.5	92.3	83	121.5	78.5
	Cd (ug/L)	0.61	0.87	0.96	1.32	0.81	0.77
	organic carbon (ug/L)	2400	3300	2100	2600	1700	1600
	sulphide (ug/L)	874	550	969	1349	448	535
Biological investigation in monitoring spots	Shannon-Weaver indexes of phytoplankton in monitoring spots	2.46	2.24	1.99	2.42	2.48	1.47
	Shannon-Weaver indexes of zooplankton in monitoring spots	3.16	2.68	2.42	2.33	2.72	1.98
	Shannon-Weaver indexes of zoobenthos in monitoring spots	0.996	1.192	2.570	3.348	3.1699	3.736
	Gross quantity of bacteria in monitoring spots(/L)	23000	17000	14000	20000	16000	13000
	Analysis result of organism samples Cd	0.234	0.205	0.147	0.356	0.127	0.109

The data shown above are 17- dimension data. Set up the PP model and use the Eq. (3) to Eq. (9). Then, employ the information entropy ACO to optimize the projection function. And in the ACO, $\alpha = 1.5, \beta = 4.0, \rho = 0.6, Q = 50$, when the information entropy is smaller than a given 0.01, the algorithm terminates. Then, we can obtain a^* = (0.10076, 0.14587, 0.1721, 0.1692, 0.10786, 0.10755, 0.1094, 0.1215, 0.1094, 0.1188, 0.1133, 0.1058, 0.1523, 0.2090, 0.1203, 0.1124, 0.1672), and $z^*(j) = (2.1563, 1.9651, 2.8927, 2.8651, 3.1622, 1.2329)$. We can draw a conclusion that No.5 spot is best, No.6 is the worst. The conclusion is same as the fact and more rational than [17].

5 Conclusion

The development of science and technology has signified the need for the study of higher- dimension data solution methods. We introduce the information entropy ACO to the PP which is an effective method for this problem. The ACO was used to optimize the projection function which is the bottleneck of PP. Engineering practice shows that this method can deal with the higher-dimension data effectively and rationally. This study provides a promising method for the clustering problems and comprehensive assessment problems.

However, as a young-stage evolutionary algorithm, ACO has many drawbacks to be studied further, for example, the strategy for continuous space optimization problems, and the convergence algorithm. We will study them in the future.

Acknowledgment

The project was supposed by the Fund of Teaching Research Project of Hebei University of Engineering, Doctoral Degree of Hebei University of Engineering and Fund of Youth's Academic Backbone of Hebei University of Engineering.

References

1. Lovchinov, V., Tsakovski, S.: Multivariate Statistical Approaches as Applied to Environmental Physics Studies. Central European Journal of Physics 4(2), 277–298 (2006)
2. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
3. Sadri, B., Har-Peled, S.: How fast is the K-means Method? Algorithmic 3, 185–202 (2005)
4. Kanungo, T., Mount, D.M., Netanyahu, N., et al.: A Local Search Approximation Algorithm for K-means Clustering. Computational Geometry: Theory and Applications 28(2), 89–112 (2004)
5. Mishra, N., Ron, D., Swaminathan, R.: A New Conceptual Clustering Framework. Machine Learning 56(1), 115–151 (2004)
6. Pinto, M.: A Generalization to Poincarés Theorem for Recurrent Systems. Appl. Anal. 85(4), 373–381 (2006)

7. Yang, X., Zhao, W., Chen, Y., et al.: Image Segmentation with a Fuzzy Clustering Algorithm based on Ant-tree. Signal Processing 88(10), 2453–2462 (2008)
8. Huber, P.J.: Projection Pursuit. The Annals of Statistics 13(2), 435–475 (1985)
9. Jimenez, L., Landgrebe, D.: Hyperspectral Data Analysis and Feature Reduction via Projection Pursuit. IEEE Trans. Geosci. Remote Sensing 37(6), 2653–2667 (1999)
10. Dorigo, M., Stutzle, T.: Ant Colony Optimization. MIT Press, Cambridge (2004)
11. Li, Y., Li, W.: Adaptive Ant Colony Optimization Algorithm based on Information Entropy: Foundation and Application. Fundamenta Informaticae 77(3), 229–242 (2007)
12. Friedman, J.H., Tukey, J.W.: A Projection Pursuit Algorithm for Exploratory Data Analysis. IEEE Trans. Comput. 23(9), 881–889 (1974)
13. Cheng, P.: Projection Pursuit of the Limit Distribution of the Distribution of Experience. Applied Probability and Statistics 3(1), I8–21 (1987)
14. Friedman, H.J., Stuetzle, W.: Projection Pursuit Regression. Jamer. Statist. Assoe. 76, 817–823 (1981)
15. Huber, P.J.: Projection Pursuit with Discussion. Ann. Statist. 13(1), 43–52 (1985)
16. Fu, Q., Xie, Y., Wei, Z.: Application of Projection Pursuit Evaluation Model Variations in the Sanjiang Plain. China Pedossphere 3, 249–256 (2003)

Ant Colony Algorithm for Multiple-Depot Vehicle Routing Problem with Shortest Finish Time

Jianhua Ma and Jie Yuan

Information Management School, Shandong Economic University,
Jinan, 250014, China
majh@sdie.edu.cn

Abstract. The objective of the multiple-depot vehicle routing problem (MDVRP) is to shorten the finish time, in the emergency management and special delivery research. In this paper, an ant colony algorithm for multiple-depot vehicle routing problem with shortest finish time (FTMDVRP) is studied. We discuss the concept and framework of FTMDVRP. The methods of making use of improved split algorithm to divide cars for given customer sequence is presented. We use the max flow algorithm allocate cars to each depot. Our experimental results confirm that our approach is effective in multiple-depot vehicle routing.

Keywords: Multiple-depot Vehicle Routing Problem with Shortest Finish Time, Improved Split Algorithm, Ant Colony Algorithm.

1 Introduction

During the late 1950s, the term of VRP (Vehicle Routing Problem) became popular and it was proposed first by Dantzig and Ramse [1]. Then it was summarized and deepened by N.Christofides [2]. Vehicle Routing Problem means that distribution center provides goods and arranges for vehicle routing to satisfy a certain amount of customers who have requirement of different quantity for goods, and to achieve a certain goal (e.g. the shortest distance, the least cost and shortest expenditure time) under some constraint conditions.

The objective of the traditional multiple-depot VRP is to minimize the total distance or total cost, which is mainly considered from the vehicle transportation costs, and the customers' waiting time is not concerned about. But in sometimes, the customers' waiting time is very important. The longer waiting time, the worse quality of service is. So sometime, meeting all the needs of customers in the shortest possible time is also very important. Such as the fast-food takeaway, EMS, and emergency supply distribution, and so on. In this situation, shortening the finish time of services is more important than reducing service costs or the total trips distance.

VRP with time windows consider the customers' time requirements. But when the time requirements of all customers are the sooner the better, this model is difficult to handle. In [3], VRP balancing the vehicles time utilization is solved. The objective of this problem is to minimize the using time difference of vehicles. But when the using time difference of vehicles is shortest, the using time of every vehicle may be

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 114–123, 2010.

long. In [4], VRP with shortest finish time is studied. The objective of this problem is to make the finish time of the last vehicle shortest. It is based on having one distribution center. But in fact, many large commercial enterprises have more than one distribution centers. So multiple-depot vehicle routing problem with shortest finish time (FTMDVRP) is studied.

The difficulty of MDVRP (multiple-depot vehicle routing problem) is the appointment of depots. The traditional method is dividing customers into several groups, each group is served by a depot. Thus converting multiple-depots problem into several single-depot problems which will be solved concurrently [5]. In [6], a new method of solving multiple-depot VRP is introduced by us. First arrange vehicle, then appoint depots. This method is better than the traditional method.

VRP is NP-hard problem [7], therefore people use heuristic algorithm or evolutionary algorithm to solve VRP. Tabu search algorithm, genetic algorithm, simulate annealing algorithm and ant colony algorithm have been used to solve this problem [8]. After dealing with, MDVRP can be converted into the problem of finding the optimal customer sequence. Ant colony optimization algorithm has obvious advantages in finding the optimal path problem [9].

In this paper ant colony algorithm for FTMDVRP is introduced. By arranging vehicles and appointing depots for given customer sequence, FTMDVRP is converted into the problem of finding optimal customer sequence, then ant colony algorithm is adopted to determine an optimal customer sequence. Improved split algorithm is used to arranging vehicles and max flow algorithm is used to appoint depots. First FTMDVRP is described in section 2. The methods of arranging vehicles and appointing depots for given customer sequence are introduced in section3. Ant colony algorithm for FTMDVRP is given in section 4. At last computational instances are given.

2 Multiple-Depot Vehicle Routing Problem with Shortest Finish Time

Multiple-depot VRP with shortest finish time can be described below:

There are M depots. Each depot owns $K_m (m = 1, \cdots, M)$ vehicles. Each of them can hold w goods. These depots are responsible to distribute goods to n customers. The quantity of the goods that customer i requires is $q_i (i = 1, \cdots, n)$, and $q_i < w$. Matrix T is time matrix. t_{ij} represents the time from i to j. Matrix T is real symmetrical matrix and $t_{ik} \le t_{ij} + t_{jk}$. By arranging customers should be served by which vehicle and depot, and determining the customer sequence, the problem should satisfy the following conditions:

(1) Each customer should only be served by one vehicle.

(2) Each vehicle can service several customers, but the total quantity of the goods that it provides can't exceed its carrying capacity.

(3) All vehicles set out from distribution center.

(4) Make the finish time shortest.

Only when each and every customer has been served, the whole mission can be end. So FTMDWRP is based on the service time of the customer who is served at last. Obviously, the customer who is served at last must be the last customer in one vehicle routing. Therefore, the finish time of the whole mission is the last vehicle having served the last customer in its routing. It is represented by T . The time of each vehicle completing its mission is represented by T_j, $j = 1,2,...,m$, then

$$T = \max_{j=1,2,..m} T_j \tag{1}$$

The more vehicles, the shorter finish time is. But the number of vehicle is limited, so FTMDVRP is under the condition of quantitative limitation of vehicles. There are different from total travel time and finish time. For VRP with shortest total time, there may have one vehicle routing which need long time but others need short time. In this way, the finish time will be long. And, the time from the last customer to depot is not need to consider in finish time. But this time must be considered in total time.

3 The Method of Vehicles Arrangement and Depots Appointment

Giving a customer sequence, in order to calculate the shortest finish time, customers are distributed to each vehicle in each depot under this sequence. Because all vehicles are the same customers are distributed to vehicles, then vehicles are allocated to depots. First of all, a virtual depot is brought in to substitute actual depots and each customer is considered as the customer of the virtual depot (actual depots are considered negligible). The time from each customer to virtual depot is the time from each customer to the actual depot which is nearest to it. After bringing in a virtual depot and ignoring actual depots, MDVRP is converted into SDVRP. The second step, the given customer sequence is divided to several groups by using the improved split algorithm, and then those groups of customer are allocated to each vehicle, thus the best solution of vehicles arrangement on the premise of giving the customer sequence can be found. The third step, by using the max flow algorithm, vehicles and their route are allocated to each depot. Thus, for such a given customer sequence, each vehicle's route and depot can be determined.

3.1 The Method of Vehicles Arrangement-Improved Split Algorithm

A method of vehicles arrangement was proposed in [7]. It is improved in this paper, then it can be used to arrange vehicle in FTMDVRP. It is called improved split algorithm. The role of improved split algorithm is dividing the given customer sequence to several groups and then allocating those groups of customer to each vehicle, thus the best solution of vehicles arrangement can be found on the premise of giving the customer sequence. Following a simple example is to illustrate the improved split algorithm. Take 8 customers for example. Assume that the given customer sequence is 6-8-7-4-3-5-2-1, the demand of them, the distance between them and the distance from them to depot is in Fig. 1.

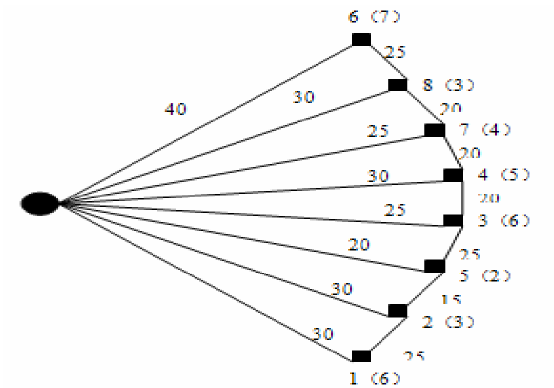


Fig. 1. The example of vehicle routing problem

Let capacity of vehicles $w=15$ and number of vehicles $m=3$. The problem is converted into the following network graph: points 0 represents the depot, point $i=1,2,\cdots,n$ represents the customer $P(i)$. Then the graph has m columns of same customer sequence. Start from point 0, if the total demands of the customer $p(1),p(2),...,p(j)$ are less than or equal to w, then add an arc between point 0 and point j, its length is total distance of routing $0 \to p(1) \to p(2) \to ... \to p(j)$. Then start from point i in the first column which is pointed by arcs. If total demands of the customer $P(i+1),P(i+2),\cdots,P(j)$ are less than or equal to w, then add an arc between point i in the first column and point j in the second column, its length is total distance of the routing $0 \to P(i+1) \to P(i+2) \to \cdots \to P(j)$, following by this rule, until draw all arcs which meet the conditions. If the last point of previous column is pointed to by arcs, then add an arc between the last point in previous column and the last point in current column. Its length is 0. Thus the network graph corresponding to this example is in Fig. 2.

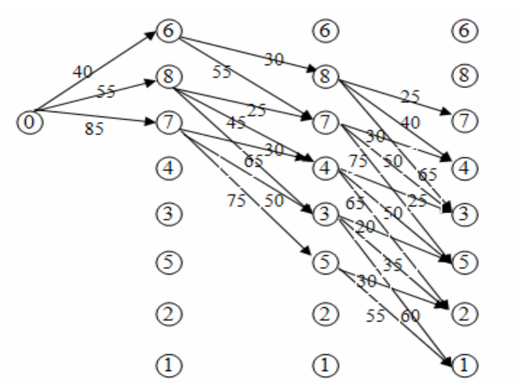


Fig. 2. Result of improved split algorithm

It is different from VRP with shortest total time. The directed graph is not used to calculate the route from point 0 to last point which total arc length are longest. But it is used to minimize the arc which length is longest. First isolated point is deleted. The second step, the longest arc is deleted, and then a new graph is got. The second step is repeated until the route from point 0 to point 1 does not exist in the new graph. Now the route from point 0 to point 1 existing in the graph in the previous graph is the optimal route. The length of the previous arc deleted is the finish time. As the graph shows, the arcs which length is 85 or 75 are deleted. A new graph can be got. And the route from point 0 to point 1 exists in the new graph. So the longest arc in the new graph is deleted. And the route from point 0 to point 1 does not exist in this new graph. So the route from point 0 to point 1 exists in the previous graph is the optimal route. The shortest finish time is 65, the optimal route is 0-6-8, 0-7-4-3, 0-5-2-1.

3.2 The Method of Depots Appointment

When arrange vehicles, the time from first customer to virtual depot is to the actual depot which is nearest to it. If the first customer is appointed to depot which is nearest to it, the number of arranged vehicles of some depots may exceed their limit. Therefore the vehicle need to be appointed to depot over again.

For MDVRP, a method of depot appointment is introduced in [7] by using the min flow algorithm. For FTMDVRP, it can't be solved by min flow algorithm. Max flow algorithm is used time after time to solve this problem.

First of all, a network graph should be constructed. Point s represents the initial point. Point t represents the termination point. Point $i = 1, 2, ..., M$ represents that depot i. Point $j = 1, 2, ..., m$ represents vehicle j. Arcs are added from the initial point to each depot. The capacity of arc from point s to depot is the number of vehicles owned by this depot, and its length is 0; Arcs are added from depots to each vehicle. The capacity of arc from depots to each vehicle is 1, its length is the time of vehicle having served its customers; Arcs are added from vehicles to termination point. The capacity of arc from vehicles to termination point is 1, its length is 0.

For example, there are 2 depots a, b. Each depot has two vehicles. There are 15 customers. Their sequence is 5-4-8-1-3-9-10-7-2-6. t_{ij} represents the time from i to j. Those customers are divided into 4 groups by the improved split algorithm, and each group of customers is served by one vehicle, thus the routing of those 4 vehicles are 0-5-4-8, 0-1-3, 0-9-10, 0-7-2-6. So the network graph structured is in Fig. 3.

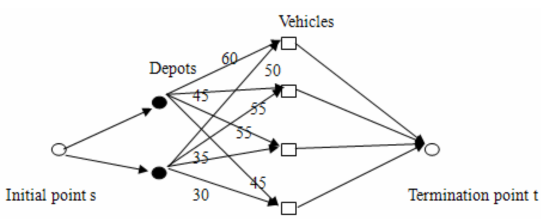


Fig. 3. Graph of the max flow

The limit of capacity is integer, so the feasible flow is integral flow. When the flow value is m, the flow value of arc from vehicles to termination point must be 1. The flow value of arc from depot to vehicle is 1 or 0, and each vehicle has exactly only one arc whose flow value is 1. If the flow value of arc from depot i to vehicle j is 1, then the vehicle i is allocated to the depot j. A solution of depots appointment is got. The finish time is the longest length of the arcs whose flow value is 1.The aim is to minimize this finish time under the condition of flow value is m.

Step 1: Calculate the feasible flow whose flow value is m. Then the longest length of the arc is got in this feasible flow, notes as l. The arc its length is more than or equal to l are deleted. Thus a new network graph is got.

Step 2: Calculate the max flow of the new network graph. If the max flow is m, go to step1. Else, the feasible flow whose flow value is m in last network graph is the best solution of depots appointment. The shortest finish time is l.

As the Fig.3 shows, the max flow is 4. The longest arc is deleted time after time. They are 60,55,55,50. Then the network graph whose max flow is 3 can be got. When the arc whose length is 50 won't be deleted, the feasible flow in the network graph is the best solution of depots appointment. Using the max flow algorithm to calculate this network graph, the best solution of depots appointment can be got: a-1-3, a-7-2-6, b-5-4-8, b-9-10. The short finish time is 50.

Because the number of the vehicles and depots is small in general vehicle routing problems, the network graph is not large in scale. Computing time is fast.

4 Ant Colony Algorithm

The shortest finish time of given customer sequence can be got by above method. So FTMDVRP is converted into finding the optimal customer sequence, what need to do is using ant colony algorithm to find the optimal customer sequence.

A full permutation of all customers is a viable solution. In each iterative process, ants build the full permutation of all customers in accordance with pheromone matrix and inspired information, and then pheromone matrix is updated. So, the key of ant colony algorithm is determining pheromone matrix. Suppose there are n customers, then pheromone matrix is a $n\times(n+1)$ matrix. The last row shows the information of vehicle setting out from depot to each customer. Row i shows the information of vehicle setting out from customer i to each customer. At the beginning, the probabilities of ants choosing the next customer are equal. Therefore, the initial pheromone matrix is the equal probability matrix. When $t=0$, the pheromone matrix is:

$$b_{ij}(0)=\begin{Bmatrix} 0 & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{n-1} & \frac{1}{n-1} & \cdots & 0 \\ \frac{1}{\mathrm{n}} & \frac{1}{n} & \cdots & \frac{1}{n} \end{Bmatrix}_{(n+1)\times n} \tag{2}$$

Where $b_{ij}, i = 1,2,...,n+1, j = 1,2,...,n$ is the pheromone of customer j being closely behind customer i. Suppose there are m ants. In periods t, m ants visit the customers, and then m feasible solutions can be got. If finish time of each solution is $L_k(t); k = 1,2,...,m$, the increased amount of pheromone in the period of t is as follows:

$$\Delta b_{ij}(t) = \sum_{k=1}^{m} \Delta b_{ij}^{k}(t) \tag{3}$$

and

$$\Delta b_{ij}^{k}(t) = \begin{cases} Q / L_k(t); & \text{if ant } k \text{ set out from } i \text{ to visit } j, Q \text{ is constant} \\ 0; & \text{else} \end{cases} \tag{4}$$

Thus the amount of information in the periods $t+1$ is as follows:

$$b_{ij}(t+1) = (1-\rho) b_{ij}(t) + \Delta b_{ij}(t) \tag{5}$$

Where, ρ is the volatile factor. Inspired information is also shown by a $n \times (n+1)$ matrix. Inspired information in the last row is $\eta_{n+1,j} = A / t_j$, others are $\eta_{ij} = A / t_{ij}$. t_{ij} is the time from two customer i to customer j. t_j is the average time from customer j to each depot. A is a constant.

To build the full permutation of all customers, firstly, the first customer is chosen in accordance with probability. Then according to the probability, the ant choose the next customer from $allowed_i$ ($allowed_i$ represents the set of customers which can be visited by ant setting out from customer i). In the period of t, the probability of ant setting out from customer i to visit customer j is as follows:

$$p_{ij} = \begin{cases} \dfrac{[b_{ij}(t)]^{\alpha} \bullet [\eta_{ij}]^{\beta}}{\sum\limits_{s \in allowed_i} [b_{is}(t)]^{\alpha} \bullet [\eta_{is}]^{\beta}}; & \text{j} \in allowed_i \\ 0 & ; \text{j} \notin allowed_i \end{cases} \tag{6}$$

Where, α is importance degree of the pheromone, and β is importance degree of inspired information. Through adjusting the value of α, β, the relative important degree of the pheromone and inspired information can be determined. After calculating the selection probability matrix, the customer is selected randomly from $allowed_i$ by the method of roulette, and then deleted from $allowed_i$. Followed by analogy, a full permutation of all customers can be got at random.

Through combining ant colony algorithm with the max flow algorithm and improved split algorithm, the basic steps of solving FTMDVRP are as follows:

Step 1: Given that the number of ant m, iterative periods T, determine the values of α, β and ρ. Input the initial pheromone matrix $b_{ij}(0)$ and inspired information matrix η_{ij}, $t = 0$. According to initial pheromone matrix and inspired information matrix, initial ant colony can be got. Permutations of all customers are built at random.

$$A(0) = \{A_1(0), A_2(0),..., A_m(0)\} \tag{7}$$

Step 2: Each customer sequence is calculated by improved split algorithm and the max flow algorithm, then the finish time of each customer sequence can be got. Write the shortest finish time $finishtime$ and the best routing $besttrip$ having appeared currently.

Step 3: Calculate new pheromone matrix by formula (3-5).

Step 4: m sequences are got at random by updated pheromone matrix and inspired information matrix.

$$A(t+1) = \{A_1(t+1), A_2(t+1),..., A_m(t+1)\} \tag{8}$$

$t = t+1$, if $t > T$, then algorithm is stopped, and the values of $finishtime$ and $besttrip$ are output, else, go to step 2.

The program of this algorithm is written by using SCILAB, parameter is set as follows: m =10, $\alpha = 1.2, \beta = 0.5, P = 0.2, \varepsilon = 0.4, \phi = 0.2$, T=200.

5 Computational Results

In order to verify our algorithm's feasibility, the date in [10] is calculated. There have been some goods needing to deliver to stricken areas. The coordinate data of depots and customers show in Table 1 and Table 2. The carrying capacity of each vehicle is 25. The speed of each vehicle is 1 kilometer per minute.

Table 1. The location and demand of customers

Customer	1	2	3	4	5
Location	(74,29)	(64,26)	(67,80)	(88,15)	(21,65)
Demand	3	9	5	6	14
Customer	6	7	8	9	10
Location	(72,42)	(92,80)	(46,38)	(76,86)	(30,46)
Demand	9	7	10	4	11
Customer	11	12	13	14	15
Location	(63,48)	(23,11)	(36,72)	(29,54)	(66,16)
Demand	10	3	8	6	5

Table 2. The lacation and vehicle quantity of depots

Depot	A	B	C
Location	(40,65)	(55,25)	(90,55)
Vehicle Quantity	2	3	2

Calculating 10 times, the results are respectively: 52.9, 52.9, 53.8, 52.9, 50.8, 51.2, 55.1, 52.9, 51.2, and 51.2. So the shortest finish time is 50.8 minutes. The best solution is: the routing is A-13-14-8, its time is 50.7; A-5-10, and its time is 40. B-1-6-11, its time is 43.3; B-15-2-4, its time is 50.8; B-12, its time is 34.9; C-3-9, its time is 44.8; C-7, its time is 25.1. The average time between two customers is 43.7 minutes. The average number of customers served by one vehicle is 2.14, so average time of each vehicle is 92.2 minutes. The finish time calculated by our algorithm is less than the average service time of each vehicle. This shows our result can be accepted.

In paper [7], through comparing, the method of first arranging vehicle, then appointing depot is better than others method in solving MDVRP. In order to compare the effect of the methods in solving FTMDVRP, the traditional method (i.e. dividing customers into several groups, each group is served by a depot, thus multiple-depots problem is converted into several single-depot problems which will be solved concurrently, then these single-depot problems are solved by ant colony algorithm. And which shortest finish time is longest in these single-depot problems is the shortest finish time of the whole problem) is used to solve FTMDVRP. This method is called algorithm 1, and the method introduced in this paper is called algorithm2.

Above example is calculated by algorithm 1. The result is 66.8 minutes. The effect is less than the effect of algorithm 2. In addition, 20 examples are built at random. The number of the vehicles is 3. The number of the customers is 15. The demand of customers is from 3 to 15. The coordinates of customers and depots are distributing in the scope of 100×100 .The number of vehicles in each depot is from 1 to 3. The carrying capacity of vehicles is from 25 to 30. Total carrying capacity exceeds the 120 percents of total requirement. These 20 examples are calculated by these two algorithms. The result is in Fig.4.

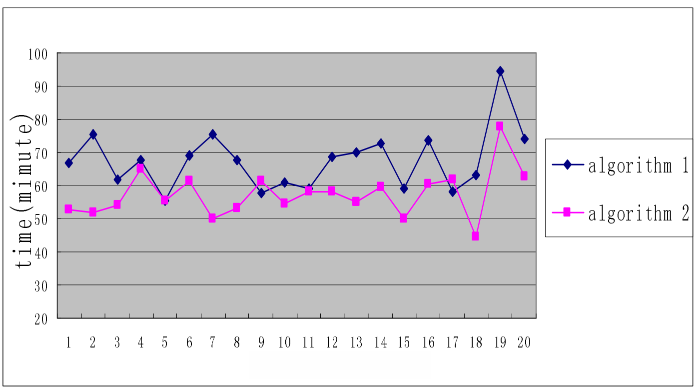


Fig. 4. The compare of two algorithms

The results of 17 examples calculated by algorithm 2 are better than algorithm 1 (figure 4). Therefore, the method of solving FTMDVRP introduced in this paper is better than the traditional method.

6 Conclusion

Ant colony algorithm for multiple-depot vehicle routing problem with shortest finish time is discussed in this paper. At first, FTMDVRP is introduced. Through arranging vehicles and appointing depots for given customer sequence, FTMDVRP is converted into the problem of finding the optimal customer sequence, then ant colony algorithm is adopted to determine an optimal customer sequence.

As FTMDVRP is a new type of VRP, it is short of classical example. The different algorithms should be studied farther in the future. Through contrast, the better algorithm can be found. The carrying capacities of the vehicle are same in this paper. In the future, the different carrying capacities vehicles can be considered.

References

1. Dantzig, G., Ramser, J.: The Trunk Dispatching Problem. Management 6(6), 80–91 (1959)
2. Christofides, N., Mingozzi, A., Toth, P.: The Vehicle Routing Problem. In: Combinational Optimization. Johnly Wiley, New York (1979)
3. Nikolakopoulou, G., Kortesis, S., Synefaki, A., Kalfakakou, R.: Solving a Vehicle Routing Problem by Balancing the Vehicles Time Utilization. European Journal of Operational Research 152, 520–527 (2004)
4. Zhang, X., Ma, J., Liu, W., Jin, F.: Ant Colony Algorithm for Vehicle Routing Problem with Shortest Completion Time. In: The Proceedings of the 13th International Conference on Industrial Engineering and Engineering Management, pp. 2928–2933 (2006)
5. Zhong, S., He, G.: On Intelligent Optimization of Multi-depots Vehicle Scheduling Problem. Journal of East China Jiaotong University 6, 25–29 (2004)
6. Ma, J., Yuan, J., Zhang, Q.: Ant Colony Algorithms for Multiple-Depot Vehicle Routing. In: The Proceedings of 2009 Conference on Systems Science, Management Science & System Dynamics, pp. 55–60 (2009)
7. Lenstra, J.K., Kan, A.: Complexity of Vehicle Routing and Scheduling Problems. Networks 11, 221–227 (1981)
8. Liu, Y., Xuan, H.: Summarizing Research on Models and Algorithms for Vehicle Routing Problem. Journal of Engineering Management 19(1), 124–130 (2005)
9. Ma, J.: A Mutation Ant Colony Optimization Algorithm of Single Batch Machine Scheduling Problem. Computer Engineering and Applications 42(3), 53–56 (2006)
10. Zou, T., Li Ning, S.D., Li, J.: Genetic Algorithm for Multiple-Depot Vehicle Routing Problem. Computer Engineering and Applications 21, 82–83 (2004)

Chaos Caused by a Transitive Map in Information

Lidong Wang[1], Jiao Peng[2], and Heng Liu[3]

[1] School of Science, Dalian Nationalities University,18# Liaohe West Road, Dalian, 116600, Liaoning, China
[2] Mathematical departments, Liaoning Normal University, 850# Huanghe Road, Dalian, 116029, Liaoning, China
[3] School of Science, Dalian Nationalities University,18# Liaohe West Road, Dalian, 116600, Liaoning, China
wld0117@yahoo.cn, pengjiao320320@126.com, hliu@dlnu.edu.cn

Abstract. In this paper, the chaotic phenomena have been discussed in an information system design process. Let (X,d_1) (Y,d_2) be compact metric spaces without isolated point and $h:X\to Y$ be a semi-conjugacy between f and g , where $f:X\to X$ $g:Y\to Y$ are continuous maps. In this paper, we investigated the expanding maps with transitivity. And we gave a sufficient condition about expanding maps are chaotic in the sense of Wiggins and Kato.

Keywords: Compact Space, Transitivity, Expand, Semi-conjugacy.

1 Introduction

"Butterfly effect" phenomenon revealed by the "sensitivity to initial conditions" that is the subtle differences in the input may soon become a tremendous difference in the output. In the information age, "Butterfly Effect" has become an increasingly common reality. A trivial gossip can rapidly spread through the Internet world, thus brings about a business disaster. To be invincible in the information age, we must pay attention to chaos. In the perspective of chaos theory, the value of information resources management will be positioned to control the chaotic state and the chaotic management of cognitive state. With chaos theory we can improve information resources management practices by the eight dimensions of the collection of information resources, organization and refining of information resources, analysis and forecasting of information resources and transmission and retrieval of information resources. So, in order to do the above work, we must understand the relationship of the different definitions of chaos.

In 1975, Li-York [1] gave the first mathematical definition of chaos, since then, people's understanding of the concept of chaos has become deeper and deeper. Although there is still no all recognized definition of chaos or classification for a chaotic system, people of various fields gave different definitions of chaos such as distribution chaos, Delaney's chaos, chaos in the sense of Wiggins and chaos in the

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 124–129, 2010.

sense of Kato and so on. Many authors found that a transitive system with a dense set of periodic points should be of sensitive dependence on initial conditions. On the one hand Huang and Ye [2] studied the transitive system with a fixed point and showed that such a system is chaotic in the sense of Li and York. On the other hand, Xiong and Chen [3] described the chaotic phenomena caused by topologically mixing map using the words different from Li and Yorker's.

The main purpose of this paper is to discuss chaoticity of extension of (Y, g).

2 Problem Statement and Preliminaries

Definition 2.1. A map f is called to be accessible if for every $\varepsilon > 0$ and every pair of non-empty open sets U and V of X, there exists $x \in U$, $y \in V$ and a positive integer n such that $d(f^n(x), f^n(y)) < \varepsilon$.

Definition 2.2. A map f is called to be sensitive if there is an $\varepsilon > 0$, such that whenever U is a non-empty open set of X ,there exist points $x, y \in U$ such that $d(f^n(x), f^n(y)) < \varepsilon$ for some positive integer n, where ε is called a sensitive constant of f.

Definition 2.3. A continuous map f is called to be chaotic in the sense of Wiggins, if 1) f is topological transitive.

2) f is sensitive dependence on initial conditions.

Definition 2.4. A map f is called to be chaotic in the sense of Kato, if f is both sensitive and accessible.

Definition 2.5. Let (X, d_1) (Y, d_2) be two metric spaces, and $f: X \to X$ $g: Y \to Y$ be two continuous maps. f and g are said to be topologically semi-conjugate if there exists a map $h: X \to Y$,such that $h \circ f = g \circ h$, h is a continuous surjection.

Definition 2.6. Let $h: X \to Y$ is a topological semi-conjugacy from f to g ,the subset $X_h \subset X$ is called $h-$ minimum cover of Y ,if

1) $\overline{X_h} = X_h$
2) $f(X_h) \subset X_h$
3) $h(X_h) = Y$
4) X_h has no proper subsets such that 1) 2) 3)

Lemma 2.1. Let $h: X \to Y$ is a topological semi-conjugacy, then there exists a $h-$ minimum cover of Y, and if g is topologically transitive, then there exists a $h-$ minimum cover of $f\big|_{X_h}$.

Proof : For a proof, see Ref [4].

Lemma 2.2. Let (X, f) be a compact system, let f be a surjective continuous map, then the necessary and sufficient condition for the transitivity of f is that there exists $x \in X$ such that $\overline{orb(x)} = X$.

Proof : For a proof, see Ref. [4].

Lemma 2.3. Suppose (X, f) is a transitive topological system, where X is a complete metric space dense in itself ,with a metric d, then there is a subset K of space X, of which the intersection with any non-empty open set in X contains a non-empty perfect set. For the set K the following conditions are satisfied:

1) Every point of K is transitive;

2) For every integer $N \geq 2$ and for any pairwise different N points $x_1, x_2, \cdots, x_N$ in K,

$$\limsup_{n\to\infty} \min\{d(f^n(x_i), f^n(x_j)) : i, j \in \{1, 2, \cdots, N\}; i \neq j\} > 0 \tag{1}$$

3) For every integer $N \geq 2$ and for any pairwise different N points $x_1, x_2, \cdots, x_N$ in K,

$$\liminf_{n\to\infty} \max\{d(f^n(x_i), f^n(x_j)) : i, j \in \{1, 2, \cdots, N\}; i \neq j\} \leq d_0 \tag{2}$$

Where d_0 is the infimum of the diameters of the invariant sets of the system (X, f), (Note that if there is a fixed point of the system (X, f), then $d_0 = 0$.)

4) For every integer $N \geq 2$ and for any pairwise different N points $x_1, x_2, \cdots, x_N$ in K,

$$\min_{1\leq i\leq N} \limsup_{n\to\infty} \min\{d(f^n(x_i), f^n(x_j)) : i, j \in \{1, 2, \cdots, N\}; i \neq j\} \geq \frac{\lambda_N}{2} \tag{3}$$

Where λ_N is the $N-$critically sensitive coefficent of the system (X, f) and for any pairwise different N points $x_1, x_2, \cdots, x_N$ in K,

$$\min_{1\leq i\leq N} \limsup_{n\to\infty} \min\{d(f^n(x_i), f^n(x_j)) : i, j \in \{1, 2, \cdots, N\}; i \neq j\} \geq \frac{\lambda_2}{2(N-1)} \tag{4}$$

Lemma 2.4. Let (X, f) be a compact topological dynamical system. If there exists pair-wise disjoint closed subsets $A_1, A_2 \cdots A_k$ satisfying

$$f(A_i) \supseteq \bigcup_{j=1}^{k} A_j, i = 1, 2 \cdots k, \tag{5}$$

then

$$\Lambda = \left\{ \bigcap_{s=0}^{\infty} f^{-s}(A_{i_s}) : (i_0, i_0, \cdots) \in \sum(k) \right\} \tag{6}$$

is an invariant subset of the hyperspace map 2^f, and $(2^f \big|_{\Lambda}, \Lambda)$ is topologically semi-conjugate to the symbolic dynamical system $(\sum(k), \sigma)$.

Reference [6] can be seen as a proof.

3 Main Results

In this paper, we obtain following main results.

Theorem 3.1. Let $(X, f)(Y, g)$ be compact systems, $h: X \to Y$ is a topological semi-conjugacy, if g is topologically transitive and there exists a non-recurrent point of f in X_h which is a minimum cover of $h(\exists x_1 \notin R(f \big|_{X_h}))$, then

1) f is chaotic in the sense of Wiggins.

2) Further more, if $\exists \overline{x_0} \in F(f \big|_{X_h})$, then $f \big|_{X_h}$ is chaotic in the sense of Kato.

Corollary 3.1. Let X be a compact metric space, D is a compact subset of X, $f: D \to X$ is continuous. If there exists mutually disjoint subsets D_1 and D_2 of D, the restriction of f to each D_i is continuous $f(D_i) \supset D_1 \cup D_2$ $i = 1, 2$, then

1) If there exists $x_1 \in K$ such that $x_1 \notin R(f \big|_K)$, then $f \big|_K$ is chaotic in the sense of Wiggins.

2) Further more, if there exists $x_0 \in F(f \big|_K)$, then $f \big|_K$ is chaotic in the sense of Kato.

4 Proof of Main Theorems

Proof: First we prove that f is chaotic in the sense of Wiggins. Since $h: X \to Y$ is topologically semi-conjugate and g is topologically transitive. By Lemma 1, there exists a minimum cover X_h of h, such that $f\big|_{X_h}$ is topologically transitive. Next we prove $f\big|_{X_h}$ is sensitive dependence on initial conditions. Since $f\big|_{X_h}$ is topological transitive, by Lemma 2 $\exists x_0 \in X_h$, such that $\overline{orb(x_0)} = X_h$. Again because of $x_1 \notin R(f\big|_{X_h})$. Thus $x_1 \notin \overline{\bigcup_{k=1}^{\infty} f^k(x_1)}$, so $d_1 = \inf d(x_1, y) > 0$, $y \in \overline{\bigcup_{k=1}^{\infty} f^k(x_1)}$.

$\forall x \in X_h$ and $U(x)$ which is the neighborhood of x, $\exists m > n > 0$ such that $f^m(x_0), f^n(x_0) \in U(x)$. Again because of the continuity of f and the compactness of X_h, $\exists \varepsilon > 0$, such that

$$\forall x' \in U(x_1, \varepsilon), d(f^{m-n}(x'), f^{m-n}(x_1)) < \frac{d_1}{4} \tag{7}$$

Denote $\varepsilon' = \min\{\varepsilon, \frac{d_1}{4}\}$. Since $\overline{orb(x_0)} = X_h$, $\overline{orb(f^n(x_0))} = X_h$, so $\exists \{n_k\}$, such that

$$d(f^{n_k+n}(x_0), x_1) < \varepsilon' \tag{8}$$

and

$$\begin{aligned} &d(f^{n_k+m}(x_0), f^{n_k+n}(x_0)) \geq \\ &d(x_1, f^{m-n}(x_1)) - d(f^{n_k+n}(x_0), x_1) - d(f^{n_k+m}(x_0), f^{m-n}(x_1)) > \\ &d_1 - {}^{d_1}\!/\!_4 - {}^{d_1}\!/\!_4 = {}^{d_1}\!/\!_2 . \end{aligned} \tag{9}$$

Therefore

$$d(f^{n_k+m}(x_0), f^{n_k}(x)) > {}^{d_1}\!/\!_4 , \quad d(f^{n_k+n}(x_0), f^{n_k}(x)) > {}^{d_1}\!/\!_4 . \tag{10}$$

So $f\big|_{X_h}$ is sensitive. $f\big|_{X_h}$ is chaotic in the sense of Wiggins.

Next we prove that $f\big|_{X_h}$ is chaotic in the sense of Kato. Since f is sensitive, it suffices to show f is accessible.

Let U', V' be two non-empty open sets of X_h, where $U' = U \cap X_h, V' = V \cap X_h$, U and V be non-empty open sets of X. Since f is transitive, by Lemma 3, there exists a subset K of X_h such that $K \cap A$ contains non-empty compact full set for each open subset A of X_h and K satisfies the conditions 3) of Lemma 3 with $N = 2$, we pick $x \in U' \cap K$, $y \in V' \cap K$, since $f\big|_{X_h}$ has a fixed point , we have

$$\liminf_{n\to\infty} d(f^n(x), f^n(y)) = 0. \tag{11}$$

Hence there is a positive integer n such that

$$d(f^n(x), f^n(y)) < \varepsilon. \tag{12}$$

This shows that f is accessible. So f is chaotic in the sense of Kato.

5 Proof of Corollary

Proof: The corollary satisfies the conditions of Lemma 4, so there exists a compact invariant set $K \subset D$ such that $f\big|_K$ is semi-conjugate to σ. So by the main theorem, we can prove this corollary.

Acknowledgment

This work is supported by the major basic research fund of the educational department of the State Ethnic Affairs Commission of PRC No.07DL05 and The NSFC No.10971245 and The Basic Research Foundation of The Educational Department of Liaoning Province of China, No.2009A141.

References

1. Li, T., Yorke, J.: Period 3 Implies Chaos. Math. Monthly 82, 985–992 (1975)
2. Huang, W., Ye, X.D.: Deveny's Chaos or 2-scattering Implies Li-York's chaos. Topology and Its Applications 117(3), 259–272 (2002)
3. Xiong, J.C., Chen, E.C.: Chaos Caused by a Strong-mixing Measure-preserving Transformation. Science of China (A) 40(3), 253–260 (1997)
4. Zhou, Z.: Symbolic Dynamical. Shanghai Scientific and Technological Education Publishing House, Shanghai (1997)
5. Xiong, J.C.: Chaos in a Topologically Transitive System. Science of China (A) 35(3), 302–311 (2005)
6. Wang, Y., Wei, G.: Conditions Ensuring that Hyperspace Dynamical Systems Contain Subsystems Topologically (semi-)conjugate to Symbolic Dynamical Systems. Chaos, Solitons and Fractals 36(2), 283–289 (2008)
7. Zhou, Z.Z., He, W.H.: Administrative Levels of Orbit Structure and Topological Semi-conjugate. Science of China (A) 25(5), 457–464 (1995)

Comparative Study of Methods for Monitoring and Controlling a Regional Economic System*

Shi Zheng[1], Wen Zheng[2], and Xia Jin[1]

[1] School of Economic and Management, Dalian Nationalities University, Dalian, 116600 Liaoning Province, China
[2] Northeastern University in Qinhuangdao, 066000 Hebei Province, Qinhuangdao, China
{zhengshi1993,xiajin1998}@yahoo.com.cn, zhengwen@mail.neuq.edu.cn

Abstract. As a measure of the regional economic development, regional economic competitiveness in a regional economic system has been monitored and controlled by two different methods: classical integration control method and the fuzzy integration control method. In this paper, we discussed the relationship of these two methods, and then present a new method combing them together. Our experimental results confirm that this new method is effective in monitoring and controlling a regional economic system.

Keywords: Regional Economic System; Fuzzy Integration Control Method; Classical Integration Control Methods.

1 Introduction

A region is a regional unity which is useful in terms of describing, analyzing, managing, planning or policy making. All kinds of economic activities and their links within a regional unity form a regional economy system. In contrast with an economic system, which is the system of production, distribution and consumption of goods and services of a national economy, a regional economic system is ranged between macro and micro economy. It can be regarded as the spatial extent of the division of the national economy. In a regional economic system model, the output of the system can be the competitiveness, productions, etc. Within these models, variables seldom can be observed by the experts. Ever since the application of system identification [1], Expert System [2], fuzzy Evaluation [3] [4], fuzzy control method [5], it is possible to monitor and control the regional economic system dynamically.

Single-loop controllers have been used in industry for a long time [6]. The controllers are currently going through an interesting phase of development where features like auto-tuning, gain scheduling and adaptations are added [7]. Since the

* This research is supported by Liaoning Social Science funds (Project No. L07DJY067) and Ph.D. support funds (20066201). Supported by the Fundemental Research Funds for the Central Universities.

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 130–139, 2010.

controllers are made and used in large number, much industrial experience of using fuzzy control etc. [8] is accumulated.

However, design details and control performance about a regional economic system has never been discussed up to the present. It is necessary to give the users more information about the application of the advanced control techniques in a regional economic system.

In this paper, we demonstrate that regional economic development indeed can be monitored and controlled dynamically. Two different methods have been discussed. Our experimental results confirm that this new method is effective in monitoring and controlling a regional economic system.

2 Methods Comparison

In this section, we adopt fuzzy complex evaluation method in the first step, and then adopt the classical integration control method and fuzzy integration control method respectively.

2.1 Regional Economic Evaluation Model

The regional economic evaluation model can evaluate whether a regional economic system is in good condition or not. Advanced methods are improvements of the evaluation method. These methods need evaluation method as a foundation. The index of regional competitiveness includes two aspects: the core elements and auxiliary elements of regional competitiveness. The index data of core elements is sourced from China statistical yearbook. The index data from the yearbook should be quantified. The quantified method adopted in this section is following this thinking, which transforms the interval variables into 5 levels. Since index variables is continuous one (e.g. technology market transaction amount etc.). The starting point of X axis is the minimum of its value scope so as to avoid the error when value of index data is zero. Due to the difficulties of acquiring the objective data of auxiliary elements of regional competitiveness, the auxiliary elements data is acquired from the expert's evaluation. And the data should be transformed into data in 5 levels.

2.1.1 Determine Evaluation Index Set

The regional competitiveness constituent element is set by $u = \{u_1, u_2, u_3, u_4\}$. u_i depicts the constituent element. Among them, $u_1 = \{x_{11}, x_{12}\}$, $u_2 = \{x_{21}, x_{22}\}$, $u_3 = \{x_{31}, x_{32}, x_{33}, x_{34}\}$, $u_4 = \{x_{41}, x_{42}, x_{43}\}$. The evaluation set of the regional competitiveness core element is set, i.e. $\gamma_1 = \{NM, NS, ZE, PS, PM\}$. The evaluation set of the regional competitiveness auxiliary element is set, i.e. $\gamma_2 = \{NM, NS, ZE, PS, PM\}$. Among them, NM, NS, ZE, PS, PM depict negative middle, negative Small, Zero, Positive Small, and Positive Middle in Fuzzy Control Theory. In this paper they depict the weak, relative weak, normal, relative strong, and strong of the regional competitiveness.

2.1.2 Depict the Quantities Value of Each Index

Adopting the above-mentioned 5 level method, the quantities value of core element of regional competitiveness was gotten. Adopting the expert's evaluation method, the scores of the auxiliary elements of regional competitiveness also can be gotten. The scores of auxiliary elements should be transformed into data in interval variables of [0, 1].

This paper regards these data as the comprehensive data of the regional competitiveness evaluation system, which can be evaluated later. Comparing the each element by the familiar (0, 1, 2) scale method, then construct a comparative matrix and calculate the sequence index of each element. In the second stage, transform the comparative matrix into determine matrix. At the same time the consistency of the matrix should be proven. This method is an improved one of Saaty' 9 scale method. It can avoid the inconsistency of the determine matrix. By this method, the weightiness of U layer to X layer can be calculated. Let the first level evaluation vector of u_i is B_i, then

$B_i = N_i \circ R_i = \{B_{i1}, B_{i2}, B_{i3}, B_{i4}\}$ .The operator in the above equation is normal real multiplier.

2.1.3 Carry on the Second Stage Evaluation

Regarding each $u_i (i = 1, 2, 3)$ as an element, i.e. $U_1 = \{u_1, u_2, u_3\}$, regarding u_4 as an element, i.e. $U_2 = u_4$, then U_1, U_2 are two element sets respectively. The single element evaluation matrix of U_1 is

$$U_{i1} = \begin{Bmatrix} B_{i1} \\ B_{i2} \\ B_{i3} \end{Bmatrix}$$

And we can also get

$$U_{i2} = B_{i4}$$

Applying the 3 scale's AHP method, the weightness of U_1 layer to A_{i1} layer is $W_{i1} = \{w_1, w_2, w_3\}$. Then the second evaluation vector can be gotten, i.e. $A_{i1} = W_{i1} \circ U_{i1}$. Among them, A_{i1} is i columns and one row vector. As we also know that $A_{i2} = U_{i2} = B_{i4}$.

2.2 Classical Control Integration Method

2.2.1 System Identification Method

In this method, regional competitiveness core-elements is regarded as system input No.1 and regional competitiveness auxiliary elements as system input No.2, and regional competitiveness is regarded as system output, then the system features can be described effectively.

Then we use the ARX parametric model to identify the regional economic system. ARX Model goes as follows:

$$A(q^{-1})y(k) = B(q^{-1})u(k - n_k) + e(k) \tag{1}$$

In equation (1), y is the output of the system and $y(k)$ denotes the output at k time. u is the input of the system and $u(k - n_k)$ denotes the input of the system with n_k period system latency. $e(k)$ is a random error term. $A(q^{-1})$ and $B(q^{-1})$ are the parameters of the model, i.e. lagged operator or translation operator polynomial. q is the unit lag operator. k denotes the discrete time. n_k is the system latency. The ARX model (1) is the dynamic model described by a difference equation. The economic meaning of $B(q^{-1})$ is the enhancement effect of the regional productive elements distributes in k period of time. $A(q^{-1})$ and $B(q^{-1})$ can be estimated by system identification method. Thus, the regional nonlinear system will have the definite forms.

The next step is parametric estimation. It is the process that estimates the parameters of ARX Model by error criterion. The mature methods are Least Square Method and Improved Least Square Method [10] (Generalized Least Square Method [11] and Extended Least Square Method). Other methods include Error Predict Estimation Method, IV Estimation Method [12] and Mathematical Approximation Method, such as Forward-backward Method, Yule-Walker Method [13], Burg Method [14], and Geometric Lattice Method [15]. Parametric estimation of ARX Model can be realized by MATLAB Toolbox.

The third step is to select a best model by appropriate algorithm. Through calculation, ARX117 model has been obtained in the following form:

$$\begin{aligned}&[1-0.8131(0.0685)q^{-1}]y(t) = -0.1022(0.128)q^{-7}u_{1t}\\&+0.08216(0.0524)q^{-7}u_{2t} + e_t\end{aligned} \tag{2}$$

The fourth step is to transform the discrete transfer function from ARX model. According to control system identification practice, ARX Model can be transformed into discrete transfer function. Equation (2) can be transformed as follows:

$$G(z) = \begin{bmatrix} -1.022/(z-0.8131) & 0 \\ 0 & 0.0826/(z-0.8131) \end{bmatrix} \tag{3}$$

Then we can construct control model based on transfer function, Equation (3) can be rewritten in the following form:

$$z = \{[\ ],[\ ];[\ ],[\ ]\};$$
$$p = [-0.8131,0;0,-0.8131];$$
$$k = [-0.1022,0;0,0.0826];$$

Transfer function can be transformed into zero pole gain model. Using the following command:

$$G = zpk(z, p, k)$$

The fifth step is to apply GUI to design the controller based on transfer function
We can get two controllers by using $sisotool$ in control box in Matlab.
Finally, we can realize control by this transfer function.

2.2.2 Data Sources

26 pairs of data come from the improved fuzzy evaluation of 26 provinces (or cities) in China in 2005. The 26 provinces are random samples among 31 provinces(or cities) in China, they are HaiNan, SanXi, TianJin, ShanDong, HeiLongJiang, AnHui, ZheJiang, JiangSu, LiaoNing, JinLin, ShangHai, GuangDong, HeBei, GuangXi, FuJian, NeiMengGu, QingHai, ChongQing, GanSu, YunNan, BeiJing, GuiZhou, NingXia, SiChuan, XiZang, XinJiang. 26 pairs of nominal measurement data, which represent regional competitiveness core elements and auxiliary elements respectively and regional competitiveness output is acquired.

2.3 Regional Economic Monitor Method Based on Fuzzy Evaluation and Fuzzy Control

Fuzzy control rule is a group of multi-condition command. It can be depicted by Fuzzy Relation from the analects of input variables to analects of output variables. In this Research, fuzzy rule depicts relationship from the analects of core element and auxiliary element to competitiveness. In order to make comparison between two groups of rules, a group of rules which has 19 sentences and the other group of rules which has 5 sentences were designed by fuzzy rule edit. Applying Matlab Fuzzy controller GUI, the process was designed.

Two inputs are core element and auxiliary element respectively. The output is competitiveness. Gauss Membership function was selected for core element and auxiliary–element, and then Trimf was selected for competitiveness in the first groups of rules. Defuzzification method adopts centroid method. The range of membership function of core–element is [0 0.2]. It includes 5 curves and their central point is 0, 0.05, 0.1, 0.15, 0.2. The range of membership function of auxiliary element is [0 0.3]. It includes 5 curves and their central point is 0, 0.075, 0.15, 0.225, 0.3. The range of membership function of competitiveness is [0 0.2]. It includes 5 curves and their central point is 0, 0.05, 0.1, 0.15, 0.2. Mamdani Fuzzy System and verbose fuzzy rule were adopted. Save fuzzy controller as rule1.fis.

Gauss Membership function was selected for core element and auxiliary –element, then Trimf was selected for competitiveness in the second groups of rules. Defuzzification method and the range of Membership function of core – element, auxiliary element and competitiveness are the same as the first group. Core element membership function includes 3 curves and their name and parameter is weak, [0.06 0]; ordinary, [0.01 0.1]; strong, [0.02 0.2]. Auxiliary element membership function includes 3 curves and their name and parameter is weak, [0.08 0]; ordinary, [0.04 0.15]; strong, [0.08 0.3]. Competitiveness Membership function includes 5 curves and their name and parameter is weak, [-0.03 0 0.03]; relatively weak, [0.03 0.065 0.1]; ordinary, [0.1 0.13 0.16]; relatively strong, [0.16 0.17 0.18]; strong, [0.18 0.2 0.22]. Mamdani fuzzy system and verbose fuzzy rule were also adopted. Save fuzzy controller as rule2.fis. In order to maintain the objective of the measurement, online design mode is recommended.

In brief, some aspects of classical integration control method were compared with fuzzy integration control method in Table 1.

Table 1. Comparison between classical integration control method and fuzzy integration control method

Control methods / Items	Classical integration control methods	Fuzzy Integration Control method
Design precondition	System models are required	System models are not required
Adjustment capacity	no	Have
Control mode/design mode	Feedback control/offline	Fuzzy control/online
Controller design	Complicated	Easy and little calculation
Controller and sampling time	Relevance	irrelevance
Control performance	Related to the accuracy of system model	Related to the membership functions and fuzzy rules

3 Simulation Details

3.1 Classical Fuzzy Evaluation Result

Evaluation on regional competitiveness among 26 provinces is conducted by random sampling method. Core element index data of regional competitiveness is derived from <2003 China's Economic Statistic Yearbook>. Expert evaluation data is acquired from the works of Fan Gang and Wang Xiaolu[16]'s <General Adoption of Market Principle>, Tang Renwu and Tang Tianwei[17]'s <2002 Efficiency Measurement of China's Provincial Government> , and scores are given by relative member in our project team(five of them are professors). Expert authorized weightness is compared and determined by professors and P.H.D from our project team. Applying the classical fuzzy evaluation method mentioned in Section 2.1, the result is as follows (see Fig. 1):

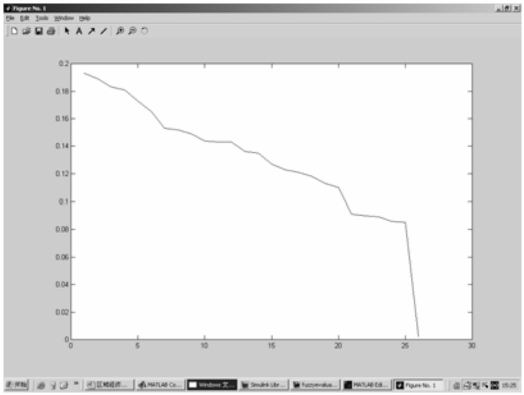


Fig. 1. Classical Fuzzy Evaluation Result

3.2 Simulation Result of Classical Control Integration Method

Using “Draw Simulink Controller” in “tool” menu, two Simulink control model can be gotten. Parallel connection of the two models can make it describes the multi-inputs and multi-outputs condition of the above ARX model. Fig.2 shows the result of the design is effective after simulating the Simulink model.

Fig. 2. Simulation result of Classical integration control methods

3.3 Fuzzy Integration Control Method Simulation Result

26 pairs of data of regional economic competitiveness elements were loaded from workspace and READFIS (‘rule1.fis‘) command were applied to load the fuzzy controller of the first group of rules (19 sentences). After designing the simulation model, the simulation output curve (see Fig.3) can be gotten. The result of regional competitiveness sequence acquired from Simulation controlled under 19 sentences is

in accordance with classical regional economic fuzzy evaluation. However, it was underestimated. It shows that the design of the first group of rules can track the trace of Regional Economic Competitiveness. But the result gotten from this controller is not satisfactory.

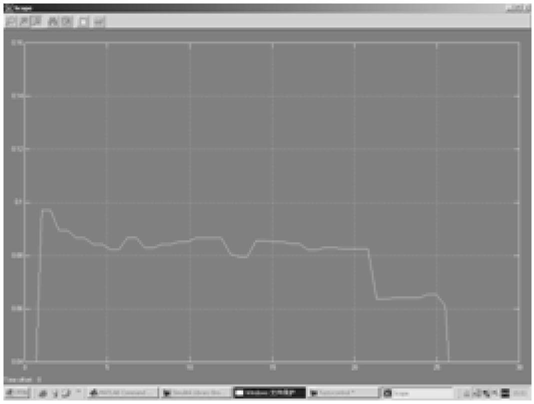


Fig. 3. Regional Economic Evaluation System Simulation Output Curve (19 rules simulation result)

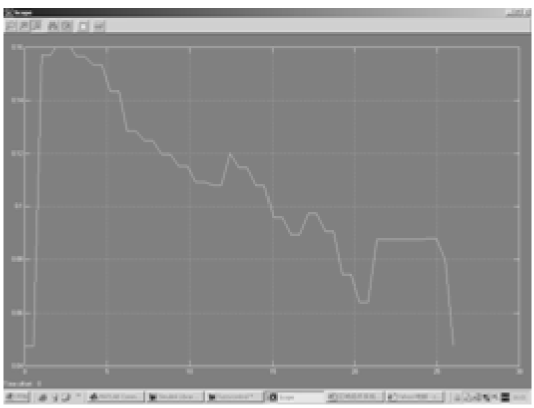


Fig. 4. Regional Economic Evaluation System Simulation Output Curve (5 rules simulation result)

READFIS ('rule2.fis') commands were applied to load the fuzzy controller of the second group of rules (5 sentences). After designing the simulation model, the simulation output curve (see Fig.4) also can be gotten. It shows that the second group of fuzzy control rules can track the regional economic competitiveness exactly (see Fig. 5). The result of regional competitiveness sequence acquired from simulation controlled under 5 sentences can realize the alarm function in the area high regional economic competitiveness, and can control the regional economic competitiveness under real time condition in the area of low regional economic competitiveness.

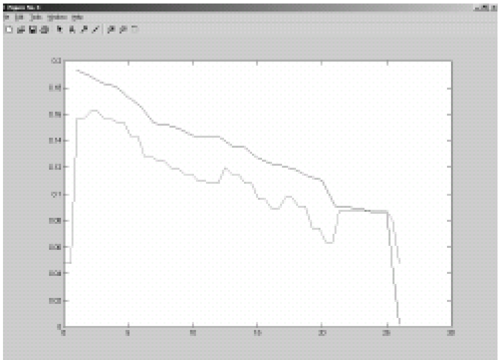


Fig. 5. Comparison result (simulation under 5 rules and output of classical fuzzy evaluation method)

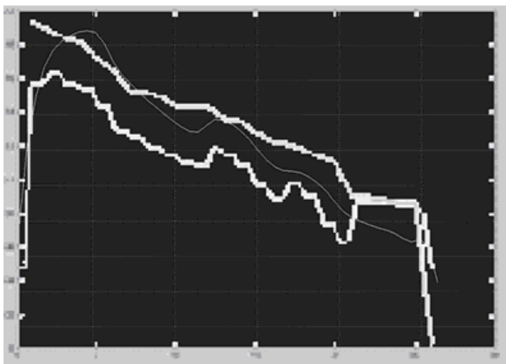


Fig. 6. Comparison result (output of classical fuzzy evaluation method, simulation under 5 rules and fuzzy integration control method)

However, the simulation results show (see Fig.6) that the curve of two simulation results is approaching the fuzzy evaluation results and has an error. If we take the fuzzy evaluation result as an objective reference system, neither method is accurate.

4 Conclusions

By demonstrating that it is possible to monitor and control the development of a regional economic system dynamically, a comparative study between the classical integration control method and fuzzy integration control method was proposed based the data from 26 provinces in China. The simulation result shows that neither the classical integration control method nor fuzzy integration control method can provide an accurate economic evaluation result, which means neither of the two methods can be employed independently. Future research work will be the construction of an expert control system by the combination of the two methods.

References

1. Zheng, S., Zheng, W., Jin, X.: An Application of System Identification Method in Regional Economic System. In: The 3rd IEEE International Conference on Wireless Communications, Networks and Mobile Computing, pp. 3870–3873 (2007)
2. Obeli, C., Urzua, J., et al.: An Expert System for Monitor Alarm Integration. Journal of Clinical Monitoring and Computing 15(1), 29–35 (1999)
3. Wu, J., Karmin, D.: A Fuzzy-Expert System-Based Structure for Active Queue Management. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3645, pp. 1013–1101. Springer, Heidelberg (2005)
4. Hadjimichael, M., Kuciauskas, A., et al.: A Meteorological Fuzzy Expert System Incorporating Subjective User Input. Knowledge and Information System 4(3), 350–369 (2002)
5. Zheng, S., Zheng, W., Jin, X.: Fuzzy Integration Monitor Method in Regional Economic System and its Simulation. In: The 26th Chinese Control Conference, pp. 2–4 (2002)
6. Astrom, K.J., et al.: Automatic Tuning and Adaptation for PID Controllers-A Survey. Adaptive Systems in Control and Signal Proceeding, 371–376 (1992)
7. Takatsu, H., Kawano, T., Kitano, K.: Intelligent self-tuning PID Controller. In: International Symposium on Intelligent Tuning and Adaptive Control, pp. 451–457 (1991)
8. Nguyen, H.T., Prasad, N.R., Walker, C., et al.: A First Course of Fuzzy and Neural Control. CRC Press, USA (2003)
9. Zhu, Y., Meng, Z., Gan, S.: Calculation of weigntness by AHP. Journal of North Traffic University 10, 119–122 (1999)
10. Charmet, G., Cadalen, T., Sourdille, P., Bernard, M.: An Extension of the 'Marker Regression' Method to Interactive QTL. Molecular Breeding 4, 67–72 (1984)
11. Arif, M., Ishihara, T., Inooka, H.: Iterative Learning Control Utilizing the Error Prediction Method. Journal of Intelligent and Robotic Systems 25(2), 95–108 (1999)
12. Cedervall, M., Stoica, P.: System Identification from Noisy Measurements by Using Instrumental Variables and Subspace Fitting. Circuits, Systems, and Signal Processing 15(2), 275–290 (1996)
13. Haubruge, S., Nguyen, V.H., Strodior, J.J.: Convergence Analysis and Applications of the Glowinski-Le Tallec Splitting Method for Finding a Zero of the Sum of Two Maximal Monotone Operators. Journal of Optimization Theory and Applications 97(3), 645–673 (1998)
14. Subasil, A.: Application of Classical and Model-Based Spectral Methods to Describe the State of Alertness in EEG. Journal of Medical Systems 29(5), 473–486 (2005)
15. Nyman, K.L.: Linear Inequalities for Rank 3 Geometric Lattices. Discrete Comput. Geom. 31(2), 229–242 (2004)
16. Fan, G., Wang, X.: Index of General Adoption of Market Principle in China. Economic Science Press, Beijing (2001)
17. Tang, R., Tang, T.: Measurement on Efficiency of Province and Local Government in China. China Administration 6, 2857–2871 (2004)

Energy Levels and Co-evolution of Product Innovation in Supply Chain Clusters

Guojun Ji

School of Management, Xiamen University, Xiamen Fujian, 361005, China
jiking@xmu.edu.cn

Abstract. In the last decade supply chain clusters phenomenon has emerged as a new approach in product innovation studies. This article makes three contributions to the approach by addressing some open issues. The first contribution is to explicitly incorporate the energy levels in the analysis. Hence, the unit of analysis is widened from sectoral systems of innovation to socio-technical systems. Hence, the unit of analysis is widened from sectoral systems of innovation to socio-technical systems. The second contribution is to suggest an analytical distinction between different evolution method, actors involved in them, and the institutions which guide actor's perceptions and activities. Thirdly, the article opens up the black box of institutions, making them an integral part of supply chain. The article provides a coherent conceptual multi-level perspective, using insights from sociology, institutional theory and innovation studies. The perspective is particularly useful to analyze long-term dynamics supply chain clusters phenomenon, shifts from one energy level to another and the co-evolution of product innovation.

Keywords: Supply Chain Cluster, Wave-particle Duality, Quantum Jump, Co-evolution.

1 Introduction

Imitation or counterfeiting by bypassing the multi-year investment, enterprises can rapidly imitate products, using the reputation of original brand and marketing strategy to reap huge profits, it has become a strategic business model of the rapid development of enterprises [1]. Counterfeit Intelligence Bureau forecasted that global trade in 2008 accounted for 5% to 7% of counterfeiting products, the value of up to $600 billion. Counterfeiting products in Chinese market share are far more than the global average level, and local protectionism results in its large presence [2-3]. It cannot be denied that the short-term effects of counterfeiting is obvious, Chongqing Lijian Motorcycle Manufacture Co. for example, Honda motorcycle brand was imitated the "Hongda", but there are firms abusing the dominant market imitation strategy to pursue profits, or by lowering the product quality and technical content to reduce costs to ensure its survival in the highly competitive market environment, then result in the disruption of the normal order of market competition, and standardized-competitive conduct is

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 140–158, 2010.

overwhelmed, product technical level is low, and the relative industrial upgrading is slow, with the inevitable impact on the country to promote innovative process.

Ecological/green-based product has become the construction key of a resource-saving and environment-friendly society. For example, the provision of pollutant emission standards indicate, enterprises which damage the natural environment will be punished, but the individual enterprise's environmental governing cost is high and its effect is limited, so we must from the whole supply chain environmental governing perspective to manage product innovation behaviors. In addition, Chinese counterfeiting products sold in the process of the world will face anti-dumping, patent protection and other multi-block, i.e., innovative enterprises usually have a variety of leading technologies and protected patents, with the international market standardization requirements and trade protection remounting, the counterfeiting product is difficult to occupy the global market and its security issues are concerned [3]. So, only through independent innovation, enterprises can gain the initiative competition in the global market. The modern enterprises in a supply chain or supply chain cluster need to actively build innovative networks, such as Dell's success is mainly due to its high efficiency throughout the value chain "network" efficiency [4]. The choice of product innovation in a supply chain and independent innovative model has become the core of the entire value chain as well as entering the world market, such as ZARA through continuous strengthening, consolidating, reengineering to reach the timely innovative capabilities to satisfy the customer demand and product innovation [5-6]. At the same time, innovation may also lead to destructive results [7-8]. Therefore, different energy-level control ability of product innovation cannot be ignored, i.e., the imbalance of innovation capacity and control ability is the reason that enterprise from the bloom to doom, as well as the loss of the root cause of vitality.

The idea of using the supply chain concept as a normative model to improve firm behavior and thus ultimately industrial innovation through the development of supply chain cluster (SCC) or integrated supply chains has been discussed since the late 1990s [9], a SCC is a geographically concentrated, self-flourishing ecosystem, including production systems. Shippers, logistics service providers, IT vendors, infrastructure providers, regulatory agencies, research institutions, consultants and other logistics-related organizations that leverage on the inter-dependencies between them, provide efficient and effective logistics solutions and create innovative new solutions. Reference [10] analyzed that simultaneous development of the SCC is possible with planning and aggressive marketing to attract multi-national companies in manufacturing and third-party logistics providers. This is important since they bring with their global experiences and also gain economies of scale. Additionally, while the supply chain process integrates all activities from product innovation to sales of a product, the integration of the service phase, physically, informationally and organizationally, with the rest of the supply chain is an emerging trend. So, a SCC is a geographically concentrated, self-flourishing ecosystem, including Shippers, Suppliers, Logistics Service Providers, IT Vendors, Infrastructure Providers of land, sea and air logistics facilities, Financial Institutions, Regulatory Agencies, Research Institutions, Consultants and other logistics-related organizations that leverages on the interdependencies between them, provide highly efficient and effective logistics solutions and create innovative new solutions. It is easy to see, a SCC almost covers all activities of society, such as “a small society”. Therefore, using by such system can improve competitiveness through linkages and integration within a cluster, as well as between industrial clusters. A SCC system consists in one or more parallel single supply chains in the agglomeration location, not only presents all enterprises in one single supply chain cooperate with one another internally, but cooperation and coordination exists across different supply chains externally too [11]. Reference [10] showed India should proactively attract investments by following the SCC paradigm where all the stakeholders in the supply chain such as manufacturers, logistics providers, financial institutions, and so on are collocated in the region creating a value chain of excellence which is difficult to replicate. New technologies adoption in a cluster often has the nature of increasing returns , because of some reasons, new technologies are usually the dominant position with the virtue of pre-emptive to lower unit costs, reach learning effects and coordinated effects, then result in popularity in the market, thereby to achieve self-reinforcing virtuous cycle in the cluster; on the contrary, one better technology because of delay introduced into market, it might be trapped in a vicious cycle, or even "Lock-in" in some invalid state. Reference [12] argued that the self-reinforcing phenomenon for the technicalogical evolution to extended to the institutional change aspects, which established the path-dependent theory of the institutional changes. The core content of the theory is that path dependence has the objective rule, the operation mechanism can be summarized in the given conditions, activating mechanism, setup state, out of lock-in and so on process. From product innovation development, energy level of product innovation has an evolution process as well as human beings recognition on the environment and the evolution of the supply chain [9]. Based on the existing research, the energy level jump mechanism of product innovation, and energy level evolution of product innovation in supply chain and other issues have not been studied.

The structure of this paper is organized as follows: the “wave-particle duality” used in a SCC is discussed, product innovation energy levels based on quantum jump theory in a SCC are presented, mechanism and model of new product innovation in a SCC are reflected in section 2; the co-evolution models are presented in section 3 to analyze the stability of the system; a case study proves our conclusions is in section 4; In the last section, some conclusions and open problems are given.

2 Product Innovation Energy Level Based on Quantum Jump

2.1 Wave-Particle Duality of A SCC

In allusion to the dispute for light is “wave” or “particle” has lasted out more than 300 years, since “wave theory” and “particle theory” of light is opposite, the final result ascribes light is neither wave nor particle, and is the substance of wave and particle with duality properties. If we call “a SCC” is “wave-motion”, and supply chains or enterprise cluster are “particle-motion”, then separation of a SCC and supply chains or enterprises is similar to antinomy for light is wave or particle on earth. The wave-motion of a SCC means that different node enterprises integrated result. There are two describing variables for “wave-motion”, i.e., “nodes” and “nodes centralized degree”. With more the social division fine, more complicated and length the supply chain is. Essence of the SCC’s wave-motion is that process of “explicit transmission” of the tangible semifinished product and “implicit transmission” of the intangible assets (e.g., information, brand, knowledge), which transmission scope is between semifinished product market and final product market or within SCC, as well as it is the process of supply chain innovation and value realization. In such transmission process, there exists the definite fluctuation like as “wave-motion” among every SCC’s nodes, so the SCC can be described as “wave-motion”. Otherwise, the organizational model of a supply chain cluster is a network structure formed by upstream and downstream correlative enterprises, every supply chain or enterprise is just like as a “particle” in a SCC. There are two describing variables for “particle-motion”, i.e., “node position” and “innovation ability”. Enterprises often are changing in a supply chain’s node position, thus results in every supply chain is changing, as well as the SCC. And enterprises whether or not find, search, hold or change their node position to reach “robustness” of the SCC, which entirely depends on their innovation ability. Movement or determination of the node position together with magnitude of innovation ability, are closely relative to cooperation and competitive competence of a SCC. Therefore, the “particle-motion” denotes the integration competence of different enterprises along with supply chains.

In quantum physics, “quantum particle” and “quantum wave” is uniform, quantum of particle must depends on “wave-motion”, and quantum of wave must depends on “particle-motion” too [13]. In fact, “wave-motion” and “particle-motion” in a SCC is uniform. “Particle-motion” indicates the production innovation of a SCC or enterprises is the drive of the cluster innovation and value actualization. The “wave-particle duality” of a SCC shows the cluster is an organization by way of social division and the division coordination to come true, which is integration of supply chains and enterprise cluster, to implement the effective disposition resources and operation in the SCC, and ensure division and collaboration favoringly. Based on the quantum physics, the Bohr atomic model has three hypotheses, i.e., energy level, jump and orbit. Correspondingly, the Bohr atomic model of a SCC also has three hypotheses as follows: (1) Energy level. Supply chains or enterprises in a SCC always are in a series of discontinuous innovative states, synchronously the product in a supply chain or its enterprises is in a series of discontinuous value state too, such case is called stationary state. Transformation from one state to another often needs definite energy, so this will engenders different energy levels. (2) Jump. Transformation of different supply chains, different enterprises or different products from one stationary state to another usually needs definite energy, in order to offset their shortage of technical and innovative competence, so this likes as jump. At the same time, the process of increasing or decreasing value likes as jump too. (3) Orbit. What is called orbit means nodes of a SCC. There exists energy of technical innovation and difference value among the nodes of a SCC, those nodes often are discontinuous. Therefore, the meaning of the model is interpreted as follows: (i) Nucleus can denote the final product supply chains or enterprises. In Bohr atomic model, proton and neutron likes as the final product supply chain or enterprises in a

SCC, there usually exists a large number of proton and neutron in the cluster. (ii) Electron can show the matching enterprise, core enterprise even supply chain or the cluster bottleneck. In a SCC, electrons located in orbit are semifinished product supply chains, enterprises or matching enterprises, which they are situated in the cluster's nodes. A node may has many enterprises, which form the different multi-level supply chains, and a SCC can be regarded as the dynamic network formed by several multi-level supply chains. (iii) Orbit can denote chain, node or matching radius. Orbit in a SCC likes as nodes of the cluster. Based in the differences of the distance from these orbits to nucleus, they in turn can hanger together from far till close to form the SCC. Since nodes often are located in different positions of the SCC, if technical innovation competence is stronger and value-added is greater, the "energy level" is higher. (iv) "Quantum jump" can denote displacement of the semi-finished product. The semi-finished product between node orbits can move via logistics activities namely "quantum jump". "Quantum jump" shows semi-finished product along its SCC to pursue consummate continuously, transmission often is from low-node (far from the final enterprise) to high-node (close to the final enterprise), then the cluster can implement the whole "jump" process from product innovation to production then to consumption. In nature, the "quantum jump" process is process of innovation and value-added in a SCC.

2.2 "Entanglement of Quantum" and "Quantum Jump" in A SCC

In the quantum physics, "entanglement of quantum" shows a phenomenon, i.e., whether how far two particles, one particle's change will affect another, they are interactional radically [14]. There is a special "entanglement of quantum" phenomenon in a SCC, whether government or enterprises themselves, only encircle some of enterprises like as "entanglement of quantum", they can foster and optimize the cluster. "Entanglement of quantum" indicates the relationship between supply chains, supply chain and enterprises or enterprises, which they are neither the entire independent market transaction relations, nor the relative close interior relations within the cluster, and is a "keep it at an arm's length" entanglement relation, such relation shows change of one supply chain or a enterprise must affect other supply chains or enterprises even the cluster, thus forming "entanglement of quantum". Change in "entanglement of quantum" incarnates the change of "quantum state". In the quantum physics, "quantum state" denotes the state of particles (such as atom, neutron, proton), can represent energy, rotation, motion, magnetic field of particles and other physical characteristics [13]. "Quantum state" in a SCC can present the state of "implicit knowledge" in supply chains or enterprises. Implicit knowledge is the supply chains or enterprises own knowledge, technology, brand, information etc. In the quantum physics, studying on "entanglement of quantum" aims to reach "implicit transmission" of quantum information. So, the objective of studying on the SCC to actualize "implicit transmission" of quantum states of supply chains or enterprises. What is called "implicit transmission" that shows just to transmit the carriers (semifinished product or product) of technology, information, brand, value etc., instead of themselves. In the SCC, the objective of "entanglement of quantum" in supply chains or enterprises is to implement “implicit transmission” of technology, band and so on. The external form of the SCC is the “explicit transmission” of semi-finished product, and the internal form is the “implicit transmission” of implicit knowledge. In addition, with the increase of the production factor price and global extension of supply chain, the government must understand its role (service, coordination or support), if it blindly issues some preferential policies, the SCC will be difficult to develop and optimize. Apparently, using by government’s preferential policies to attract enterprises, the ultimate goal is to promote the formation of “entanglement of quantum” among supply chains or enterprises. In this way, a perfect SCC can be built only according to the requirements of “entanglement of quantum”, Otherwise the cluster relationship will not stable for long.

In the quantum physics, in order to carry out long-distance quantum cryptography communication or quantum state implicit transmission, we must allow long-distance particles together with short-distance particles can hold maximal “quantum entanglement state” [14]. Similarly, in a SCC, we must try hard to impel the “quantum entanglement state” formed among supply chains and enterprises, i.e., such that “quantum state” presents entanglement state. In addition, the influence of technical standards to the SCC is self-evident, once a technical standard changes, supply chains or enterprises in the SCC will change to fit a new standard. The process of constituting standard is the process of “quantum entanglement state” formed. And the process of implementing technical standards is the process of SCC formed via the standard or implicit knowledge transmission. For example, Microsoft Inc. always perseveres in innovation over time. In fact, the technical innovative process is the process of “quantum entanglement state”, once the “quantum entanglement state” engendered, it will drive other supply chains or enterprises change virtually, the “entanglement of quantum” will appear, then results in further integration of the “explicit transmission” of Microsoft Inc. and the cluster. In the quantum physics, because there are diversified unavoidable environmental noises in the communication channel, the quality of the “quantum entanglement state” will decrease as the transmission distance increase [15]. In a SCC, the quality of “entanglement of quantum” among supply chains and enterprise also decreases as their distance increase and environmental noises. In fact, many reasons can cause the environmental noise of the SCC, such as the social economic environment, business strategies, matching model of enterprises etc. Therefore, keeping the “quantum entanglement state” and achieving “quantum jump” are correlative important to the SCC. The process of “quantum jump” is the process of the “quantum entanglement state” maintenance.

2.3 Quantum Energy Level Theory for New Product Innovation

By using “energy levels of quantum jump”, we can analyze the innovative mechanisms in a SCC. The product innovation has the following theory assumption of quantum energy levels: (1)The product innovation in members of a supply chain reflects characteristics that is interrupted by non-consecutive, the long-term, gradual change; (2) the contribution and effect of product innovation on a particular subject can be characterized by the concept of innovation energy levels; (3) Differences exists in the innovation energy-level of different participants of product innovation in the supply chain, the core enterprises are often the dominant, other members can adjust behaviors based on the comprehensive effect of environmental factors and internal mechanism; (4) Innovative state owned by individual member at a time in the supply chain has the quantum state, "uncertainty" but with a statistical expression; (5) The influence of quantum innovation on each other also has quantum state, which can be characterized through transitions of discrete energy levels. Therefore, the product innovation in a supply chain will be expressed into a four-phase of the quantum innovation model.

Phase I: Product innovation is in quantum shock-storage phase. In this phase, the relationship among supply chain members is dynamic, the coordination capability among the member enterprises is poor, interruption frequently occurs in the chain, the supply chain strategy is essentially a niche-oriented; Product innovation in the supply chain are mostly counterfeiting, mainly no cost of product innovation, and environmental awareness considered in product innovation is almost zero (the position I is shown in Figure 1), the chain belongs to the counterfeit product supply chain, energy level of product innovation is in small shocks and stable state, there is no significant energy-level jump in product innovation from macro-level, the main task of product innovation is dominated by a single core enterprise, information sharing of product innovation is low, R&D level is low, primarily depends on resources, mainly through the cognition of new products on the market and the accumulation of historical product knowledge, imitation is introduced into their own supply chain, the product innovation aims to maximize the overall benefit and market share rate; From the appearance, each firm joining the supply chain, reflects varying degrees of cooperation and innovation, but from the effect and efficiency, product innovation is relatively low-level imitation, a considerable of members do not really understand the innovation, of course, they are difficult to adapt to innovative behavior. This phase is widespread early in 1990s in China, industrial development is immature, most products are non-brand, the operation of SMEs are reflected OEM or OAM fashion, the product is often from OEM or even purely by taking the counterfeiting or low-level road.

Phase II: Innovation energy level of quantum starts jump. In this phase, supply chain strategy is still niche-oriented together with quality-driven, the coordination of members in the supply chain starts to enhance; product innovation begins to focus on the upstream and downstream cooperation, mainly prior to the integration of the forward supply chain and part of the reverse supply chain, not only through counterfeiting or original counterfeiting innovation to produce quality reliable products, but also by tracking innovation and independent innovation, begin to pay attention to the conduction effects of customer participation, experience and service in the product supply chain, innovation costs begin to increase, however the product innovation process often do not consider the product recovery and remanufacturing, with a larger external negative effects on the environment; The technological level in product innovation is still low, depends on resources all the same, but a greater energy level jump appears, industrial clusters have begun to emerge, began to adjust to each

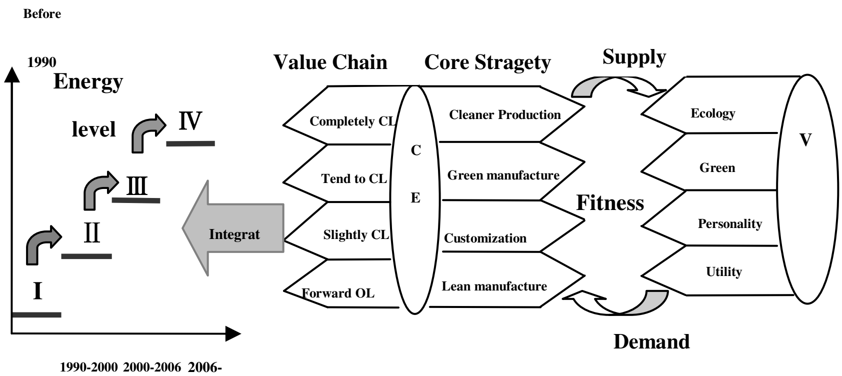


Fig. 1. Energy levels of quantum innovation (where E denotes energy, CE means the core enterprise, V denotes the customer value)

other adaptive behavior (transition process from I to II in the Figure 1). As the level of coordinative innovation understanding increasing, coordination apperception with other mechanisms result greater energy-level jump in the supply chain in innovation process, innovative knowledge has been gradually formed networking, with the innovative smart of "comprehend by analogy", the members as a whole takes on a clear innovative effects, there are some innovation product brands with high quality at home and abroad. In China, this stage is mainly reflected from 1990 to 2000.

Each member in the supply chain digests and absorbs the accumulated external knowledge from phase I, changing its own innovation tacit knowledge, where parts of the "tacit knowledge" will be continued to ferment in their minds, constantly generating new ideas, and strengthening their internal knowledge networks, so as to promote collaborative innovation; the other parts which are based on market-driven mechanism through improving customer satisfaction, building brand, to enhance the core competitiveness etc. The knowledge will be shared with other members through the network or the public, and organization communication, thus further stimulating the innovation networking and regeneration of their own and others. As a result, the adaptive behavior of each member enterprise together to form a positive feedback loops to promote continuous innovation and innovation energy-level jump.

Phase III: The quantum state of innovation energy-level belongs to relative steady-state higher phase. Along with environmental concerns, green supply chain appears, green product innovation strategy becomes one of the Government strategy, innovation energy level is relatively stable, innovation activities are mainly related to the whole forward supply chains and most reverse supply chains, not only focus on innovation and discontinuous conduction effect for customer participation, experience and service in the product supply chain, but begins to focus on the environmental governance among various members in the supply chain. Although the green supply chain management focuses on waste products and packaging recycling etc., but the consideration of reusing waste and energy is relatively lack. This phase is collaborative innovation decision-making stage or to adapt to the environment behavior process, cost of product innovation is in increasing trend (higher energy state represented III in Figure 1). Each member in the supply chain has a stable internal control mechanism, the initiative social responsibility become the main objective of product innovation, the innovation emergence period is a non-steady-state process, each member must adjust their behavior to adapt to changes of the environment, to implement the environmental governance as feasible as possible, and therefore such behavior tends to close to their preliminary environmental adaptation, and the organizational structure tends to the closed-loop. This phase is from 2000 to 2006 in China, the supply chain strategy towards value-oriented or environment-driven, energy-level jump of product innovation is not so clear as it from I phase to II phase, but the recognition of corporate social responsibility is the appropriate motivation of organizational innovation decision-making. Environmental consciousness innovation begins to encourage each member to conclude their own innovation experiences and adjust their behavioral process, while other kinds of green products theory, Government environmental regulation (e.g., WEEE) etc. are also actively pushing the outcome of the entire innovation process to change as much as possible into supply chain knowledge database for a long-term sharing in supply chain members.

Phase IV: The quantum innovation energy-level belongs to most steady-state phase. Since 2006, environmental management and industrial ecology concepts have been blended into the supply chain, then formed the highest energy-level, i.e., ecological supply chain (ESC), involving the complete closed-loop system formed by entire forward supply chain and the reverse supply chain. A ESC uses the basic principles of industrial ecology and recycling economic thought, modeling on natural ecological processes in the materials closed-loop flow and material echelons, not only concern the cost savings and internal business environment improvement and product innovation, but also give full consideration to the impact on the environment from manufacturing to distribution process, as well as how to deal with waste and emissions, recycling and reusing and so on, eliminating the external diseconomies caused by environmental pollution. The main work of product innovation in this phase is modeling on natural ecological processes reflect the innovative products of naturalization, resources refining technology, human nature (IV is shown in Figure1), supply chain strategy is toward value-oriented and ecology-driven. As with the scarcity and timeliness of social resources, as time goes on, the impact of new technologies will enable product innovation to comply with the limited nature operation. There is harmonious co-existence relationship between human and nature, the importance of innovation in the supply chain is concerned that the timeliness and reduce external diseconomies, and innovation product factors are directly proportional to zero pollution. On the other hand, only the ecological concept is applied borderless, full closed-loop operation, ecological innovative products and people, nature or enterprises can be harmonious coexistence, so the efficient evolvement of supply chain can be guaranteed. Eco-products have gradually become geared to the demand of consumers, the supply chain cluster is also contributing to the ecological operation of product innovation.

3 Co-evolution Models

Cluster supply chain is the coupling body of supply chains and industry clusters, which formed by a complete or relatively completes network-based supply chain system in the same industry cluster, the external performance has the network characteristics of industrial cluster, but also has the organizational structure of supply chain characteristics. In this system, the horizontal scale has a complete supply chain structure system from suppliers to manufacturers, distributors and end customers; the vertical scale has the highly developed division of labor and cooperation among enterprises. If the supply chains lack collaborative innovation characteristics, the enterprises in the cluster will produce homogeneous, non-differentiated products, eventually lead to vicious competition among enterprises; if no cluster features of eco-innovation networks in the cluster, supply chain organizations will because only cooperation but not of competition were eventually reduced to the low efficiency of the organization, i.e., the system refers to in the form of the composition interrelated, co-evolved continuum, which is formed in a particular space, the economic, social and ecological environment related enterprises or organizations by means of material flow, energy flow, capital flow, information flow and knowledge flow.

According to the similarity of cluster and ecosystem, co-evolution in cluster members of the ecosystem refelects inter-firm competition, cooperation, parasitism, predators and prey models to promote product innovation energy level transition, This co-evolution depends not only on cash flow, material flow, energy flow, but also depends on the information flow, knowledge flow (Jouni Korhonena et al., 2004). The ecosystem development of the supply chain cluster complies with Logistic model. Co-evolution shows that any supply chain in cluster can affect the maximum output level of another chain, considering there are two supply chains in the cluster, and then we have:

$$\dot{x}_i = r_i x_i \left(1 - \frac{x_i}{N_i + \omega_i\left(x_j\right)}\right), i \neq j, i, j = 1,2 \ . \tag{1}$$

Where x_i denotes that the level of innovation output for i supply chain is a function of time t, i.e., with t change, the level of output subjects to their own resources, access to information, innovation ability, innovation cost and other changes; $\dot{x}_i$ denotes the change rate of output level at per unit time for i supply chain; r_i is the growth rate of output level under full use of a variety of resources for i supply chain, with the continuously raising the level of output it declines and ultimately tends to zero; N_i is the limiting value of output level for i supply chain, i.e., the maximum level of output under full use of all resources; When i supply chain subjects to the impact of j supply chain, with expressed by $N_i + \omega(x_j)$, $\omega(x_j)$ means the effect or earning or loss of i supply chain by j supply chain.

When there is competitive co-evolution between the supply chains in the cluster, according to Gause's competitive exclusion theory, two species occupy the same niche will be bound to happen fierce competition (F. Ge, 2002). In the cluster system, the supply chains occupied similar ecological niche to obtain more resources or market shares, they compete with each other, the follow reflects:

$$\begin{cases} \dot{x}_1 = r_1 x_1 \left(1 - \dfrac{x_1}{N_1 - \alpha_1 x_2}\right) \\ \dot{x}_2 = r_2 x_2 \left(1 - \dfrac{x_2}{N_2 - \alpha_2 x_1}\right) \end{cases} \tag{2}$$

Where $\alpha_1, \alpha_2 > 0$, because of competition, negative α_1, α_2 means supply chain B can inhibit level of innovation output in suplly chain A; and vice versa. Equation (2) has four equilibrium points: $E_1(0,0), E_2(0, N_2), E_3(N_1, 0), E_4\left(\dfrac{N_1 - \alpha_1 N_2}{1 - \alpha_1\alpha_2}, \dfrac{N_2 - \alpha_2 N_1}{1 - \alpha_1\alpha_2}\right)$. Suppose the impact on the competition evenly matched between two supply chains, i.e., $\alpha_1 = \alpha_2$. Then, the stable point is $S_1\left(\dfrac{N_1 - \alpha_1 N_2}{1 - \alpha_1\alpha_2}, \dfrac{N_2 - \alpha_2 N_1}{1 - \alpha_1\alpha_2}\right)$, and the stability condition satisfies $\alpha_1\alpha_2 < 1, \alpha_1 < \dfrac{N_1}{N_2}, \alpha_2 < \dfrac{N_2}{N_1}$. To describe the interaction co-evolution between the supply chains, using by non-dimensional transformation, assume that $x_1 = N_1\omega_1, x_2 = N_2\omega_2, t = \dfrac{\tau}{r_1}$, and take the initial values $x_1^0 = 0.25, x_2^0 = 0.75, \alpha_1 = \alpha_2 = 0.35$, based on the simulation analysis (see Figure 2), we can see under the impact in the competition factors being equal and less than 1 (weak fair competition), two competing supply chains will eventually tend to a stable point, that point is less than $\min(N_1, N_2)$. Therefore, co-evolution of competition limits the largest level of output of two supply chains.

When there is the cooperative co-evolution between two supply chains in the cluster, for example, in the cluster there is research and development, sharing of technologies etc. The cluster can improve technological innovation to enhance production efficiency, thus promote the rapid development of the cluster, the cooperative co-evolution model follows:

$$\begin{cases} \dot{x}_1 = r_1 x_1 \left(1 - \dfrac{x_1}{N_1 + \alpha_1 x_2}\right) \\ \dot{x}_2 = r_2 x_2 \left(1 - \dfrac{x_2}{N_2 + \alpha_2 x_1}\right) \end{cases} \tag{3}$$

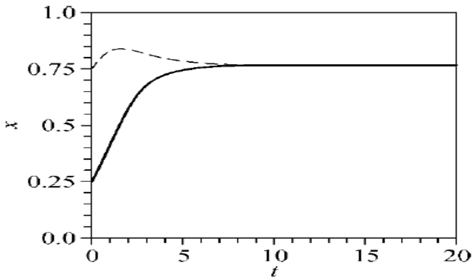


Fig. 2. The co-evolution between the competitive supply chains

Where positive α_1, α_2 means that in cooperative period, which will co-promote growth of innovation output level for two supply chains; and vice versa. Equation (3) has four equilibrium points: $E_1(0,0), E_2(0,N_2), E_3(N_1,0), E_4'\left(\frac{N_1+\alpha_1 N_2}{1-\alpha_1\alpha_2}, \frac{N_2+\alpha_2 N_1}{1-\alpha_1\alpha_2}\right)$, and the stable point is $S_2\left(\frac{N_1+\alpha_1 N_2}{1-\alpha_1\alpha_2}, \frac{N_2+\alpha_2 N_1}{1-\alpha_1\alpha_2}\right)$. Choosing the initial value $x_1^0 = 0.25, x_2^0 = 0.75, \alpha_1 = 0.15, \alpha_2 = 0.85$, using by the simulation analysis (see Figure 3), it is easy to see the final stable point of two supply chains is greater than $\max(N_1, N_2)$, therefore, cooperation is conducive to improving the innovation output capacity of two supply chains.

When co-evolution of the cluster exists wandering relations, wandering relation refers to the level of innovation output for some wandering enterprises depends on the supply chains in the cluster, while the latter level of output without much relationship with the former, the former is called wandering enterprise, The latter can be called the host supply chain, there is the following co-evolution model:

$$\begin{cases} \dot{x}_1 = r_1 x_1\left(1 - \dfrac{x_1}{N_1 + \alpha x_2}\right) \\ \dot{x}_2 = r_2 x_2\left(1 - \dfrac{x_2}{N_2}\right) \end{cases} \tag{4}$$

Where α is influencing factor for the host supply chain impact on the wandering enterprises, considering the latter little effects on the level of innovation output of the former, it can be negligible, equation (4) has four equilibrium points $E_1(0,0), E_2(0,N_2), E_3(N_1,0), E_4''(N_1+\alpha_1 N_2, N_2)$, and the stable point

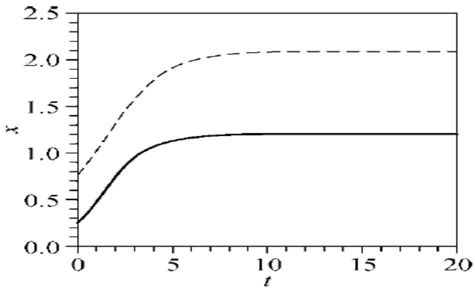


Fig. 3. The co-evolution between the cooperative supply chains

is $S_3(N_1+\alpha_1 N_2, N_2)$. Choosing the initial value $x_1^0=0.25, x_2^0=0.75, \alpha=0.4$, using by of simulation analysis (see Figure 4), it can be seen that the level of innovation output for the wandering enterprise x_1 increases with α increases, and the final stable point exceeds its maximum output level N_1, which may develop into a supporting enterprise of the host Supply Chain; and when the host supply chain x_2 keeps alone development, it reaches the maximum output N_2. The wandering relation reflects without prejudice to the host supply chain development circumstances. It can promote the wandering enterprise development.

When there is predator-prey co-evolution in the cluster, usually reflect the cluster to achieve further development among the core enterprise and the supporting auxiliary enterprises through mergers or "eating into the market", the co-evolution model is as follows:

$$\begin{cases} \dot{x}_1 = r_1 x_1\left(1-\dfrac{x_1}{N_1-\alpha_1 x_2}\right) \\ \dot{x}_2 = r_2 x_2\left(1-\dfrac{x_2}{N_2+\alpha_2 x_1}\right) \end{cases} \tag{5}$$

Where x_1, x_2 are the prey and predator enterprise, respectively, negative α_1 means a predator will impact on a prey enterprise, which can inhibit the latter's level of innovation output, positive α_2 means a prey enterprise will impact on a predator enterprise, which can promote the predator enterprise to increase the level of innovation output. Equation (5) has four equilibrium points: $E_1(0,0), E_2(0,N_2), E_3(N_1,0), E_4'''\left(\dfrac{N_1-\alpha_1 N_2}{1-\alpha_1\alpha_2}, \dfrac{N_2+\alpha_2 N_1}{1-\alpha_1\alpha_2}\right)$,

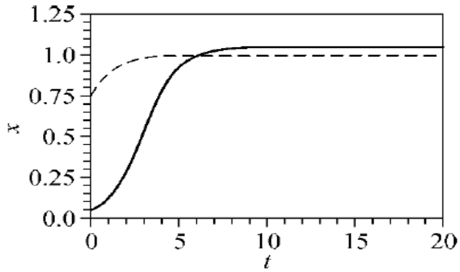


Fig. 4. The co-evolution between the host supply chain and wandering enterprises

and the stable point is $S_4\left(\frac{N_1-\alpha_1 N_2}{1-\alpha_1\alpha_2}, \frac{N_2+\alpha_2 N_1}{1-\alpha_1\alpha_2}\right)$. Choosing the initial value $x_1^0=0.25, x_2^0=0.75, \alpha_1=0.8, \alpha_2=0.2$, using by simulation analysis (see Figure 5), it can be seen that the level of innovation output of predator enterprise is greater than the maximum value of its individual development, while the level of innovation output of the prey enterprise is less than the maximum value of its individual development, and may eventually become bankrupt, i.e., the predation model can be conducive to large enterprise merge small and medium enterprises and enhance the status of core enterprise in the cluster.

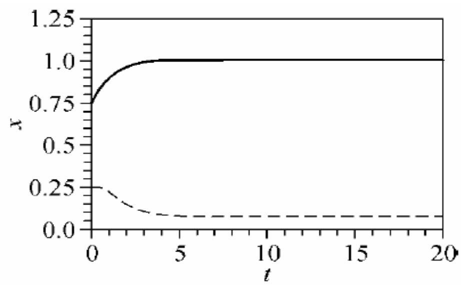


Fig. 5. The co-evolution between predator and prey

4 Case Study-Guitang Group

As the largest stock company in the Guigang city of Guangxi province in China, Guangxi Guitang Group (GTG) was established in 1954 to produce cane sugar initially. Today it is the largest sugar-making company in China which has several other industrial enterprises, such as a pulp-making plant, an alcohol plant, a cement mill and a fertilizer plant together with forming an ecological supply chain cluster, which is a single centralized SCC. The annual total production includes sugar (120,000 tons), paper (85,000 tons), alcohol (10,000 tons), cement (330,000 tons) and fertilizer (30,000 tons) so on (Zhu and Cote , 2004). All these plants are based on by-products generated from the sugar refinery.

The sugar industry has been a major polluter, especially in China, as most the refineries with a small scale can not meet the existing environmental standards. In fact, GTG development path in line with the above-mentioned energy-level jump phase of product innovation in its supply chain. Since "National Eco-industrial (sugar) Demonstration Park-Guigang" project was approved, which is first large-scale enterprises as a leading eco-industrial park construction plan. Nowadays, GTG sugar product innovation has achieved conversion of industrial pollution prevention from the end governance to the whole process control. Through these years development, GTG has formed the circular economical rudiment, and built a circular product innovation system in which product involves sugar, manufacturing paper, alcohol, light calcium carbonate etc., sugar production is generated by bagasse, molasses, filter mud and other waste after proper processing, all achieved overall recycling, waste utilization rate of 100% and the comprehensive utilization of products have been much higher than sucrose value of the main product. It is leading domestic level that has enjoyed a number of independent intellectual property rights with environmental protection leading domestic level. GTG product innovation attaches great importance to cleaner production, putting in a lot of special funds for environmental protection, widely utilized new environmental technologies and new techniques and new equipment, with emphasis on comprehensive governing the emission of slop, waste water recycling, flue gas desulfurization etc., the use of high-tech and advanced applicable technologies to transform traditional industries, and constantly enhance the efficient use of resources and environment protection. To reduce resource consumption and environmental pollution, the sugar production innovation based on GTG is established to recycle these product wastes and returns from the end of pipe to gain simultaneously economical, environmental, and social benefits. Initially, the SCC was built within GTG and slowly extended to external firms. Now the sugar ecological SCC consists of an agriculturally ecological farm, a sugar refinery, sugar distributors and retailers and other firms acting as the reclaiming agents. Reclaiming agents include an alcohol-processing plant, a compound-fertilizer plant, a pulp plant, a thermoelectricity plant, a cement mill and other recyclers. The Figure 6 shows three main innovation production flows in the ecological SCC model: (1) the forward product flow from sugarcane farmers to end users, (2) the reverse supply chain from customers to suppliers. Some returns from consumers are processed within the sugar refinery, some useless wastes are send to recyclers, such as water recovering plant.(3) the by-products flow from sugar refinery to reclaiming agents, in which the by-products generated by the sugar refinery are reused.

In order to optimize the environmental and economic performance of the whole network of companies within GTG, GTG complied with the ecological design principles to develop new technologies, optimized the production processing and adopted ecological management. Within the sugar chain, three innovation approaches can be seen to treat the residual products, i.e., reuse, volume reduction and disposal.

By-products such as sugar slag, spent molasses and filter sludge, become the raw material of the pulp plant, the alcohol plant and the cement mill separately to produce pulp, alcohol and cement. The by-products of alcohol plant, residual alcohol can also be reused by the fertilizer plant to produce compound fertilizer that is sold to the raw material producer, the sugarcane farmers. To reduce the amount of residual products, cleaner production technologies are employed. And new technologies to improve water efficiency are developed, which is expected to reduce the wastewater between 30% and 40%. As other by-products produced during sugar refinery which are the most severe pollution problems for the sugar industry, GTG has collected them, and began to use them as raw materials. In addition, the sugar refinery sent sugar slag to the thermoelectricity plant, and the reverse flow is electricity and vapor. The recycle resources from the recyclers, such as water, are also sent reversely to sugar refinery. Hence, there are some recycling flows in the sugar ecological SCC, which not only minimize the damage to environment, and maximize the utilization of resource, but also improve the whole SCC financial performance.

GTG product innovation also maintains close relationships with his primary suppliers, the sugarcane farmers. As mentioned above, GTG sells the fertilizer produced from residual products back to the farmers. Thus, the chemical fertilizer which can decrease the quality of sugarcane is avoided to be used by farmers. In addition, GTG gives technological and financial supports to farmers to improve the quality of sugarcane and resolve the production problems, encourage them to develop scale economics. The long-term contract with farmers also ensures the quantity of sugarcane and the benefit of farmers. All these efforts make innovation product

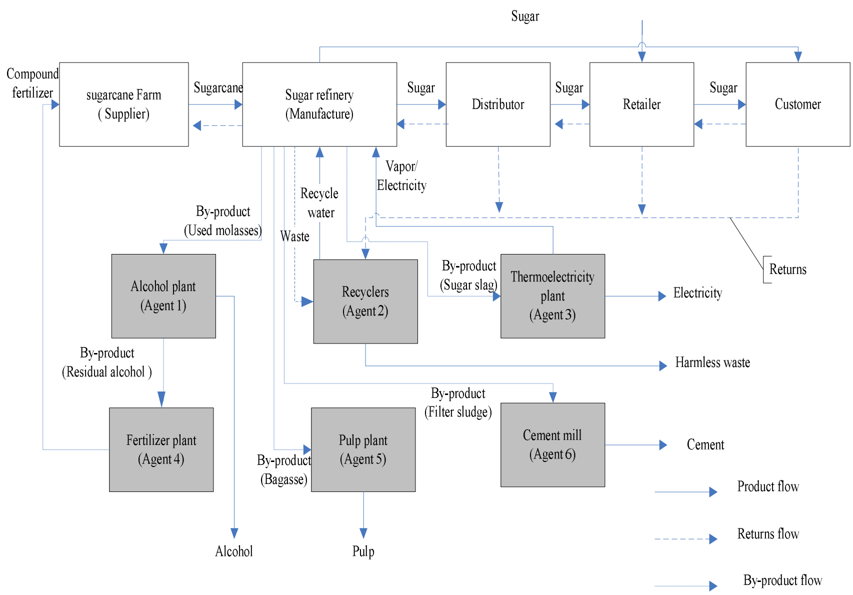


Fig. 6. The innovation production flow of Guangxi Guitang Group

(sugar) with a high quality and a low cost, which increase the competitive advantage of sugar in the international market. At the same time, GTG has worked on establishing better relationships with their customers. GTG produces the best quality sugar in China according to color, sulfur and impurity content, and thus sharing a large market. Its average sugar price was between 30% and 35% higher than that made by other Chinese sugar refineries because of good quality.

By taking full advantages of by-products and product following innovation in an environmentally, socially and economically responsible manner, GTG has realized multi-win of human, nature and society. In the product innovation of GTG, it considers traditional ‘waste’ as resources, which reduce the material cost. Besides ensuring cooperation of upstream and downstream, the ecological SCC also helps GTG to decrease the transaction costs arising from interactions within different members in the SCC. They can share updated information and easily establish cooperative relationships. As a result, some transaction costs are reduced. Since all sugarcane is produced by the farmers in the Guigang city, GTG can also benefits from the low transportation cost. Coca-Cola and Pepsi-Cola who used to purchase sugar from other countries have established joint ventures in China, now begin to buy sugar from GTG. Many domestic soft drink companies such as the Wahaha Group buy sugar only from GTG. The key reason is that the sulfur content in the sugar made by GTG is lower than that made by other Chinese sugar plants. This is due to improved environmental technologies developed by GTG and the resulting higher quality of “green” product. Barrier related to information dissemination and communication is the key obstacles to ensure the effective ecological innovation of the supply chains. Fortunately, in our case, most enterprises in the SCC are linked tightly around GTG, which facilitates communication among the members.

The government and employees of GTG also can benefit from the ecological SCC innovation management, as they can get more revenue and salary. The problem of lacking water resource has been resolved and the quality of water from the rivers is ensured as a result of by-products reusing. Similarly air quality is improved as most CO_2 and other toxic gases are processed. The rapid development of the firms in the sugar SCC drives the development of the relate service industry and increases the job opportunity and the living level of local people. Therefore, GTG can keep the “quantum entanglement state” and promote “quantum jump”, and institutional arrangement on division in such SCC is integrating continuously, GTG has made progress both economically and environmentally while most sugar companies in China are still struggling for survival.

5 Conclusions

Based on the physical theory and the wave-particle duality, a SCC is the special organization whose characteristic is wave-particle duality. The product innovation energy levels based on quantum jump theory is analyzed. We discuss mechanism and model of new product innovation in a SCC. By constructing the co-evolution models, the stability of the system is analyzed. Finally, a case study proves our conclusions.

There are some open problems include: how to determine new product innovation matching model based on innovation energy levels in a supply chain? What is the relationship between cost of product innovation and innovation energy levels? What is the relationship between product innovation energy levels and the compensation mechanism of government and enterprises? What are compensation options for ecological consumption? Whether there is isomorphic relationship among Product creations, ecological innovation and zero pollution production? What mechanism is forward and reverse coordination under reverse Engineering? What is co-opetition mechanism in the ecological homotopy innovation?

Acknowledgment

This work is supported by the National Natural Science Foundation project “A study on in continuous innovation energy level based on the product supply chain” (70971111) and Natural Science Foundation of Fujian Province project “Logistics security mechanism for emergency management research in Fujian Province”.

References

1. Cantisani, A.: Technological Innovation Processes Revisited. J. Technovation 26(11), 1294–1301 (2006)
2. Wang, C., Yan, H.: Some Thinking on Talent Management in Enterprise Internationalization Strategy. J. Special Zone Economy 7, 230–232 (2006)
3. Wang, C., Wang, X., Xu, R.: Industrial Cluster Technological Innovation Means based on Knowledge Exterior. J. Economic Tribune 15, 7–9 (2007)
4. Chopra, S., Sodhi, M.S.: Managing Risk to Avoid Supply-chain Breakdown. J. MIT Sloan Management Review 46(1), 53–61 (2004)
5. Chow, D.: Organized Crime, Local Protectionism, and the Trade in Counterfeit Goods in China. J. China Economic Review 14(4), 473–484 (2003)
6. Malerba, F.: Innovation and the Dynamics and Evolution of Industries: Progress and Challenges. International Journal of Industrial Organization 25(5), 675–699 (2007)
7. Frizelle, G., Woodcock, E.: Measurement of Complexity as an Aid to Developing Operational Strategy. International Journal of Operation and Production Management 15(5), 26–39 (1995)
8. Ji, G.: Ecological Supply Chains Performance Evaluation and Disruption Risk Management Strategies. J. Human and Ecological Risk Assessment: An International Journal 15(2), 351–370 (2009)
9. Ji, G.: Value Delivery Systems under the Instantaneous Competition. Journal of Service Science and Management 1, 29–49 (2008)
10. Ji, G.: Logistics and Matter Element Models based on Firm Innovative Supply Chains. J. International Journal of Operational Research 1(3), 283–301 (2006)
11. Wang, J.: Innovation Space: Enterprise Cluster and Regional Development. Beijing University Publishing House, Beijing (2001)
12. Li, J.: Inventory Management in Cluster Supply Chain. Press of Chinese Economy, Beijing (2006)

13. Shannon, C.E.: Mathematical Theory of Communication. Bell System Technology 27, 623–656 (1948)
14. Shinn, T., Lamy, E.: Paths of Commercial Knowledge: Forms and Consequences of University-enterprise Synergy in Scientist-sponsored Firms. J. Research Policy 35(10), 1465–1476 (2006)
15. Chen, Z., Ma, S., Chen, R.: Research on the Lean Supply Chain. J. Computer Integrated Manufacturing Systems 5(5), 11–16 (1999)

A Typology Framework of Loyalty Reward Programs

Yuheng Cao, Aaron Luntala Nsakanda, and Inder Jit Singh Mann

Sprott School of Business, Carleton University,
1125 Colonel By Drive, Ottawa, ON, Canada K1S 5B6
ycao2@connect.carleton.ca

Abstract. Loyalty reward programs (LRPs), initially developed as marketing programs to enhance customer retention, have now become an important part of customer-focused business strategy. With the proliferation and increasing economy impact of the programs, the management complexity in the programs has also increased. However, despite widespread adoption of LRPs in business, academic research in the field seems to lag behind its practical application. Even the fundamental questions such as what LRPs are and how to classify them have not yet been fully addressed. In this paper, a comprehensive framework for LRP classification is proposed, which provides a foundation for further study of LRP design and planning issues.

Keywords: Loyalty Reward Program, Typology Framework, Classification.

1 Introduction

Since their first debut as American Airlines' 'AAadvantage' program in 1981 [1], loyalty reward programs (LRPs) have been employed by a wide range of consumer goods and service industries. In the airline industry alone, more than 130 companies currently have a LRP, and 163 million people throughout the world collect loyalty-based air miles [2]. LRPs are also being offered in a number of other industries such as retail, travel, financial service, telecommunication, gaming and entertainment. Geographically, LRPs have been quite popular in the United States, United Kingdom, Canada and a host of other countries. Some recent studies show that 90% of Americans, 92% of UK consumers and 95% of Canadians have at least one loyalty card. Many of them are enrolled in multiple loyalty programs [3].

LRP was initially introduced as a marketing tool that aimed to enhance customer retention by providing incentives to loyal customers. After more than twenty years' innovative development in practices, LRP has been widely adopted as an important component of firms' customer relationship management (CRM) strategy to increase customer loyalty and long-term profitability (e.g., [1], [4]) because both practitioners and researchers find that retaining customers is less expensive than attracting new

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 159–170, 2010.

ones, and that the "best" customers are the most profitable, especially in the long-run (e.g., [5], [6]). Developing or joining a LRP is of strategic importance to a firm because such programs require the firm's long-term commitment and significant investments. Also, as many researchers have pointed out, the success of loyalty program depends on the choice of its type (e.g., [7], [8]). That requires an in-depth understanding of LRPs and their typology. However, our research shows that critical questions such as what are LRPs and how to classify them have not yet been fully addressed in academic literature.

In this paper, we seek to address these questions and propose a comprehensive typology framework. This framework would help researchers and practitioners to classify various LRPs, to study the differences among various classes of LRPs and to analyze the potential impact of those differences on their research questions or business operations. Furthermore, this framework lays a foundation for further studies in LRP design and implementation. The paper is organized as follows. In section two we present a review of previous research on LRP classification. It is followed by section three in which we elaborate on our typology framework and discuss relevant research on each classification criteria. We conclude, in section four, with a summary of the study and possible future research directions.

2 Literature Review

2.1 Definitions of LRPs

Loyalty reward programs (LRPs) are also known as loyalty programs, rewards programs, "frequent-shopper" programs or "frequent-flier" programs in the literature. "Loyalty" and "reward" are the core concepts in these programs. More specifically, "loyalty" is the primary objective of LRPs and "reward" is the key instrument for attaining that. According to [9], customer loyalty is a deeply held commitment to re-buy or re-patronize a preferred product or service consistently in the future. Although the debate regarding the meaning of customer loyalty (e.g., [10], [11], [12]) still exists in the academia, there is no doubt that customer loyalty has become an important asset to a firm (e.g., [7], [13]). It has long been proved in psychology studies (e.g., [14]) that "reward" has strong impact on person's decision making behavior, and also on behavior modification. In LRPs, rewards refer to all kinds of incentives such as discounts, cash rebate, free goods, or special services. These rewards are designed to encourage customers to keep doing business with one firm or a group of firms sponsoring the same LRP rather than with competitor firms (e.g., [15]).

It is observed that in the reviewed literature researchers have provided conceptual definitions of LRPs (see Table 1), but none of them have given a specific description of what LRPs are

Table 1. Definitions of Loyalty Reward Programs in the Literature

Literature	Definition of LRPs
Sharp and Sharp (1997) [16]	A loyalty rewards program is defined as a program that allows consumers to accumulate free rewards when they make repeated purchases with a firm. Such a program rarely benefits consumers in one purchase but is intended to foster customer loyalty over time.
Bagdonienė and Jakštaitė (2006) [8]	Enterprises create loyalty reward program in order to develop loyalty and to reward. It is an effective marketing instrument helping to create such a situation where all interested sides win.
Berman (2006) [3]	Loyalty programs are offered by both retailers and manufacturers to stimulate continued patronage among consumers through discounts, cash, free goods, or special services (such as free magazines on specialized topics of interest to loyalty program members).
Banasiewicz (2005) [17]	Brand loyalty program is a broad category of reward-based initiatives aimed at stemming customer attrition. Taking advantage of rapidly advancing informational technologies, these marketing initiatives are emerging as the most effective means of finding and retaining brands' most profitable buyers.
Lewis (2004) [4]	Loyalty programs that base rewards on cumulative purchasing are an explicit attempt to enhance retention. Such programs encourage repeat buying and thereby improve retention rates by providing incentives for customers to purchase more frequently and in larger volumes.
Sayman and Hoch (2005) [18]	A frequency (or loyalty) program is an intertemporal promotion which offers some reward or benefit to customers based on their history of purchases. Frequency programs are widely used and important promotional tools in many industries.

Here, we propose a definition of LRPs, which provides a more specific description of LRPs from a practical perspective.

For the purposes of this study a business/marketing program is considered as a loyalty reward program (LRP) when it displays all of the following features:

- The program targets a customer's long-term profitability or a customer's lifetime value.
- Customer enrollment is required.

- The program collects customer information, and records customer's purchase history of products/services through use of membership cards, co-branded credit cards, or identification numbers (e.g., login ID) by the customer.
- The program rewards repeated customer purchase behavior on the basis of customer's purchase history.
- The program has a clear rewards scheme explicitly stating how customers will be rewarded and the benefits that customers can obtain after they join the program.

2.2 Classifications of LRPs

Despite the increasing use LRPs worldwide, and proliferation of a large variety of LRPs in recent years, few studies have addressed LRP classification, except [19], [8], and [3] (see Table 2).

Table 2. Classifications of Loyalty Reward Programs in the Literature

Literature	Classification of LRPs
Kadar and Kotanko (2002) [19]	- exclusive one-company program - inclusive company-specific programs - cross company programs
Bagdonienė and Jakštaitė (2006) [8]	- open or closed LRPs - direct or indirect reward LRPs - LRPs for end customers or for intermediate customers
Berman (2006) [3]	- members receive additional discount at register - members receive 1 free when they purchase *n* units - members receive rebates or points based on cumulative purchases - member receive targeted offers and mailings

Kadar and Kotanko [19] studied LRPs in terms of organizational structure and classified LRPs into three categories: exclusive one-company programs, inclusive company-specific programs, and cross-company programs. Exclusive one-company programs refer to programs operated and controlled solely by one firm; inclusive company-specific programs refer to programs which have been extended to include a number of partners; and cross-company programs refer to the programs created by a group of companies together. Focusing on the customer perspective, Bagdonienė and Jakštaitė [8] classified LRPs into open or closed programs, programs for end customers or for intermediate customers, and direct or indirect programs. In an open program all customers are eligible to join the program, as long as they purchase goods or services from the LRP firm. In a closed LRP program only desired or invited customers can join. Programs for end customers are the business to consumers (B2C) LRPs, whereas those for intermediate customers are business to business (B2B) LRPs. Direct LRPs provide customers some financial benefits as rewards, such as permanent discounts, gift cards, or free products. In contrast, indirect LRPs provide customer rewards that are non-financial privileges (e.g., pre-board services or free access to the business lounge at airports). Berman [3] in his study discussed four types of LRPs in terms of reward scheme: members receive additional discount at register, members receive one free when they purchase *n* units, members receive rebates or points based on cumulative purchases, and members receive targeted offers and mailings.

We find that although these classifications touch different aspects of LRPs, none of them captures the full expanse of the variety of LRPs in today's marketplace. Therefore, we address this issue and propose a comprehensive framework for LRP classification in the next section.

3 A Typology Framework

The typology that we propose here is based on the relationship view of LRP system. In general, there are five fundamental entities involved in a LRP system (see Figure 1): LRP host who launches or owns the program; the program itself; LRP partners who join the program or become agents of the program by taking some management roles; LRP members (e.g. end consumers, business entities) who participate in the program; and third-party service providers who provide service or technical support for the program, but do not communicate with the LRP members directly.

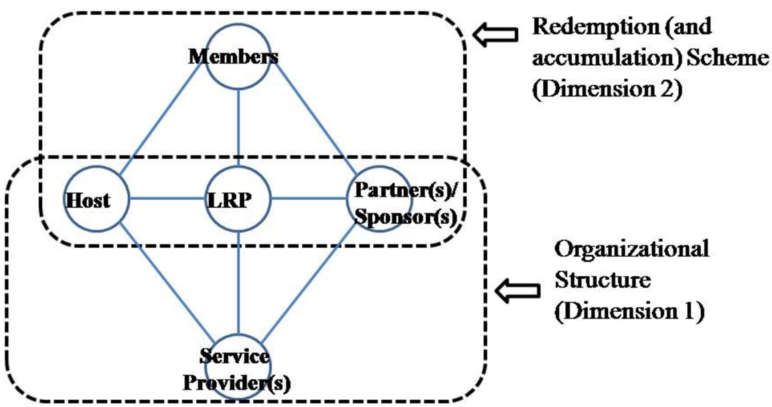


Fig. 1. A General LRP System

The variety of LRPs can be viewed as arising from the different relationships among these entities. As both organizational structure and redemption (and accumulation) scheme are the key determinants of the relationships among these entities we treat them as two dimensions of the framework. Organizational structure defines the relationships among entities in the lower triangular; while the redemption (and accumulation) scheme defines the relationships among the entities in the upper triangle, especially among members and the other three entities (i.e., host, LRP, and partners).

3.1 Typology Dimension One—Organizational Structure

Organizational structure defines the relationships among LRP host, LRP service providers, and LRP partners. Several researchers have noticed differences of LRPs along this dimension. Gudmundsson *et al.* [20] identify three structure models of airline frequent flier programs (FFPs). We find that these structure models are also common in LRPs other than FFPs. Along this dimension, LRPs are grouped into three categories: type A, single sponsor programs; type B, multi-sponsor programs; and type C, joint programs. Under types A-B, there is a sub-dimension, which we adopted from [20]. The graphs in Figure 2 show the differences among these structure models.

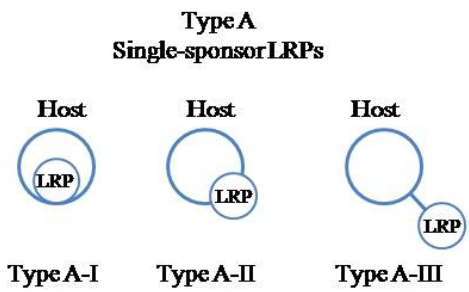


Fig. 2. Organizational Structure Models: Type A (I, II, III)

Types A-I, A-II, and A-III illustrate structures where an LRP is fully and solely owned by an LRP host. In type A-I, the LRP is an internal unit of the LRP host and fully managed by the host. Type A-II represents the structure where an LRP is partially managed by an LRP host and some of the management functions are outsourced to other firms or a third party; whereas in the type A-III model, all of the management functions of the LRP are outsourced. The "exclusive one-company programs" mentioned in [19] are quite similar to the Type A-I LRPs that we define here.

In the above models, the LRP host is the sole sponsor offering accumulation and redemption to LRP members. In general, such LRPs are limited in flexibility and are narrow in scope. We noticed that in recent years many existing LRPs have been restructured to contribute to value growth. Some of those third-party service agents in Type A-III model replaced the traditional LRP host enterprises (e.g., airline companies or retail companies) and became LRP hosts themselves. Meanwhile, in order to compete effectively, LRP hosts started to offer products and services in different categories through partnership with other non-LRP firms ([21]). Therefore, Type B and Type C structure models (see Figure 3) have appeared in recent years and have become more and more popular in large-scale LRPs.

Type B models (I, II, and III) are extensions of Type A models (I, II, and III). In these Type B models, LRPs follow the same ownership and management structures as those in Type A models, but have multi-redemption and/or multi-accumulation partners/sponsors. The type B-IV model represents the structure where LRP and related services are the focal business of the host firm (e.g., Aeroplan, Air Miles). Type B LRPs are known as multi-sponsor programs or coalition programs. The "more inclusive company-specific program" described by [19] is similar to Type B-I LRPs in this typology.

In contrast to these structure models, the Type C model represents the structure in which an LRP does not belong to any individual firm (i.e., no sole host) and is formed when a number of firms band together to develop a joint program, and is known in the literature as joint LRP program (e.g., [22]) or cross-company program (e.g., [19]).

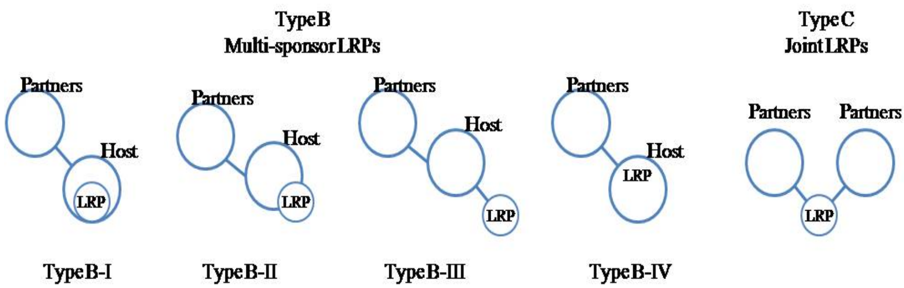


Fig. 3. Organizational Structure Models: Types B and Type C LRPs

In Types B and C LRPs, customer loyalty is no longer built around a product or a company but around the LRP program and the associated reward system. Such programs not only possess significant advantages on an operational scale, and offer a wider range of benefits to members, but more importantly they can be leveraged for cross-selling to customer bases that belong to other partners ([19]). In these LRPs, each company brings different capabilities to the table and each may take away a different form of enhanced value. In this way they are structured as win-win solutions for both LRP host and LRP partners.

3.2 Typology Dimension Two—Redemption (and Accumulation) Scheme

The second dimension of the typology framework is based on redemption (and accumulation) scheme. Redemption (and accumulation) scheme primarily defines the relationship between LRP members and LRP. It is also identified in the literature as an element essential to the administration and positioning of LRPs.

Fundamentally, there can be two strategies, static and dynamic, for designing redemption (and accumulation) schemes. Term 'static' refers to those schemes that are not changed over time. They usually take the form of "one scheme applicable for all LRP members". This type of scheme is used quite often in large retail industries such as the Bay, Shoppers' Drug Mart and Loblaws. Common rewards offered to LRP members are cash back and gift cards. These firms have a large customer base and they essentially deal with products. Their marginal cost and marginal revenue per customer are low. In these firms, LRPs are purely viewed as marketing tools and usually operated as cost centers. Dynamic reward schemes are quite often used by enterprises that specialize in LRPs or the big companies in service industries that have a large number of partners. The pricing, type and timing of redemption (and accumulation) in these schemes change over time, and are structured differently for different member segments. Dynamic schemes offer LRP members more accumulation and redemption choices. But on the other hand, members are required to put in much more effort to be constantly familiar with the 'ever-new' schemes in order to get most benefits out of it. Although the design and administration costs are higher, dynamic schemes have much more flexibility and capability to make the LRP a real profit center for the firms. Compared to the dynamic schemes, the static schemes are much simpler, requiring lower design and administration costs and lesser learning effort on the part of members.

Overall, LRPs can be classified as either LRPs using static scheme or LRPs using dynamic scheme. In addition, no matter which design strategy is used in a LRP, the LRP can be further classified based on four other scheme-related criteria: reward medium, redemption (and accumulation) timing, reward type and redemption (and accumulation) grid.

3.2.1 Reward Medium

In many LRPs, the relationship between member's purchase effort and a final outcome is mediated by the presence of an intermediate currency known as reward 'medium' (i.e. points or other exchange units). As Duffy [1] pointed out, the communication between a LRP host and LRP members may get confusing when no proper unified medium is used. Points, miles or vouchers are the most popular mediums that are used to link members' spending to rewards.

Most existing LRPs use a single-medium (e.g. expenses → points → rewards), the rest are either no-medium LRPs (e.g. expenses → rewards) or multi-medium LRPs (e.g. expenses → points → vouchers → rewards). A few researchers have started to examine the impact of medium of LRPs. Shi and Soman [23] examined the impact of single-medium vs. multi-medium on members' evaluation of LRP. They found that multi-medium reward program functionally outperform the equivalent single-medium program. The multi-medium LRP is more attractive to LRP members, which in turn results in a positive effect on members' purchasing behavior. Therefore, LRPs can be classified based on reward medium, as: no-medium LRPs, single-medium LRPs or multi-medium LRPs.

3.2.2 Reward Type

Reward type refers to the type of rewards offered by an LRP. This criterion has been studied in previous LRP literature. Dowling and Uncles [24] classified LRPs into either direct-reward LRPs or indirect-reward LRPs. Direct-reward LRP refers to the LRP offering rewards which directly support the value proposition of the product/service that LRP host or sponsors provide (e.g. gift card, free tickets). The main purpose of offering direct-reward is to keep customer loyalty on one product, one company or one LRP brand (e.g. Aeroplan). Indirect-reward LRP is defined as the LRP presenting rewards that causes the LRP members indirectly to buy the products or service. This type of reward has no linkage with the product/service.

The typical indirect-reward is cash-back. Kim *et al.*[25] examine the decisions on selecting direct or indirect rewards (firm's own product or service vs. cash) when the firm faces different customers (heavy user vs. light user, price sensitive user vs. insensitive user).

In today's marketplace fewer LRPs offer indirect-reward, because it is believed that direct-reward is more appropriate for creating loyal customers. Some LRPs offer bundles of direct and indirect rewards to their members (e.g., [26]). Therefore, under this criterion, LRPs can be grouped into LRPs offering indirect rewards, LRPs offering direct rewards, or LRPs offing mixed rewards (i.e. bundles of direct- and indirect- rewards).

3.2.3 Redemption (and Accumulation) Timing

Redemption (and accumulation) timing is used by some researchers to refer to whether a LRP offers immediate or delayed rewards (e.g., [27], [28]). Delayed rewards are benefits and incentives that are obtained or are redeemable at a later date from the point of sale. Conversely, immediate rewards refer to benefits that are availed at the point of transaction. Examples of immediate rewards include direct-mail coupons, discounts or price cuts offered to customers at the point of transaction. However, as immediate rewards do not relate to customer's cumulative purchasing behavior over time, they are less effective in retaining consumers than delayed rewards ([29]).

From our point of view, redemption (and accumulation) timing involves the variable of time in the redemption (and accumulation) scheme. For example, in some LRPs, points accumulated or rewards available for redemption expire after some days. Obviously, using redemption (and accumulation) timing can increase flexibility in LRP operations; however, it creates time pressures on LRP members. Members must remain active in order to keep their status in the program. For instance, members are required to acquire or redeem points, sometimes up to a certain minimum level within a certain time period. This type of requirements may have a negative impact on members' valuation of the program. Atahk [30] found that the potential of an LRP to attract members is determined by the value of the rewards it offers, as well as by the timing of the rewards available. In reality, whether to use redemption (and accumulation) timing or not depends on an LRP's overall strategy. Some LRPs use redemption (and accumulation) heavily, while others choose not to use it at all ([31]). Therefore, based on the redemption (and accumulation) timing criterion, LRPs can be classified into two categories: LRPs using reward timing or LRPs not using reward timing.

3.2.4 Redemption (and Accumulation) Grid

Redemption (and accumulation) grid refers to the detailed reward prices (and accumulation options) offers for members. Redemption (and accumulation) grid can be stated in terms of number of points or miles, amount of LRP members' spending, members' purchase frequency, or members' portfolio. Here portfolio refers to, for instance, the combination of members' spending, product category, and membership 'status' or 'tier'. Therefore, based on this criterion, LRPs can be classified as LRPs using an amount-based grid, LRPs using a frequency-based grid, or LRPs using a portfolio-based grid. To our knowledge, earlier research work has not discussed this criterion formally.

Based on the above discussion, we present a two-dimension typology framework in Figure 4 below.

Fig. 4. A Typology Framework for LRPs

4 Summary and Future Research

In this paper first we have reviewed previous research on loyalty reward programs typology. Then, two classification dimensions, organizational structure and redemption (and accumulation) scheme, have been identified based on the analysis of LRP system. A two-dimension typology framework for loyalty reward programs has been developed and explained. This framework would help academic researchers and practitioners to distinguish different facets of LRPs systematically and to identify the key characteristics that are important for LRP design and implementation. Using this typology framework, it will be possible in future to examine the status of LRPs in the real business world, identify the recent trends of LRP development and explore the critical factors that influence firms' decision on LRP selection. This framework has been developed in the context of business to customers (B2C) LRPs. As a future research, it would be interesting to explore business to business (B2B) LRPs and verify the presented framework in B2B context.

References

1. Duffy, D.: Customer Loyalty Strategies. Journal of Consumer Marketing 15(5), 435–448 (1998)
2. Funny Money. Economist 377(8458), 102–103 (2005)
3. Berman, B.: Developing an Effective Customer Loyalty Program. California Management Review 49(1), 123–150 (2006)

4. Lewis, M.: The Influence of Loyalty Programs and Short-Term Promotions on Customer Retention. Journal of Marketing Research XLI, 281–292 (2004)
5. Reichheld, F.: The Loyalty Effect, the Hidden Force Behind Growth, Profits and Lasting value. Harvard Business School Press, Boston (1996)
6. Nunes, J., Dreze, X.: Your Loyalty Program is Betraying You. Harvard Business Review 84(4), 124–128 (2006)
7. Liu, Y.: Long-Term Impact of Loyalty Programs on Consumer Purchase Behavior & Loyalty. Journal of Marketing 71(4), 19–35 (2007)
8. Bagdonienė, L., Jakštaitė, R.: Customer Loyalty Programs: Purpose, Means and Development. Organizacijų vadyba:sisteminiai tyrimai 37, 21–35 (2006)
9. Oliver, R.: Whence Customer Loyalty. Journal of Marketing 63, 33–44 (1999)
10. Dick, A., Basu, K.: Customer Loyalty: toward An Integrated Conceptual Framework. Journal of the Academy of Marketing Science 22(2), 99–113 (1994)
11. Palmer, A.: Relationship Marketing: A Universal Paradigm or Management Fad. The Learning Organization 3(3), 18–25 (1996)
12. O'Malley, L.: Can Loyalty Schemes Really Build Loyalty? Marketing Intelligence & Planning 16(1), 47–55 (1998)
13. O'Brien, L., Jones, C.: Do Rewards Really Create Loyalty? Harvard Business Review 73(3), 75–82 (1995)
14. Ebert, J.: Psychological Approaches to Future Rewards: Sequences, Valuation, Effort, and Frequency Programs. Advances in Consumer Research 30, 68–71 (2003)
15. George, C., Cawthorne, J.: Systems and Methods for Operating Loyalty Programs, `http://www.freepatentsonline.com/EP1483708.html` (last accessed, March 2010)
16. Sharp, B., Sharp, A.: Loyalty Programs and Their Impact on Repeat-Purchase Loyalty Patterns. International Journal of Research in Marketing 14, 473–486 (1997)
17. Banasiewicz, A.: Loyalty Program Planning and Analytics. Journal of Consumer Marketing 22(6), 332–339 (2005)
18. Sayman, S., Hoch, S.: Dynamic Choice in Frequency Programs. working paper, University of Pennsylvania (2005)
19. Kadar, M., Kotanko, B.: Designing Loyalty Programs to Enhance Value Growth. Mercer on Transport and Travel 8(2), 28–33 (2001)
20. Gudmundsson, S., de Boer, E., Lechner, C.: Integrating Frequent-Flier Programs in Multilateral Airline Alliances. Journal of Air Transport Management 8, 409–417 (2002)
21. Hofer, D.: Reinforcing the Value of Frequent Flyer Miles. White paper, CEO Loylogic Inc. (2008)
22. Mortimer, R.: A Tang of Bitter-Sweet Loyalty. Brand Strategy 164, 16–19 (2002)
23. Shi, M., Soman, D.: Multi-Medium Reward Programs, `http://www.rotman.utoronto.ca/bicpapers/pdf/04-07.pdf` (last accessed, March 2010)
24. Dowling, G., Uncles, M.: Do Customer Loyalty Programs Really Work? Sloan Management Review 38(4), 71–82 (1997)
25. Kim, B., Shi, M., Srinivasan, K.: Reward Programs and Tacit Collusion. Marketing Science 20(2), 99–120 (2001)
26. Nunes, J., Park, C.: Incommensurate Resources: Not Just More of the Same. Journal of Marketing Research 40, 26–38 (2003)
27. Yi, Y., Jeon, H.: Effects of Loyalty Programs on Value Perception, Program Loyalty, and Brand Loyalty. Journal of the Academy of Marketing Science 31(3), 229–240 (2003)

28. Keh, H., Lee, Y.: Do Reward Programs Build Loyalty for Services? The Moderating Effect of Satisfaction on Type and Timing of Rewards. Journal of Retailing 82(2), 127–136 (2006)
29. Zhang, Z., Krishna, A., Dhar, S.: The Optimal Choice of Promotional Vehicles: Front-Loaded or Rear-Loaded Incentives? Management Science 46(3), 348–362 (2000)
30. Atalık, Ö.: The Impact of the Reward Program Elements on Airlines Customers: Case of Turkish Airlines, `http://www.e-sosder.com/dergi/1417-33.pdf` (last accessed, March 2010)
31. Wells, J.: Exporting the Loyalty Business: Hoarding, Frustrating, Winning, `http://www.globeadvisor.com/servlet/ArticleNews/story/gam/20080823/RCOVER2` (last accessed, March 2010)

The Knowledge Transfer Factors under Different Perspectives*

Yuanhua Hu and Lili Xue

College of Business Administration, Zhejiang University of Technology,
Hangzhou 310014, China
sunnylili1983@hotmail.com

Abstract. Based on a study of knowledge transfer within different types of business, between both domestic and international R&D partners, knowledge transfer success was found to be associated with several key variables: communications, information technology, behavioral and comprehensive. This article offers a process framework of knowledge transfer. The framework identifies stages of transfer and factors that are expected to correlate with difficulty at different stages of the transfer. Each stage of the transfer is discussed to explore the predictive power of different factors at different stages of the process.

Keywords: Knowledge Transfer, Influence Factors, Knowledge Success.

1 Introduction

Teece (1977) has put forward the concept of knowledge transfer in the research on international transfer of technology [1]. Since the rapid development of information technology in 1980s, the leading value of knowledge has been generally approved. Knowledge transfer and its influence factors have gradually become the focuses of research.

In the past few years, scholars have made a large amount of research on the influence factors of knowledge transfer from different angles, and have formed various results, which have essential inner-relation. On the basis of analyzing and reviewing existing achievements, this paper abstracts out the following several visual angles, and their inner-relation is explored.

2 Influence Factors of Knowledge Transfer under the Visual Angle of Communication

From the mechanism of knowledge transfer, this kind of scholars mainly studies the series process of encoding, sending, propagating, accepting, and decoding systematically,

* This paper is one of the staged achievements of the technology plan program of "Zhejiang Study of the influence factors and countermeasures of the personal knowledge organization strategy of private enterprise of Zhejiang"(project No. 2008C25026).

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 171–179, 2010.

and the influence factors such as characteristics of knowledge, subject of knowledge transfer, communication media, etc., which play important roles on this process.

2.1 Characteristics of Knowledge

Characteristics of knowledge mainly include tacitness, specificity, complexity and usefulness of knowledge.

Tacitness of knowledge mainly means that knowledge is difficult to be encoded. Implicit knowledge is the experience and know-how accumulated by individuals, groups and organizations for a long time, and it is often attached to corresponding carrier, such as organization individual, service of goods, database, organization action convention or course, etc., so it is difficult to express completely in characters and language, etc., and its communication needs the participation of more social factors, and learning by doing or master and apprentice system, etc. is favorable to the transfer of this kind of knowledge.

Specificity of knowledge mainly means the relation between transferred knowledge and background knowledge. The knowledge of strong specificity needs the assistance of knowledge of other relevant fields for the aim of being understood. Szulanki (1996) has found out that, tacitness or heterogeneity of knowledge makes it possible that causal ambiguity might be produced during knowledge transfer, which increases the transfer difficulty [2].

The complexity of knowledge means whether it is possible to use carriers such as characters, charts or languages, etc., to transmit the subject of knowledge during the process of knowledge transfer. Kogut & Zander (1993) thought that, the complexity of knowledge will influence the recipient's assurance to the integral nature of knowledge, damage the movement of knowledge, and make the transfer inefficient. Galbraith found out after comparing the productivity of the recipient and the productivity of knowledge source during knowledge transfer that, when technology is uncomplicated, the recipient's productivity resumes quickly [3]. Cummings & Teng (2003) also found out through research that, the more difficult to express the knowledge is, the more difficult the transfer process is [4-5].

Usefulness of knowledge means influencing the transfer efficiency mainly through influencing the motive and enthusiasm of the participants of knowledge transfer. The historical record of knowledge being used to solve problems, etc. can strengthen the transfer confidence of the participants, and finally form benign guide to the transfer process.

2.2 Subject of Knowledge Transfer

Subject of knowledge transfer is the core of knowledge transfer activity, and its influence runs through the whole transfer process.

2.2.1 The Encoding, Communication Ability and Experience of Knowledge Source

The encoding, communication ability and experience, etc. of knowledge owner directly determine whether knowledge will be distorted in the course of transfer, and thus influence transfer efficiency and cost. The transfer experience of the knowledge owner influences the sensitiveness to the "existing knowledge" and the choice efficiency of transformation way of the owner. The encoding and communication ability can effectively reduce the information loss in the transfer process.

2.2.2 Absorbing and Keeping Ability of the Knowledge Receptor

Knowledge receptor is the key to determine whether knowledge can be utilized well and kept continuously and even spread again and increased of value. Cohen and Levinthal (1990) have put forward the concept of absorptive capacity, thinking that absorptive capacity is a kind of ability to understand new information and absorb it and further apply it to commercial purpose, and lack of absorptive capacity will hinder the effective transformation of knowledge [6-7]. And the relevant intensity between knowledge stock and new knowledge and the "homogeneity" between knowledge source and knowledge receptor have great influence on absorptive capacity. Meanwhile, the experience of knowledge receptor influences the searching cost of the knowledge receptor directly, and certain relevant experience of history can impel the knowledge receptor to differentiate and analyze and lock knowledge better. Szulanki (1996) further found out that, the weaker absorbing and keeping capacity of knowledge receptor will cause the viscidity of knowledge, therefore reduce the result of transfer [2].

2.3 Transfer Media

The transfer media of knowledge refers to all tools and means used for data and information transfer. It includes knowledge encoding and transfer channel.

Any transfer of knowledge includes the encoding course, which is done through languages, or behavior, or characters and data, etc. The encoding choice of knowledge source and the acceptance of knowledge receptor to encoding will influence the transfer efficiency. So, rational, effective and scientific encoding way is helpful to the smooth transfer of knowledge.

Transfer channel refers to the transfer means of knowledge after encoding, such as demonstration, face-to-face talk, characters, telephone, and network communication, etc. Channel determines the efficiency and cost of transfer to a certain extent. However, all knowledge is not applicable to the same kind of channel. For example, for explicit knowledge, characters or network, etc. can make very good result, but for some knowledge difficult to explain, demonstration or face-to-face talk may make better result. According to the study of knowledge transfer among organizations, Xu Zhanchen and He Mingsheng (2005) have found out that, the retardancy of media have a very obvious influence on knowledge transfer. So it is of great significance to have high-efficient and diversified smooth transfer channels [7-9].

3 Influence Factors under the Visual Angle of Information Technology

As information application to knowledge transfer is deepened constantly, many domestic and international scholars have begun to study information technology, for instance, the influence of management system and knowledge database, etc., on knowledge transfer.

The knowledge transfer in modern society regards information technology as carrier or assistant resource. Ripe and open information technology system and information data facilitate improving the efficiency of knowledge sharing. Nonaka (2001) discovered that, information technology is helpful to enterprises to realize the socialization and externalization of the implicit knowledge of source customers and work partners [4-5]. Li Sheting (2003) thought that, technological environment is the prerequisite for the externalization of implicit knowledge. If staff can have easy access to high-efficient knowledge spread systems such as knowledge database, digital library and BBS that can show one’s own impressions, the recycling of knowledge will be facilitated, therefore work time and exchange cost will be saved [10]. Mc DonoughⅢ and Kahn (1996) found out that the innovative enterprises with high performance usually have high-quality hardware communication technology [11]. But existing with such research results are some negative sound. Johannessen, etc. (2001) pointed out that: information technology will cause enterprises to pay over-attention to explicit knowledge while ignore implicit knowledge, and finally cause the miss-management of implicit knowledge [12]. But according to general views, information technology is greatly helpful in improving the efficiency of knowledge spread and the broadness of knowledge extension.

4 Influence Factors under Behavior Visual Angle

These scholars mainly study the factors influencing knowledge transfer from the visual angle of individual or organizational behavior, and they pay close attention to individual's motive and demand even more.

4.1 Transfer Motive

Transfer motive includes not only the transfer motive of knowledge source but also the receiving motive of the receptor. Since knowledge transfer may cause the fact that individuals or organizations may lose the exclusive right of knowledge, and there is extra cost of transfer, the organizations or individuals with unique knowledge all incline to implement “information monopolization”. And from the viewpoint of knowledge receptor, the existence of NIH (not invented here) psychology and the uncertainty of knowledge absorbing cost and knowledge value, etc. all have restraining function to knowledge transfer. Especially for implicit knowledge, its own problems such as difficultly to be decoded and situation compatibleness make the value of knowledge difficult to be weighed and uncertain. This also influences the participation enthusiasm of the transfer subject. Then, what is the motive of knowledge transfer subject in participating in the knowledge transfer?

Davenport & Prusak found out that: reciprocity, reputation and interest play a role of paying mechanism on the internal knowledge market of enterprises, and knowledge contributors hope to receive “feedback, appreciation and smile”. This has well explained why “knowledge owners like to answer questions of colleagues lying on the dressing room wall, but are unwilling to write down the same knowledge in a database". According to the research of Constant *et al*, Kiesler & Sproull, the participation of individuals and organizations in knowledge sharing comes out of their self-interest, direct or general reciprocity, self-expressive needs, and organizational citizenship behavior. Vno Krogh found out that, the empathy among organizational members can promote the knowledge sharing inside enterprises. These researches indicate, people share knowledge for the realization of such internal motives as reciprocity, empathy, self-expressive needs, etc., besides such external motives as self-interest, etc.

Scholars also found out that, external motives and internal motives are not totally independent. The interaction between such two kinds of motives are called "crowding effects" economically. I.e. forced external reward will produce weakening function on internal motives such as inherent interest, etc., and it is called "hidden costs of reward" or "the corruption effect of extrinsic motivation" in psychology.

According to the study of the four knowledge sharing mechanisms including knowledge base, knowledge exchange, informal exchange, and practicing association, Ksthryn M.Bartol (2002) found out that, economic encouragement has important influence on the knowledge sharing behavior based on knowledge base [13]. Hippel (1987) also found out that, individual's economic income expectation influences the transfer decision of knowledge source [7]. The researches of these scholars have offered theoretical foundation for economic encouragement means. But some other scholars' research supports social encouragement. Gee Woo Bock& Young-Gul Kim (2003) pointed out that: the view of expected cooperation and contribution is the decisive factor that makes individuals forms the view of knowledge sharing, which does not have too great relation with material reward [8]. Sandra Jones (2002) also made research on the relation between improving staff's work conditions, staff's participation in decision-making and staff's willingness to share knowledge, and he pointed out that only by improving staff's work conditions and giving staff the chance to participate in decision-making can staff's motive of sharing knowledge be excited [8]. So, from this angle, it seems that social encouragement is more effective.

4.2 Situational Factors

Knowledge transfer situations mainly include individual relation and organization relation, etc., which influence the whole course of transfer.

Individual relation mainly refers to the trust and common understanding between individuals and the identity sense among groups. Garnovetter (1973) made such an original proposition that, the difference of relation intensity among individuals causes the difference of information interchange, and compared with weak relation, strong relation is more favorable to the transfer of new information [13]. As the transfer of implicit knowledge, must require the frequent and much interaction between knowledge owners, harmonious relation and mutual trust can promote the knowledge exchange among individuals. Meanwhile, cognitive theory thinks that: The higher the knowledge transfer subject's identical degree of cognitive structure is, the easier the communication between them is and the smoother the transfer is. So the common understanding formed between knowledge transfer subjects on the basis of knowledge background, work experience and moral identity, etc. can also facilitate the mutual knowledge absorbing. But Lu Bing (2007) found out that, the influence of difference in knowledge and thinking between people on the efficiency of knowledge transfer is a kind of reversed U relation, because it reduces the subject efficiency of the knowledge obtained, but at the same time, rational difference in knowledge and thinking can increase the efficiency of abundance of knowledge [14]. According to his study, Bouty (2000) held that, the familiarity, competitiveness and trust degree between individuals are the three key variables that influence the knowledge transfer between individuals inside organizations. Organization relation includes organizational culture and leadership style, etc [13]. The study results of Davenport and Prusak (1998) indicate that, if the culture has no sharing knowledge, even if the knowledge transfer mechanism is ripe, knowledge transfer may not reach the anticipated result [14]. Kostova (1999) thought that, similar organizational situation, i.e. the organizational culture supporting the receptor to study and learn, can influence the performance of transfer [7]. Wei Jiang and Wang Tong'an (2006） found out after making empirical research on 10 scientific research enterprises in Xi'an: Policymaker's attitude greatly influences the transfer of knowledge from individual to organization [15]. Policymaker's style influences the forming of organizational culture to a certain extent, and the organizational culture encouraging innovation and sharing will promote the knowledge transfer inside organizations greatly. Meanwhile, the promotion of open and trustable atmosphere inside organizations to knowledge transfer is extremely crucial. Diane Ford held that: individual's trust to organizational system and individual's relation trust with superior administrators will produce more knowledge application.

5 Influence Factors of Transfer under Comprehensive Visual Angle

As the study is deepened, scholars discovered that the influence factors of knowledge transfer coexist and interact, and study of them purely from just one angle is not complete. So, the influence factors are more studied from broader visual angles. There are such main views as process view, marketing, relation quality view, etc.

Process view summarizes the research results of communication and more motive situations. Knowledge transfer exhibits the complicated course of interaction among transfer subject, transfer content and transfer media in particular environment. Domestic and international scholars made many kinds of divisions to knowledge transfer process, but they are identical in essence. Generally knowledge transfer process is divided into: beginning stage, implementation stage, adjusting stage and combing stage. According to study, the main influence factors of the several stages of transfer process can be summarized as follows: the beginning stage is mainly influenced by the causal ambiguity of knowledge application, knowledge usefulness and the degree of knowledge source being trusted; the main influence factor of the implementation stage is the relation between cooperators; the adjusting stage is mainly related to the absorbing capacity of receptor and the support and help of knowledge source; and the main influence factor of the combining stage is the capacity to change environment and accept organization. Among them, the causal ambiguity determined by knowledge characteristics influences the whole process of knowledge transfer, and with the deepened launching of knowledge transfer, the influence of organizational characteristic of knowledge source fades gradually, and the influence of knowledge receptor's characteristic is gradually obvious. While the influence factors in the transfer background gradually develops from whether knowledge owner and receptor have good relation at the beginning stage to whether there are the system and culture suitable for new knowledge application.

Sun Ting (2007) made a research on the influence factors of knowledge transfer process by employing relation matching theory [8]. She mainly divided transfer process into the following several stages: preparatory stage, transfer stage and knowledge combining stage, and pointed out that the main activities at preparatory stage are forming of transfer motive, determination of transfer knowledge and transfer organization and negotiation, so this stage mainly involves the emotion matching that influences the transfer motive and the cognition matching that influences the determination of knowledge; while at the transmission stage of knowledge which is the core course of knowledge transfer, complicated interaction begins to generate among transfer subject, transfer knowledge and transfer media, and the interaction mainly includes decoding, sending and absorbing of knowledge. The main influence factor of this stage is the matching of capacity; at the combining stage when organization makes internalization and storage of knowledge, it mainly carries on internalization and absorption in knowledge system by individuals and then carries on re-innovation. So, the biggest influence factor of this stage is the matching degree of cognition.

The scholars of market view combined such theories as behavior motives and communication, etc. Wang Kaiming and Wan Junkang (2000) pointed out that, the size of knowledge transfer cost is related to knowledge itself, the talent of knowledge sender and receptor, and their knowledge accumulation [9]. And the information management cost after transfer has certain influence on the whole transfer process. So, whether the transfer subject can obtain enough motive force and encouragement is the key to the emergence of transfer.

The scholars holding relation quality view mainly study the influence of the relation between transfer environment and transfer individual on the transfer, such as organizational culture, organizational physical and structural distance, and cultural distance, etc. Choi& Lee (1998) found out that the key factors influencing alliance enterprises' knowledge transfer is culture, including corporate culture, national culture and commercial culture [9]. Roy Lubit (2001) also thought that, individual's tendency to knowledge transfer is influenced by a lot of aspects of organizational culture, for example, whether enterprises encourage innovation, tolerate faults, pay attention to the role of knowledge and talent, support the development of personal relation network, etc [14]. Xu Xiaojun and Wang Yuanyuan (2008) also found out that, national culture have influences on each stage of the knowledge transfer inside transnational enterprises [15].

Physical and structural distance of organization means geographical distance and organization operation mode separately. Geographical distance influences exchanges will, exchange way and exchange cost. While the organizations in such forms as chains and strategic alliance, etc. can transfer knowledge high-efficiently compared with the organizations in other forms. Cultural distance of knowledge means the similarity among the task and situation between knowledge source and knowledge receptor and the knowledge structure. This mainly influences the transfer absorbing and application stage of knowledge. Especially for implicit knowledge, the similarity of task and situation between groups makes it easier for both sides to distinguish and confirm the knowledge source and knowledge demander. Meanwhile, certain background similarity improves the capacity of knowledge transfer, receiving and absorbing.

6 Research and Review of the Influence Factors of Knowledge Transfer

As stated above, the existing researches on the influence factors of knowledge transfer have made great achievement, and there are the following characteristics:

First of all, because of the limitation of research approach and understanding, the existing researches on the influence factors of knowledge transfer are more concentrated on the stages of theory study and regularity search, and are in lack of combination with practice. The existing researches made multi-angle analysis on the influence factors of enterprise's knowledge transfer, and also extracted various influence factors theoretically, but they did not make thorough demonstration or only carried on simple demonstration through some models, so, it is unable to weigh the function size of each influence factor to the transfer process. Especially the researches of some domestic scholars lean more to the exploration of theory but lack case study. So its guidance to enterprise's knowledge management practice is obvious not enough.

Secondly, knowledge has different transfer levels with individuals, groups and organizations, etc., and these three levels restrict and promote each other. While, relevant research at present failed to offer sufficiently clear definitions to individual knowledge, colony knowledge and organization knowledge, and even failed to make more systematic study on the relation among the three. During the study of knowledge transfer process, there are such problems as intentionally indistinct transfer levels, or confused transfer levels. To make the essence of knowledge transfer really clear, it is necessary to consider particular variable according to particular level, and further investigate its inherent mechanism, so that comparatively true and referential research result could be got.

Finally, since people have generally realized the complexity and tacitness of knowledge, most researches both at home and abroad are concentrated on implicit knowledge, while little on explicit knowledge. And the duplicating and spreading efficiency of explicit knowledge is high and enormous energy has been accumulated through lots and lots of years of human development. How to propagate, employ and combine this kind of knowledge fast is an enormous subject of the mankind. Meanwhile, existing research mainly pays close attention to the transfer process from knowledge source to knowledge receptor, and the study on such aspects as the emergence, appreciation and popularization and application of knowledge is relatively scarce. So, the aspect of knowledge appreciation deserves more exploration and study.

References

1. Teece, D.: Technology Transfer by Multinational Firms: the Resource Cost of Transferring Technological Know-how. J. The Economic Journal 87, 242–261 (1977)
2. Galbraith, C.S.: Transferring Core Manufacturing Technologies in High Technology Firms. J. California Management Review 32(4), 56–70 (1990)
3. Cumming, J.L., Teng, B.S.: Transferring R&D Knowledge: the Key Factors Affecting Knowledge Transfer success. J. Journal of Engineering and Technology Management 20(2), 39–68 (2003)
4. Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company. Oxford University Press, New York (1995)
5. Nonaka, I., Toyama, R., Konno, N.: SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation. J. Long Range Planning 33(1), 5–34 (2001)
6. Davenport, T.H., Prusak, L.: Working Knowledge: How Organizations Manage What They Know. Harvard Business School Press, Boston (1998)
7. Constant, D., Kiesler, S., Sproull, L.: What's Mine is Our, or is It? A Study of Attitudes about Information Sharing. J. Information Systems Research 5(4), 400–421 (1998)
8. Vno Krogh, G.: Care in Knowledge Creation. J. California Management Review 40(3), 133–152 (1998)
9. Bock, G., Kim, Y.: Breaking the Myths of Rewards: An Exploratory Study of Attitudes about Knowledge Sharing. J. Information Resources Management Journal 15(2), 37–42 (2002)
10. Lu, B., Liao, X., Yu, L.: Analysis on the Knowledge Transfer inside an Organization. J. Science Research Management 28(6), 22–30 (2007)
11. Tan, D., Huo, G.: The General Model of Knowledge Transfer Process. J. Contemporary Economy & Management 28(3), 11–14 (2006)
12. Sun, T.: The Study of Relationship between Stage and Effect of Knowledge Transfer from Employee to Organization. Master Thesis, Zhejiang University of Technology (2008)
13. Choi, D., et al.: Latecomer Strategies: Evidence from the Semiconductor Industry in Japan and Korea. J. Organization Science 9(4), 489–505 (1998)
14. Lubit, R.: Tacit Knowledge and Knowledge Management: The Keys to Sustainable Competitive Advantage. J. Organizational Dynamics 4, 164–178 (2001)
15. Xu, X., Wang, Y.: National Culture Influence on Internal Knowledge Transfer of Multinational Corporations-based on the Perspective of Hofstede's National Culture Dimensions. J. Science of Science and Management of S.&.T 29(4), 86–91 (2008)

IT Operational Risk Measurement Model Based on Internal Loss Data of Banks*

Xiaoling Hao

Department of Information Management and Engineering,
Shanghai University of Finance and Economics,
Shanghai, China, 200433
haoxiaolingsh@gmail.com

Abstract. Business operation of banks relies increasingly on information technology (IT) and the most important role of IT is to guarantee the operational continuity of business process. Therefore, IT Risk management efforts need to be seen from the perspective of operational continuity. Traditional IT risk studies focused on IT asset-based risk analysis and risk-matrix based qualitative risk evaluation. In practice, IT risk management practices of banking industry are still limited to the IT department and aren't integrated into business risk management, which causes the two departments to work in isolation. This paper presents an improved methodology for dealing with IT operational risk. It adopts quantitative measurement method, based on the internal business loss data about IT events, and uses Monte Carlo simulation to predict the potential losses. We establish the correlation between the IT resources and business processes to make sure risk management of IT and business can work synergistically.

Keywords: IT operational risk; risk measurement, Monte Carlo simulation.

1 Introduction

Risks in IT or computer-based information systems are key issues in most organizations, especially in banking industry where almost no operations of business processes can be performed without IT support. Organizations have become so dependent on IT that even a relatively short loss of the availability of a critical system or the breakdown of a network can lead to a total failure of the business. Many banks suffer from great losses because they lack the daily operational risk management which in the long run will adversely influence the reputation of the institution. Therefore, the operational continuity of business process is a major role of IT, and IT Risk management efforts need to be seen from the perspective of operational continuity. In this paper, we define IT operational risk as: the risk that IT can't adequately support the business processes in achieving the business goals, due to the

* Supported by National Science Foundation of China: 70971083 and Leading Academic Discipline Program , 211 Project for Shanghai University of Finance and Economics (the 3rd phase).

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 180–191, 2010.

unavailability or misuse of IT resources. IT resources include software, hardware, network, database, people, data, systems, etc. Breakdown of the network, malfunctions of information systems, software bugs, or unauthorized change of the database, are all factors that can cause an interruption in the business process and bring about economic losses.

IT operational risk management is becoming a growing concern for regulators and institutions as it can help to add business value. Hence it is important to measure and control the underlying IT risk that can affect the performance of the business processes. As a branch of operational risk, IT related operational risks can be studied with the well-proven methods found in operational risk research. Under Basel II, banks put a high premium on the collection of loss data to satisfy the requirement of quantitative computation of regulation capital. Continuing this trend, it is sensible to establish a methodology based on the business loss data caused by IT events.

Traditional asset-threat-vulnerability IT risk analysis methods are subjective. The estimation of the threat and vulnerability are based on the knowledge and judgment of a risk analyst rather than on objective loss data. In addition, they can't combine the IT risk with the business process operation, and cannot adequately reflect the loss due to disruption of operations. Studies on operational risk mainly focus on the methods of how to compute the capital charge, but few studies consider how to reduce the operational risk and how to estimate the effectiveness of the control mechanism.

To overcome the difficulties of previous studies, this paper presents a comprehensive model to measure IT risk. Since IT operational risk is highly correlated with the business process operation, it should change with the operation process, and it should be adjusted dynamically to business changes. Therefore, we first establish the correlation matrix between IT resources and the business processes, and determine how to collect data. Then we use the probability theory to describe the probability distribution of the severity and frequency of risk cells for each information technology resource. The estimation of the parameters of these distributions can be computed with the internal observation during a certain period. Then Monte Carlo simulation can be applied to predict future business losses and identify the key loss area, and control mechanisms can be designed and carried out accordingly. To clarify the methodology, we invented a hypothetical bank and designed the scenarios to apply the model to the case. In actuality, the model should be loaded with real data which could be applied in current practice.

This model differs from previous IT risk studies in several respects. First, it builds the correlation between IT resources and business processes, which enables correct diagnosis of the potential IT and business area. Second, the method is established on objective observation data, which is relatively rigid and systematic, and which can be run through the lifecycle of IT operational risk management. Third, it presents a method of predicting risk factors with Monte Carlo simulation, which provides the basis for control mechanisms.

2 Literature Review

There are three main academic perspectives on IT operational risk: IT perspective, business perspective, and risk perspective.

Earlier studies from the IT perspective are based on the asset-threat-vulnerability analysis and the loops of risk identification, analysis, evaluation and control, where the commonly used model is the IT risk matrix. The main defects of such methods are that it’s difficult to identify or assign a value to assets and all the data are based on subjective judgments and lack statistical information that would make it possible to determine frequency. Few of these IT risk models in the IS risk literature explicitly consider business goals and expectations. These IT risk models are established from the viewpoint of the IT department, considering the technology-related risks, which have limited usefulness for business managers. Thus it’s hard to combine the IT risk with the business process operation, which results in the failure of IT risk management. In addition, it cannot adequately reflect the loss due to disruption of operations.

With the infusion of IT into the business processes, how to combine IT risk and business risk has drawn great attention from many scholars. Sharon Halliday et al.(1996) put forward a method with a focus on risks that threaten critical business processes, rather than each individual IT asset [12]. This method uses value-chain analysis from an IT perspective (availability and functionality), to ensure the continuity of essential business processes. Bomil Suh, Ingoo Han (2003) advanced an IS risk analysis approach based on a business model, which adds organizational investigation to traditional risk analysis, and uses quantitative approach to measure the value of IS assets from the perspective of operational continuity [2]. Steven Alter, Susan A. Sherer (2004) presented a general, but broadly adaptable model of system-related risk, which encompasses: goals and expectations, risk factors and other sources of uncertainty, the operation of the system or project whose risks are being managed, the risk management effort, the possible outcomes and their probabilities, impacts on other systems, and the resulting financial gains or losses [13, 15]. Michael zur Muehlen, Michael Rosemann (2005) addressed the topic of risk management in the context of business process management, and presented a taxonomy of process related risks and discussed how this taxonomy can be applied in the analysis and documentation of business processes [9]. Jordan, Ernestl (2005) built an integrative IT risk governance model that meets the wider needs of corporate governance [5]. James L. Worrell， Ashley A. Bush (2007) conducted a survey of the perceptions of 13 information technology risk factors, among which “lack of organizational alignment between businesses” is relatively high [8]. Hannu HS Salmela (2008) adopted the business process analysis approach to analyze the business losses caused by information system risk, which associated information systems availability risk with potential losses [7]. Most of these methods are based on theoretical analysis, not experimental observations or data analysis, and can’t support the cost-benefit decision.

Since the approval of new regulatory guidelines known as Basel II for banking, quantitative operational risk studies have become popular. This agreement includes a regulatory capital charge for OR, under which Banks should adequately manage their OR in order to assume lower levels of capital. There are three approaches to set capital charges for operational risk: (1) The Basic Indicator Approach, (2) The Standardized Approach, and (3) The Advanced Measurement Approaches. Each approach requires a greater investment in processes and procedures than the one that precedes it. One of advanced is Internal Measurement Approach (IMA), under which the capital to be allocated is computed as a quantile (expected loss + unexpected loss) but rather than modeling losses to a particular distribution. Another advanced approach is Loss Distributional Approach (LDA). The idea of LDA is to fit approaches severity and frequency distributions over a predetermined time horizon, typically annual. Popular choices include exponential, weibull, lognormal, generalized Pareto, and g-and-h distributions [4]. The best fitting model is then used to produce compound processes for the annual loss distribution, from which VaR and capital estimates may be derived. Many researchers established or improved more sophisticated methodology with mathematics and statistics in order to accurately calculate the capital charge.

Recent studies on the financial industry are looking for combination of qualitative and quantitative models for operational risk. Peters, G. W. and Sisson, S. A. (2006) extended the range of models admissible for Bayesian inference under the LDA operational risk model, as it provided a mathematically rigorous paradigm to combine observed data and expert opinion [11]. Dominik, D. et al. (2009) advanced a new approach, based on a Bayesian inference method, which allows for a combination of these three sources of information to estimate the parameters of the risk frequency and severity distributions [1]. Feng Cheng etc. (2005) proposed a methodology modeling operational risk based on business process models. By connecting the generation of a probabilistic network with the business process model, this approach enables changes in the operational risk model whenever different aspects of the business process in the financial institution changes [5].

With the requirement of quantitative measurement under Basel II, many banks begin to collect data on loss which can provide statistical evidence. Thus increased studies are based on the operational risk data collected over many banks by business line and event type. Therefore, it is feasible and sensible to adopt quantitative risk method based on internal data, which can provide meaningful, repeatable and consistent results in the future.

3 IT Operational Risk Measurement Model

In this section, we establish a systematic operational risk measurement model, which includes a six-stage close-loop procedures to measure IT operational risk and evaluate the effectiveness of controls. It can be combined with business process management. As shown in figure 1.

3.1 Decompose of Business Processes and Classify Information Technology Resource

Since IT is embedded in the business processes, business processes are the most important part of IT risk management. IT event results from the breakdown or unavailability of IT resources which leads to the interruption of the business process. Thus, the continuity of the business processes depends on the availability of IT resources. In order to better manage IT related risk, business managers have to combine the IT risk closely with the business process. Without doing so, IT risk management is limited to the IT department and can’t get the attention of the business department.

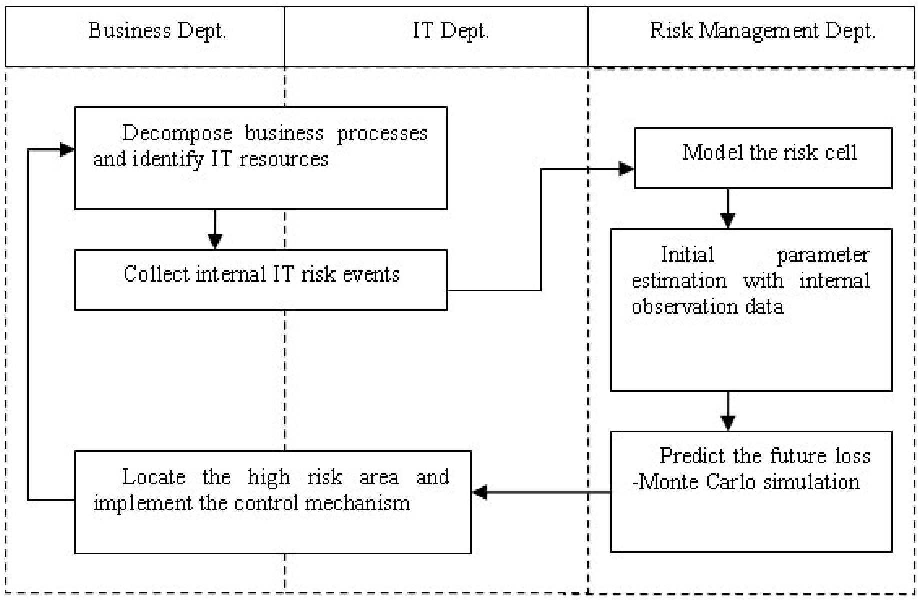


Fig. 1. IT risk measurement model

In this model, the first step is to identify and sort the business process under different business functions. IT operational risk analysis should be based on business activities of the organization. According to Basel II, there are 8 business lines in banks. There are a large number of business processes in every business line. The commonly used approach is functional decomposition, where each business line is divided into several functions, and then each function is further divided to several business processes. Each process can further decompose into the activities and tasks. The division granularity is determined by business size or complexity. For brevity, we only use the business process as the lowest level. Meanwhile, this work can be done in coordination with the IT control structure. Under the Sarbanes Oxley Act of 2002, most banks have established their internal control structure, where the processes have been similarly decomposed.

Another critical work is to classify IT related resources. The resources can be further divided into sub-classes, for example, there are many information systems in a bank, each can be regarded as a sub-class. Similarly, the granularity should be in line with the business requirements.

Here we establish the relevance matrix to describe the dependency of business processes on IT resources. Suppose a bank has n business processes, and m kinds of IT resources, Rij means the IT risk cell that has close relation with the ith business process and the jth IT resources. In this way, the two-dimension risk cell matrix is established, each risk cell corresponds to a particular business process and IT resources. In practice, banks can have their own classification according to their practical business lines, so every bank has its own IT risk profile.

Table 1. IT risk profile according to business process and IT resource

Type of IT resources (1…m) Business processes(1…n)	1 Software	…	J Hardware	…	m network	Total
1	Ri1 (risk cell)		R1j		R1m	R11 ;…R1j ;…R1m
…						
i	Ri1		Rij		Rim	Ri1 ;…Rij ;…Rim
…						
n	Rn1		Rnj		Rnm	Ri1 ;…;Rij ;…Rim
Bank-wide	R11 ;..Ri1..Rn1		R1j ;…;Rij ;…;Rnj		R1m ;…;Rim ;…;Rnm	

3.2 Collect Internal IT Risk Events

In this stage, IT events loss data are collected based on internally-defined business processes and IT resources. In order to ensure the consistency of loss data collection, it's important to establish the two-dimensional risk record mechanism. That is, when an event happens, the relevant business operator relates it to affected business process, while the IT department should take responsibility of the problem diagnosis, and then relate it to the corresponding IT resources. The advantage of this method is that it ensures accurate record of the IT events, since both the IT department and the business department make a coordinated decision. It is also easier to make parameter estimation respectively.

In collecting the data, we should consider the case that the occurrence of an IT event will impact over two business processes. A survey [2] clearly highlighted the fact that individual losses often cut across business lines and information technology resources. A couple of loss events affected up to 8 different business lines, while most of the cross-business events affect two business line, which account for 87%. For the sake of correlation between IT and business, we regard an event that influences n business lines as n events, recorded in n business lines respectively.

3.3 Model the Risk Cell

In this section, we provide some preliminary statistical analyses to describe the frequency and severity distribution of IT operational losses. We chose the commonly used Loss Distribution Approach (LDA) methodology to establish the loss distribution. The LDA approach is computationally intensive but has the advantage of taking into account the frequency and severity distributions. The two distributions are first estimated individually, and then the aggregate loss is naturally defined as a random sum of individual losses, and finally the appropriate risk measures are added across all business processes and information technology resources. In applying the LDA methodology, we assume that each type of IT operational risk satisfy the following assumptions: the frequency and the severity of the IT events are two independent random variables and each has its own probability density function, and two different losses within the same homogeneous business line or information technology resource are independent and identically distributed.

For each event unit, we can calculate an annual loss in a single risk cell. And the total loss is the sum of individual losses, as shown in formulas 1.

$$L = \sum_{k=1}^{N} F_k * S_k \tag{1}$$

Where Fk is the annual number of events frequency, and Sk, k = 1, . . . , N, are the severity of each event.

According to previous studies, we can establish the severity model for a particular IT risk cell of a bank. Suppose the loss frequency follows a Poisson distribution with the parameter λ, as shown in formulas 2. At the beginning, the parameters of these distributions are unknown and should be computed.

$$P(F = f) = \frac{\lambda^f}{f!} e^{-\lambda} \tag{2}$$

Let us assume the following severity model for a particular operational risk cell of a bank. S is the severity of the risk cell. $\ln s$ is normally distributed random variable with parameters u and σ, as shown in formulas 3.

$$P(S; u, \sigma) = \frac{1}{\sigma s \sqrt{2\pi}} e - \frac{(\ln s - u)^2}{2\sigma^2} \tag{3}$$

3.4 Initial Parameter Estimation with Internal Observation Data

With the observation data of IT events during the second stage, each risk cell is estimated with best prediction. Maximum likelihood estimation is a valid method to estimate the parameter, as described in literature of Shevchenko and Wüthrich (2006). It is adopted in this paper.

For the frequency distribution, the initial parameter of lambda can be calculated as formulas 4:

$$\hat{\lambda}_{ij} = E\ (\ X\ ) \tag{4}$$

Where, X is the average event frequency during a fixed period, usually one year.

Equivalently, parameters μ and σ can be obtained with maximum likelihood method, as shown in formulas 5 and 6.

$$\hat{u} = \frac{\sum_s (\ln s_k)}{n} \tag{5}$$

$$\hat{\sigma}^2 = \sum_s \frac{(\ln s_k - \hat{u})^2}{n} \tag{6}$$

Where S_k is the severity of the k^{th} event.

3.5 Predict the Future Loss Distribution—Monte Carlo Simulation

Once the best estimate of the parameters of severity and intensity has been selected, then if the data about frequency and the severity of IT events are given, the total loss can be predicted. Since the distribution of the total loss is complicated and does not follow a specific distribution, it may only be calculated by simulation.

Monte-Carlo approach is mature and has a wide range of application in risk management. In this paper, we use Monte Carlo simulation to forecast the distribution of the aggregated losses for each type of event separately. Then we can add up the losses according to each business line and information technology resource. The steps to apply Monte Carlo simulation are as follows:

(1) Create a parametric model for frequency and severity distribution of IT events.
(2) Estimate the parameter for each risk cell according to the sample data.
(3) Generate a set of random inputs with the estimated parameters for each risk cell.
(4) Repeat step 3 for multiple times.
(5) Summarize the annual loss for the whole bank according to the business lines or IT resources.
(6) Adding quantiles over the risk cells to find the quantile of the total loss distribution.
(7) Analyze the results using histograms, summary statistics, confidence intervals, etc.

3.6 Locate the High Risk Area and Implement the Control Mechanism

Through the Monte Carlo simulation, we can compute the potential loss of every IT risk cell, and it's easy to know where the potential loss is high. The final purpose of risk measurement from the management focus is to design controls to lower the losses from events. The managers should identify those crucial IT resources that may influence the current business objectives, and then design control mechanism accordingly. Meanwhile, it also enables the cost-benefit analysis, since the potential loss can be calculated.

After the internal control is implemented, a new round of observations begins. Then the parameters will be updated respectively with subsequent observation data in the new cycle.

4 Case Study

In this section we illustrate the methodology by applying it to a hypothetical bank. The case is not intended to be a general statement regarding the risk status of a bank. The purpose of using this generalized bank model is to explain how to apply the approach to the practice. The methodology can be used in actual practice of the banking industry on the condition that enough data can be gathered, where the best estimates of various parameters for banks can be computed.

Suppose there are only 3 business processes, and 4 information technology resources for a certain bank. The observation data over the last year is listed as follows. As for those events that influenced over 2 business processes, it is recorded as 2 events.

Table 2. Record of IT events of the last year

IT Event ID	Severity ($10,000)	Business process affected	Event type
1	1.17	BP1	ITR1
2	1.29	BP3	ITR4
3	1.00	BP1	ITR3
4	1.02	BP2	ITR3
5	1.06	BP1	ITR1
6	1.02	BP2	ITR2
7	1.40	BP3	ITR2
8	1.59	BP2	ITR1
9	2.02	BP3	ITR1
10	0.60	BP1	ITR3
11	3.02	BP3	ITR2
12	2.08	BP1	ITR1
13	0.50	BP3	ITR4
14	1.34	BP1	ITR2
15	6.26	BP1	ITR1
16	2.00	BP2	ITR2
17	1.00	BP2	ITR3
18	3.03	BP1	ITR1
19	0.80	BP3	ITR2
20	1.26	BP1	ITR2

Next step is to sort up the above data according to the information technology resource and business process. Thus, the data after handling are listed as follows.

Table 3. Frequency of IT events of record of the last year

	ITR1	ITR 2	ITR3	ITR4	Total
BP1	5	2	2	0	9
BP2	1	2	2	0	5
BP3	1	3	0	2	4
Total	7	7	4	2	20

Table 4. Severity of IT events record of the last year

	ITR1	ITR 2	ITR3	ITR4	Total
BP1	13.6	2.6	1.6	0	17.8
BP2	1.59	3.02	2.02	0	6.63
BP3	2.02	5.22	0	1.79	9.03
Total	17.21	10.84	3.62	1.79	33.46

According to above model, the frequency follows the poisson distribution, and the severity follows lognormal distribution. We adopt the maximum likelihood estimation to calculate the parameter $\hat{\lambda}$, $\hat{u}$, $\hat{\sigma}$, which are listed in table 5, table 6 and table 7.

Table 5. Parameter of $\hat{\lambda}$ for each IT risk cell

	ITR1	ITR2	ITR3	ITR4
BP1	5	2	2	0
BP2	1	2	2	0
BP3	1	3	0	2

Table 6. Maximum likelihood of parameter $\hat{u}$ for each IT risk cell

	ITR1	ITR2	ITR3	ITR4
BP1	0.7781	0.2619	-0.2554	x
BP2	0.4637	0.3565	0.0099	x
BP3	0.7031	0.4255	x	-0.2193

Table 7. Maximum likelihood of parameter $\hat{\sigma}$ for each IT risk cell

	ITR1	ITR2	ITR3	ITR4
BP1	0.6528	0.0300	0.2553	X
BP2	0	0.3366	0.0100	X
BP3	0	0.5695	X	0.4739

Through the simulation, we can know that the potential high risk areas are risk cell R11, R23, R32, which also shows the high correlation between the BP1 and ITR1, BP2 and ITR3, BP3 and ITR2. As table 7 shows. In this way, the IT problems can be diagnosed quickly; the manager can trace the high risk information technology resource to the business process and the IT event, and then take some preventive measures to strengthen the IT resources or to improve business processes.

Table 8. Mean loss for each IT risk cell

	ITR1	ITR2	ITR3	ITR4
BP1	13.5553	2.575958	1.595312	X
BP2	1.586607	2.998822	13.33978	X
BP3	2.011117	5.361535	X	1.798566

In addition to separate analysis of each risk cell, the distribution of total loss can also be described by the graph, where managers can have a clear vision about the potential economic losses.

In addition, Monte Carlo simulation also enables a bank to determine the percentile. The percentile level can be determined at the management level. Usually, the higher the percentile, the higher reserve capital that banks should set aside to cover the potential operational risk. As shown in table 9.

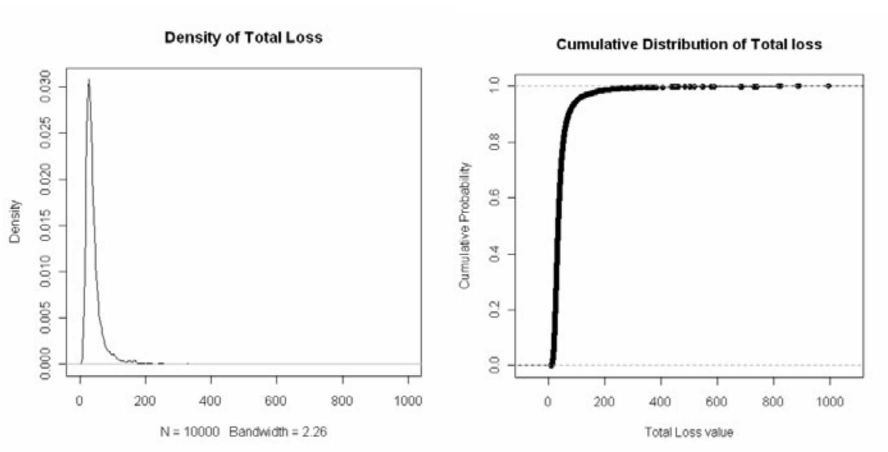


Fig. 2. Density and cumulative distribution of bank-wide losses using Monte-Carlo simulation with 10,000 scenarios

Table 9. Analysis of IT operational risk

	IT operational risk analysis Percentiles(10,000$)			
	50%	75%	95%	99%
BP1	14.45723	21.89838	41.05868	68.32410
BP2	6.926683	12.614357	66.942575	224.682532
BP3	8.135205	11.767453	19.758756	29.550277

5 Conclusion

In this paper, we attempt to develop a general, but readily adaptable model of IT operational risk which can be applied to the practice of daily operation of organizations in banking industry. This model is based on the internal loss data and binds IT events with business process and IT resources together. It is applied to the quantification of the frequency and severity distributions in IT operational risk, and uses Monte Carlo to simulate the future aggregated losses. This study has implications for banking practitioners. It provides a systematic method to bridge the IT department and business department and form an integrated model; it will benefit the banks in the long run.

There are some limits to this method. The prerequisite of the application of this model is that the internal loss data can be obtained. In addition, it requires a lot of calculation if the business processes are numerous. The development of an information system to support this method is then a task for future research.

References

1. Basel Committee on Banking Supervision. Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Model (2004)
2. Suh, B., Han, I.: The IS Risk Analysis based on a Business Model. Information & Management 41(2), 149–158 (2003)
3. Lambrigger, D., Mario, P.V., Wüthrich, V.: The Quantification of Operational Risk Using Internal Data, Relevant External Data and Expert Opinions. J. Journal of Operational Risk 2(3), 3–27 (2007)
4. Dutta, K., Perry, A.: Tale of Tails: an Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital. Working Paper of Federal Reserve Bank of Boston (2006)
5. Cheng, F., Gamarnik, D., Jengte, N., Min, W., Ramarchandran, B.: Modeling Operational Risk in Business Processes. IBM Research Report (2005)
6. Post, G.V., Diitz, D.: A Stochastic Dominance Approach to Risk Analysis of Computer Systems. MIS Quarterly 10(4), 363–375 (1986)
7. Salmela, H.: Analysing Business Losses Caused by Information Systems Risk: a Business Process Analysis Approach. Journal of Information Technology (Palgrave Macmillan) 23(3), 185–202 (2008)
8. Worrell, J.L., Bush, A. A.: Perceptions of Information Technology Risk: a Delphi Study. In: Conference of AMCIS (2007)
9. Muehlen, M.Z., Rosemann, M.: Integrating Risks in Business Process Models. In: 16th Australasian Conference on Information Systems Integrating Risks in Business Process Models (2005)
10. Shevchenko, P.V., Wüthrich, M.V.: The Structural Modeling of Operational Risk via Bayesian Inference: Combining Loss Data with Expert Opinions. J. The Journal of Operational Risk 1(3), 3–26 (2006)
11. Peters, G.W., Sisson, S.A.: Bayesian Inference, Monte Carlo Sampling and Operational Risk. Journal of Operational Risk 1(3), 27–50 (2006)
12. Halliday, S., Badenhorst, K., Solms, R.V.: A Business Approach to Effective Information Technology Risk Analysis and Management. Information Management & Computer Security 4(1), 19–31 (1996)
13. Sherer, S., Alter, S.: Information System Risks and Risk Factors: Are They Mostly about Information Systems? Communications of the AIS 14(1), 29–64 (2004)
14. Shevchenko, P.V., Wüthrich, M.V.: The Structural Modeling of Operational Risk via Bayesian Inference: Combining Loss Data with Expert Opinions. Journal of Operational Risk 1(3), 3–26 (2006)
15. Alter, S., Sherer, S.A.: A General, but Readily Adaptable Model of Information System Risk. Communications of the Association for Information Systems 14(1), 1–28 (2004)
16. Dimakos, X.K., Aas, K.: Integrated Risk Modeling. Statistical Modeling 4(4), 1–13 (2004)
17. Bai, X., Krishnan, R., Padman, R.: On Risk Management in Business Information Flow: A Process-Modeling Approach. In: Proceedings of the 5th Workshop on Secure Knowledge Management (SKM 2008), pp. 20–21 (2008)

Prediction Task Assignment of Multi-UAV Approach Based on Consensus

Chen Chen[1], Zheng Qin[1,2], and Jian-kuan Xing[2]

[1] School of Software, Tsinghua University, Beijing 100086, China
[2] Department of Computer Science and Engineering, Tsinghua University, Beijing 100084, China
mengfeicc@gmail.com

Abstract. In the battlefield, UAV (Unmanned Aerial Vehicle) needs to assign tasks dynamically by using the radar tracking information, but when the poor weather comes, or UAV enters blind area, the radar cannot gather the accurate information which severely impair the assignment result. A new task assignment prediction method was proposed in this paper. We first predict the UAV information using UKF (Unscented Kalman Filter) algorithm, and then clear up the collision by auction algorithm. Finally, we assign the tasks. The results show that the algorithm performs better than greedy task assignment and consensus based auction algorithm.

Keywords: Multi-task Assignment, Prediction, Distribution System, UAV, Tracking.

1 Introduction

UAV is widely used in modern warfare, and the task assignment methods used in multi-UAV coordination military have been the hotspot in UAV research area. The aim of task assignment methods is to make the UAV warfare be more efficient and consume less.

The methods of multi-UAV task assignment can be categorized into centralization ones and distributed ones. In the centralization method [1-3], one central UAV is responsible for assigning tasks to all the UAVs based on the global information. The best task assignment solution can be gained by this method, but it depends on the central UAV all the time, and is not flexible and robust enough; the distributed task assignment method [4-6] is paid attention because of its high-efficiency, robustness and flexibility, and it can give the assignment solution quickly because there is no need to collect all the information.

For the distributed task assignment problem, some method [7] instantiates a central machine on each UAV to increase the task range, and avoids the single point failure, but this method is based on unlimited bandwidth and globally fluent communication. A classical method is the greedy algorithm, which means each UAV chooses a task which

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 192–199, 2010.

is the nearest to attack without communication and coordination. The greedy algorithm is the best on task assignment strategy, but there is lots of resource wastage and time delay. RDTA [8] (Robust Decentralized Assignment) algorithm enables UAV to exchange the task assignment plan as well as the circumstance information which improves the robustness. Sujit [9] proposes a sequential auction method based on the market auction mechanism, which can resolve the confliction by adding a validation process before auctioning. The resource consumption is reduced by two coordination processes and it ensures the confliction can be resolved. But there is lots of time consumption and large communication load. Luc Brunet [10] proposes the CBAA (Consensus Based Auction Approach) algorithm, which both resolves the confliction and reduces the communication load during coordinated battle. But during the dynamic battle process, many factors greatly restrict the theoretically optimized algorithm in the real battle process such as the delay of radar positioning the UAV position, weather influence, blind area and so on.

This work improves the CBAA algorithm, and proposes a prediction task assignment algorithm PCAA (Predicted Consensus Auction Approach) based on Consensus. Firstly, it tracks the UAV to relocate the information of UAV, then chooses a target and resolves the confliction if exists, finally it assigns the tasks. This work introduces the formula description of task assignment problem first and second paragraphs introduce the detail of tracking, choosing target and resolving confliction process, the third paragraph compares the PCAA algorithm with the greedy algorithm and the CBAA algorithm through the experiment, I also analyze and validate the high efficiency and high performance of the PCAA algorithm.

2 Formula Description of Task Assignment Problem

The goal of task assignment is that, for the target set T and UAV set U, where $|T| = n$, $|U| = m$, to find a good method assigning the task to UAV, making the total consumption low and the effect good without confliction. No confliction means there are no more than one UAV assigned the same target. The communication radius of UAV is r_c, and the detection radius is r_s, this work assumes that $r_c = r_s$ 。State representation of UAV is introduced as below:

At time t, the state of U_i can be represent as

$$S_i(t) = \{z_i, w_i, v_i, N_i, G_i, c_i\} \tag{1}$$

Where $z_i = [x_i, y_i]$ is the current position coordinate of U_i, w_i is the current angular velocity of U_i, v_i is the current linear velocity of U_i, N_i is the UAV set in the communication rage, G_i is the target set in the detection range, c_i is the target task of U_i.

3 Prediction Task Assignment Based on Consensus

Combining the tracking algorithm [11,12] and consensus task auction algorithm, this work proposes a predicted task assignment algorithm based on consensus. The detail algorithm flow is presented in figure 1. UAV first collects the target set in the detection range, if there are targets to be attacked, then we predict the information of UAV using tracking algorithm, and choose target according to the target chosen strategy. Secondly, UAV sends the auction request to the neighbor UAVs in the communication range, finally UAV made the target assignment according to the feedback information of UAVs.

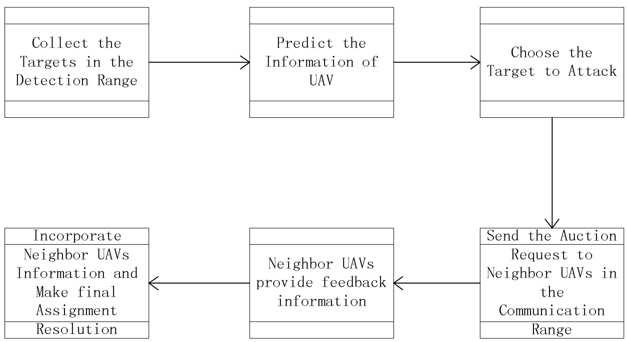


Fig. 1. Task Assignment Mechanism

3.1 Predict UAV Information

UKF (Unscented Kalman Filter) is a tracking algorithm which is mainly used to track the dynamic target on the ocean, land and air. It is a nonlinear algorithm with accurate result. Because of the strength of UKF, UKF is paid attention and is widely used in the military area.

We will introduce the prediction process in the rest of this section.

For the system below:

$$\begin{cases} x_{k+1} = f(x_k) + v_k \quad (a) \\ z_{k+1} = H_{k+1} x_{k+1} + \tau_{k+1} \, (b) \end{cases} \tag{2}$$

Where equation (a) is the linear state formula, v_k is the noise added to the state formula; equation (b) is the nonlinear state formula, τ_{k+1} is the noise.

The initial data is as follows:

$$\widehat{x_0} = E(x_0) \tag{3}$$

$$\widehat{P_0} = E((x_0 - \widehat{x_0})(x_0 - \widehat{x_0})^T) \tag{4}$$

We will introduce just one time calculation process of UKF algorithm:

1) Calculate Sigma Points

$$\chi_{k-1} = [\hat{x}_{k-1} \quad \hat{x}_{k-1} + \sqrt{(n+\lambda)\widehat{P}_{k-1}} \quad \hat{x}_{k-1} - \sqrt{(n+\lambda)\widehat{P}_{k-1}}\,] \tag{5}$$

2) Update Time

$$\chi_{k|k-1} = f(\chi_{k-1}) \tag{6}$$

$$\hat{x}_{k|k-1} = \sum_{i=0}^{2n} W_i^{(m)} \chi_{i,k|k-1} \tag{7}$$

$$\widehat{P}_{k|k-1} = \sum_{i=0}^{2n} W_i^{(c)} (\chi_{i,k|k-1} - \hat{x}_{k|k-1})(\chi_{i,k|k-1} - \hat{x}_{k|k-1})^T + Q_k \tag{8}$$

$$\hat{z}_{k|k-1} = H_k \hat{x}_{k|k-1} \tag{9}$$

3) Update Measurement

$$\widehat{P}_{zz.k} = H_k \widehat{P}_{k|k-1} H_k^T + R_k \tag{10}$$

$$\widehat{P}_{xz.k} = \widehat{P}_{k|k-1} H_k^T \tag{11}$$

$$K_k = \widehat{P}_{xz.k} \widehat{P}_{zz,k}^{-1} \tag{12}$$

$$\hat{x}_k = \hat{x}_{k|k-1} + K_k (z_k - \hat{z}_{k|k-1}) \tag{13}$$

$$\widehat{P}_k = \widehat{P}_{k|k-1} - K_k \widehat{P}_{zz,k} K_k^T \tag{14}$$

According to the iteration of above steps, the position coordinate of UAV, linear speed and angular velocity can be predicted correctly. It prepares for the task assignment below.

3.2 Task Assignment

After the prediction process, task should be chosen to be auctioned next. Each UAV will choose a target to auction asynchronously.

1)Cost Function

U_i uses Dubin's distance to decide the cost to attack target T_j^i, Dubin's distance can be calculated according to the initial position, initial angular velocity and the final position. The cost is dependent on the number of targets assigned to UAV, let $l = |S_i|$, T_l^i is the final target in the target list, then the cost to attack target T_k^i can be represented as

$$C_{ik} = \begin{cases} C_{il} + d(T_l^i, w_l^i, T_k^i) & \text{if } l \neq 0 \\ d(z_i, w_i, T_k^i) & \text{otherwise} \end{cases} \tag{15}$$

Where C_{il} is the cost of attacking target T_l^i, $d(T_l^i, w_l^i, T_k^i)$ is the Dubin's path length from the initial position T_l^i to the target T_k^i, U_i chooses target according to the length, while other neighbor UAV gives feedback and validates information according to the length.

2)Target Chosen

U_i collects all the target set in the detection range r_s, if there is no assigned task, then calculate the cost between U_i and each target in the range according to the cost function, choose the target who has the least cost as the auction target. If in the target set, there are some assigned targets, then get rid of these targets, calculate the attack cost between U_i and other targets, choose the target who has the least cost as the auction target. It can be represented as follows:

$$c_i = \max(d(U_i, T_k)) \tag{16}$$

Where target T_k is not assigned in the detection range, $d(U_i, T_k)$ is the Dubin's distance between U_i and target T_k.

3.3 Conflict Resolving

After the auction target has been chosen, here comes the confliction resolving process. Because UAVs assign targets in the limited range, maybe one target is assigned to two or more UAVs. To avoid the waste of resource, UAV needs to send the information to neighbor UAVs. When the neighbor UAVs give the feedback information, the UAV can incorporate all the information including itself, according to the confliction resolving strategy; the UAV assigns the target to the UAV who has the least cost; finally the UAV broadcasts the assignment resolution to the neighbor UAVs in the communication range.

4 Experimental Analysis

The experiment platform of this work is “UAV Battle Coordination Simulation Platform UAVSim Based on Distributed TupleSpace” based on the national defense 115 pre-research projects. The UAV battle area of the simulation circumstance is

1000*1000m. In the environment, we randomly setup 6 UAVs and 20 Targets. The turnaround radius r is 10m, the communication radius r_c = detection radius r_s =100, 200. The angular velocity w of UAV and Target, and initial position are generated randomly in each experiment. This work finishes 200 experiments using PCAA, CBAA and greedy algorithm when the communication radius r_c = detection radius r_s =100,200m. We gain the target attacked and cost consumption during the dynamic task assignment process, and compare and analyze the performance of the three algorithms.

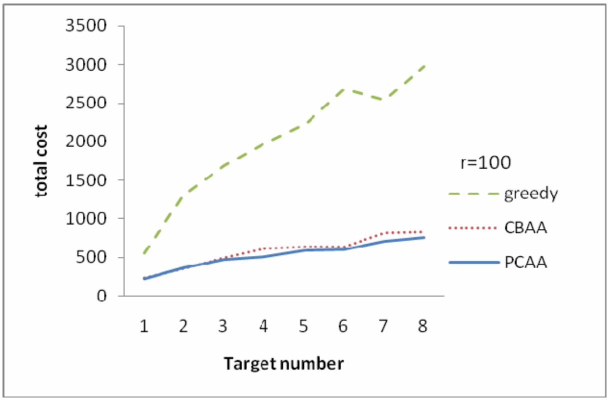


Fig. 2. Cost consumption of three algorithms when r=100

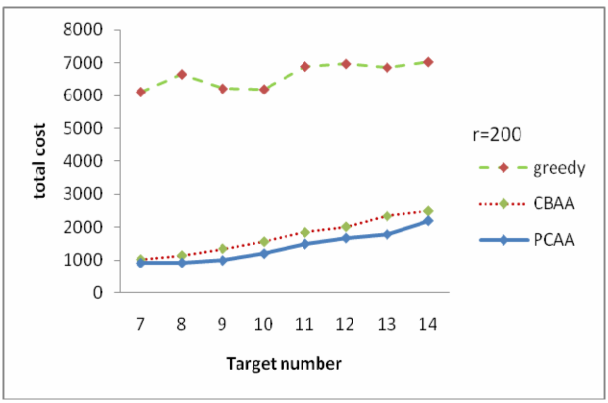


Fig. 3. Cost consumption of three algorithms when r=200

Figure 2 represents the three algorithms performance when the communication radius and detection radius is 100m, where green line represents greedy algorithm, red dotted line represents the CBAA algorithm, and blue line represents the PCAA algorithm.

When the number of targets is small, the performance of CBAA and PCAA is close, but they are better than the greedy algorithm.

The Figure 3 is the performance of the three algorithms when the communication radius and detection radius is 200m. We can see that, when the number of targets is getting lager, the performance of PCAA is much better than the CBAA algorithm, and the two algorithms are better than the greedy algorithm.

Because the mechanism of greedy algorithm is no-coordination strategy, there may be several UAVs attack one target; the total cost is very high. Both the CBAA and PCAA have the confliction resolving process, the cost is low. When there is blind area, CBAA don't have the tracking process, one target may be assigned to two or more UAVs in a period of time. So the PCAA algorithm which has the tracking algorithm is better.

5 Closing Remark

This work designs a PCAA auction algorithm distributed task assignment mechanism. It advantage is that when the radar cannot get the correct UAV information, such as strong wind and so on, it can predict the information of UAV, use the reasonable communication mechanism and resolve confliction, which make the efficiency of task auction higher, and optimize the global performance of task assignment. Comparing it with the greedy algorithm and CBAA algorithm, PCAA algorithm uses the prediction algorithm in the distributed auction problem, and improves the task assignment performance when radar cannot gain the correct UAV information.

References

1. Sujit, P.B., George, J.M., Beard, R.: Multiple UAV Task Allocation Using Particle Swarm Optimization. In: AIAA Guidance, Navigation and Control Conference, Honolulu (2008)
2. Edison, E., Shima, T.: Genetic Algorithm for Cooperative UAV Task Assignment and Path Optimization. In: AIAA Guidance, Navigation and Control Conference, Honolulu (2008)
3. Darrah, M.A., Niland, W.M., Stolarik, B.M.: Multiple UAV Dynamic Task Allocation using Mixed Integer Linear Programming in a SEAD Mission. In: AIAA Guidance, Navigation and Control Conference, Honolulu (2005)
4. Alighanbari, M., How, J.P.: Task Assignment for Unmanned Aerial Vehicles. In: Decision and Control Conference, pp. 5668–5673 (2005)
5. Dionne, D., Rabbath, C.A.: Multi-UAV Decentralized Task Allocation with Intermittent Communications: the DTC algorithm. In: Proceedings of the 2007 American Control Conference, New York (2007)
6. Sujit, P.B., Sinha, A., Ghose, D.: Multiple UAV Task Allocation Using Negotiation. In: International Conference on Autonomous Agents, Japan (2006)
7. McLain, T.W., Beard, R.W.: Coordination Variables, Coordination Functions, and Cooperative-timing Missions. Journal of Guidance, Control, and Dynamics 28(1), 150–161 (2005)
8. Alighanbari, M., How, J.P.: Decentralized Task Assignment for Unmanned Aerial Vehicles. In: Proceedings of the 44th IEEE Conference on Decision and Control, USA (2005)

9. Sujit, P.B., Beard, R.: Distributed Sequential Auctions for Multiple UAV Task Allocation. In: Proceedings of the 2007 American Control Conference, pp. 3955–3960 (2007)
10. Brunet, L., Choi, H.L., How, J.P.: Consensus-Based Auction Approaches for Decentralized Task Assignment. In: AIAA Guidance, Navigation and Control Conference, Honolulu (2008)
11. Hao, Y., Xiong, Z., Sun, F., Wang, X.: Comparison of Unscented Kalman Filters. In: Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation, China (2007)
12. Guo, R., Qin, Z., Chen, C.: An Adaptive Unscented Particle Filter for Tracking Ground Maneuvering Target. In: Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation, China (2007)

A Macroscopic Timing Optimization Model for Signalized Intersections Based on an Extended Internal State Node Model

Ximin Liu[1,2], Shoufeng Lu[2], and Shiqiang Dai[1]

[1] Shanghai Institute of Applied Mathematics and Mechanics, Shanghai University, Shanghai 200072, China
[2] Traffic and Transportation Engineering College, Changsha University of Science and Technology, Changsha 410114, China
itslusf@gmail.com

Abstract. This paper aims to construct the macroscopic timing optimization model for signalized intersections with considering traffic flow dynamics. First, Lebacque's internal state node model is extended to the situation of signalized intersections. The formulas of travel time are derived respectively for one section and multiple sections by using cumulative curve method. Then a macroscopic timing optimization model for signalized intersections is presented. Finally, some examples concerning under-saturated and over-saturated intersections are presented, which give support to the proposed model.

Keywords: Signalized intersection, macroscopic optimization model, internal state node model.

1 Introduction

In signal timing optimization formulation, traffic flow dynamics has been simply considered. For the movement of vehicles at the link, platoon dispersion phenomena have been described by statistical formulas, for example, the Roberston formula and the Pacey formula [1]. Recently, Lo [2] used the cell transmission model (CTM) to develop a cell-based signal-control model, which covers the entire fundamental diagram by encapsulating the CTM. Lo's model is a microscopic one, which needs to calculate traffic flow evolution second by second so that the optimized timing plan could be obtained. For the queue along the stop-line, a point queue paradigm was used. Until recently, the physical queue paradigm has been studied (for example, see Chow and Lo [3]). Minimum formula was derived by Lebacque [4] which is equivalent to the LWR model [5, 6]. Later on, an internal state node model for non-signalized intersections was established by Lebacque [7], which is a macroscopic model. Chevallier and Leclercq [8] developed a macroscopic model for single lane roundabouts. The paper aims to develop the macroscopic models for signalized intersections and give signal timing optimization formulas based on an extended internal state node model, so that traffic flow dynamics can be considered for signal timing optimization. For under-saturated and over-saturated intersections, some numerical examples are given to illustrate the effectiveness of the proposed model.

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 200–214, 2010.

2 Minimum Formula and Internal State Node Model

Lebacque [4] analyzed the Godunov scheme for the LWR model and obtained a minimum formula of the equilibrium supply and demand functions. For example, the local supply function is described as

$$\Sigma(x,t) = \Sigma_e(\rho(x+,t), x+) \tag{1}$$

The demand function is

$$\Delta(x,t) = \Delta_e(\rho(x-,t), x-) \tag{2}$$

In which, $\rho(x,t)$ is the density. Then the flow of q_x at point x is the minimum of supply and demand, i.e.,

$$q_x = \min[\Sigma(x,t), \Delta(x,t)] \tag{3}$$

Khoshyaran and Lebacque [9] extended the minimum formula for the link to the intersection and constructed the internal state node model for the non-signalized intersections. Assume that $N_z(t)$ is the sum of the number of vehicles and $N_{z,j}(t)$ at per exit links j inside the intersection. Total demand and total supply are the functions of total number of vehicles $N_z(t)$. The node supply is split by the constant split coefficient, $\sum_{z,i}(t) = \beta_i \sum_z(N(t))$. The node demand is split proportionally into directional demands, $\Delta_{z,j}(t) = \dfrac{N_{z,j}(t)}{N_z(t)} \Delta_z(N_z(t))$. Then node state dynamics can be described by the following formulas:

$$q_i(t) = \mathrm{Min}\left[\delta_i(t), \Sigma_{z,i}(t)\right] = \mathrm{Min}\left[\delta_i(t), \beta_i \Sigma_z\left(N_z(t)\right)\right] p_j(t) = \mathrm{Min}\left[\Delta_{z,j}(t), \sigma_j(t)\right] = \mathrm{Min}\left[\frac{N_{z,j}(t)}{N_z(t)}\Delta_z\left(N_z(t)\right), \sigma_j(t)\right] \frac{d\dot{N}_{z,j}}{dt}(t) = -p_j(t) + \sum_i \gamma_{ij}(t) q_i(t) N_z(t) = \sum_j N_{z,j}(t) \tag{4}$$

Where γ_{ij} are the assignment coefficients of fraction of users entering from i bound to j.

3 Macroscopic Model for Signal Intersection and Signal Timing Optimization Model

Introduce the following notations:

$\Sigma(t)$ the supply function of the signal intersection.

$\delta(t)$ the demand function.

$\sigma_j(t)$ the supply function at the exiting link.

$N_i(t)$ the number of vehicles at link i at time t.

q the flow rate of inflow dots.

p the flow rate of outflow dots.

3.1 Macroscopic Model for Signalized Intersection

a) Demand and supply functions at the intersection

As is illustrated in Fig. 1, the intersection connects the inflow links and the outflow links. For the inflow links, the intersection is the supply, while for the outflow links, the intersection is the demand. The internal states of the intersection are described by the green dots. For clarity, the dots connecting with the inflow links are marked as inflow dots, whereas the dots connecting with the outflow links are marked as outflow dots. As the signal is red, the supply of the inflow dots is zero, while as the signal is green, the supply of the inflow dots is the saturated flow rate, shown as

$$\Sigma(t) = \begin{cases} 0, & t \in red \\ S, & t \in green \end{cases}$$
$$q(t) = \min\left[\delta(t), \sum(t)\right] \tag{5}$$

The demand at the outflow dots is the sum of flow rate at inflow dots bound to the outflow dots, that is,

$$\Delta_j(t) = \sum_i \gamma_{ij} q_{ij}(t)$$
$$P_j(t) = \min\left[\Delta_j(t), \sigma_j(t)\right] \tag{6}$$

Where γ_{ij} is the correlation coefficient between inflow dots and outflow dots.

For outflow links, blue dots are the supply at the intersection. The supply is the sum of flow of inflow dots bound to the outflow dots. But the sum is subject to the signal phase setting. The inflow rate at green dots in the same phase can be superimposed. For example, as is shown in Fig. 2, if the south left traffic flow and east through traffic flow are in the same signal phase, then

$q_{so} = \min\{f_{si}, \Sigma_s\}$, $q_{eo} = \min\{f_{ei}, \Sigma_e\}$, $f_{wi} = q_{eo} + q_{so}$, $p_{wo} = \min\{f_{wi}, \Sigma_w\}$, in which, Σ denotes the supply; $f_{ji}(j = s, e, w)$ the demand; $q_{jo}(j = s, e)$ the flow rate of inflow dots; p_{wo} the flow rate of outflow dots.

b) The evolution of vehicle number at the inflow links

For the green period,

$$N(t+1) = N(t) + \delta(t) \cdot \Delta t - q(t) \cdot \Delta t$$
$$\text{so } \frac{N(t+1) - N(t)}{\Delta t} = \delta(t) - q(t)$$

$$\frac{dN_i(t)}{dt} = \delta_i(t) - q_i(t), t \in green \tag{7}$$

For the red period,

$$N(t+1) = N(t) + \delta(t) \cdot \Delta t$$
$$\text{so } \frac{N(t+1) - N(t)}{\Delta t} = \delta(t)$$

$$\frac{dN_i(t)}{dt} = \delta_i(t), t \in red \tag{8}$$

At each instant t, we can derive the density by dividing the number of vehicles by the link length. Then we can calculate the demand of the link through the demand function.

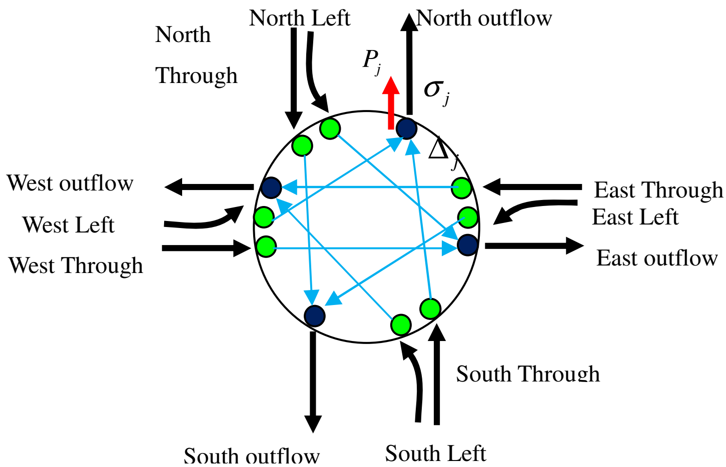


Fig. 1. The internal node relationship at a signalized intersection

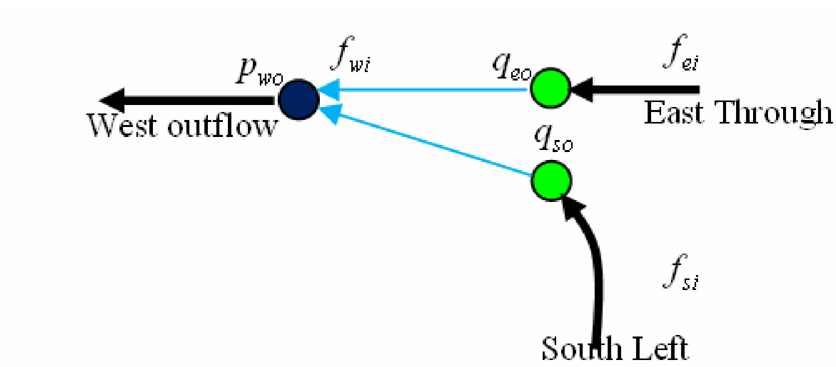


Fig. 2. Internal node relations in the same signal phase

c) The evolution of vehicle number at the outflow dots

$$N_j(t+1) = N_j(t) + \Delta_j(t) - P_j(t) \tag{9}$$

Where the variable $N_j(t+1)$ describes the delayed vehicles inside the intersection node. When the value of the variable is greater than zero, it means that its downstream link is congested. So this variable is very useful for the signal coordination.

d) The evolution of vehicle number at the outflow links

$$N_o(t+1) = N_o(t) + P_j(t) \cdot \Delta t - P_{oj}(t) \cdot \Delta t \tag{10}$$

Where $P_{oj}(t)=\min\left[\delta_{oj}(t),\sum_{down}(t)\right]$, $P_{oj}(t)$ denotes the flow rate from outflow link into the downstream intersection. $\sum_{down}(t)$ is the supply of downstream link. When there is no downstream intersection, $\sum_{down}(t)$ is set as infinity.

3.2 Macroscopic Signal Timing Optimization Models

Introduce the following notations:

A_i the arrival number of vehicles in the ith cycle;

Cy the cycle length;

R_i the residual number of vehicles in the ith cycle in inflow link;

AD_i the allowed departure number of vehicles in the ith cycle;

g the green period;

r the red period;

f_{ijr} the flow rate from section i to section j in the red period;

f_{ijg} the flow rate from section i to section j in the green period;

n_{1r} the arrival number of vehicles at section 1 in the red period;

$n_i(i=2,\cdots,m)$ the initial number of vehicles;

D_i the demand function of section i, $i=1,2,\cdots,m$;

S_i the supply function of section i, $i=1,2,\cdots,m$;

$S_{intersection}$ the saturation flow rate of the intersection in the green period;

L_i the length of section i;

N_i the dissolved number of vehicles in the ith cycle.

The cumulative curve method can employed to analyze queue and travel time between any two locations. The details about the principles of this method can refer to Gazis and Potts [10] and Newell [11]. In this paper, we use the cumulative curve method to derive the travel time for the link between the link upstream end and the downstream stop-line. Total travel time of all the vehicles in each cycle is derived by the cumulative arrival curve and departure curve. The macroscopic method describes the effect of link length on the travel time by average flow rate. If the link is longer, the density will be lower, so that average flow rate will be smaller and the departure number of vehicles will be smaller. If the link is shorter, the density will be higher, so that average flow rate will be larger and the departure number of vehicles will be larger. Therefore, from the qualitative viewpoint, the macroscopic method can describe the effect of link length on the travel time.

The objective function is used to minimize the total travel time of all vehicles. The total travel time is the sum of travel time of dissolved vehicles in each cycle, given as

$$\begin{gathered} Min\ T=\sum_{i=1}^{m} t_i \\ s.b.\begin{cases} g_{\min}\le g_1\le g_{\max} \\ g_{\min}\le g_2\le g_{\max} \end{cases} \end{gathered} \tag{11}$$

Where m is the total cycle number. At the mth cycle, all the vehicles are dissolved. The derivations of travel time t_i are illustrated in the following. The essential difference between macroscopic signal timing optimization model and Lo's cell-based traffic control model is that the former minimizes the total travel time based on cumulative arrival curve and cumulative departure curve, and the latter minimizes the delay.

The whole time period is partitioned into red time period and green time period. The total space is partitioned into some equal or unequal sections. In the red period, the vehicles move into the forward section. The vehicles queue at the section latest to the intersection. In the green period, the vehicles can pass through the section containing the stop-line. The cumulative arrival curve at the link upstream end and cumulative departure curve at the downstream stop-line for signalized intersection are illustrated in Fig.3. The cumulative departure curve is calculated by the macroscopic intersection model, which calculates the average cumulative departure flow rate.

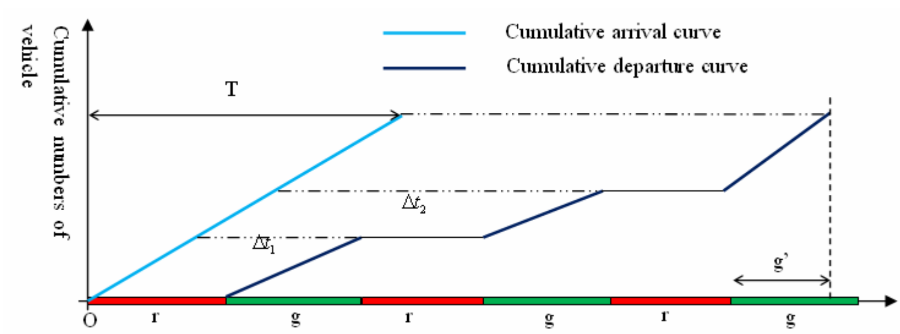


Fig. 3. The derivation of travel time formulas

3.2.1 Derivation of Travel Time Formulas for One Section

For the first cycle, the arrival number of vehicles is

$$A_1 = Cy \times \lambda \tag{12}$$

in which λ is the arrival rate. Using the macroscopic model with signalized intersection, we can calculate the departure flow rate

$$q_1 = \min\{D(A_1 / L), S_{intersection}\} \tag{13}$$

So, the allowed departure number of vehicles in the first cycle is

$$AD_1 = \int_0^g q_1(t)dt \tag{14}$$

If $AD_1 > A_1$, then the actual departure number of vehicles $N_1 = A_1$. If $AD_1 < A_1$, then $N_1 = AD_1$. The residual number of vehicles is $R_1 = A_1 - N_1$. The total travel time of all the dissolved vehicles in the first cycle is the area between cumulative arrival curve and cumulative departure curve in the first cycle. $\Delta t_1 = \frac{N_1}{\lambda}$, so

$$t_1 = (\Delta t_1 + r) \times N_1 \times 0.5 \tag{15}$$

For the second cycle, the arrival number of vehicles is

$$A_2 = Cy \times \lambda \tag{16}$$

The total number vehicles at the link is

$$C_2 = A_2 + R_1 \tag{17}$$

Applying the macroscopic intersection model, we can calculate the departure flow rate

$$q_2 = \min\{D(C_2 / L), S_{intersection}\} \tag{18}$$

So, the allowed departure number of vehicles in the second cycle is

$$AD_2 = \int_0^g q_2(t)dt \tag{19}$$

If $AD_2 > C_2$, then $N_2 = C_2$. If $AD_2 < C_2$, $N_2 = AD_2$. The residual number of vehicles is $R_2 = C_2 - N_2$. The total travel time of all the dissolved vehicles in the second cycle is the area between cumulative arrival curve and cumulative departure curve in the second cycle. $\Delta t_2 = \dfrac{N_1 + N_2}{\lambda}$, so

$$t_2 = (\Delta t_1 + r + \Delta t_2) \times N_2 \times 0.5 \tag{20}$$

For the nth cycle, if $n \times Cy < T$, then the arrival number of vehicles is

$$A_n = Cy \times \lambda \tag{21}$$

If $n \times Cy > T$, then no vehicle arrives. The total number of vehicles at the link is

$$C_n = A_n + R_{n-1} \tag{22}$$

With the macroscopic intersection model, we can calculate the departure flow rate

$$q_n = \min\{D(C_n / L), S_{intersection}\} \tag{23}$$

So, the allowed departure number of vehicles in the nth cycle is

$$AD_n = \int_0^g q_n(t)dt \tag{24}$$

If $AD_n > C_n$, then $N_n = C_n$. If $AD_n < C_n$, $N_n = AD_n$. The residual number of vehicles $R_n = C_n - N_n$. The total travel time in the nth cycle is the area between cumulative arrival curve and cumulative departure curve in the nth cycle. $\Delta t_n = \dfrac{N_1 + N_2 + \cdots N_n}{\lambda}$, so

$$t_n = (\Delta t_{n-1} + r + \Delta t_n) \times N_n \times 0.5 \tag{25}$$

For the last cycle m, the residual number of vehicles $R_m = T\lambda - \sum_{i=1}^{m-1} N_i$, in which T is the total vehicle arrival time. Using the macroscopic intersection model, we can calculate the departure flow rate

$$q_m = \min\{D(R_m / L), S_{\text{intersection}}\} \tag{26}$$

Thus we can know $g' = \frac{R_m}{q_m}$, and $\Delta t_m = (m-1) \times Cy - T + g'$.

$$t_m = (\Delta t_{m-1} + r + \Delta t_m) \times R_m \times 0.5 \tag{27}$$

3.2.2 Derivation of Travel Time Formulas for Multiple Sections

The link is spatially partitioned into m equal or unequal sections, as shown in Fig. 4.

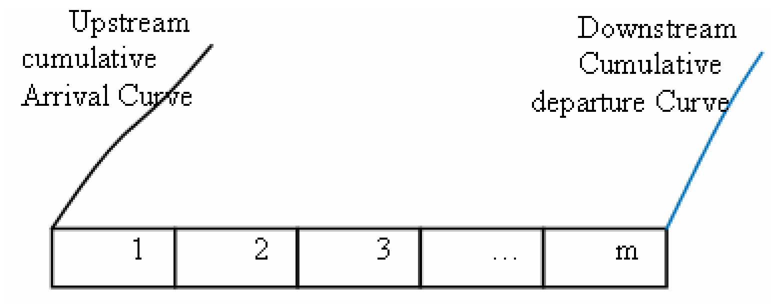


Fig. 4. Link with the multiple sections

For capturing the queue effect at the stop-line, we partition the link space into multiple sections. Set the flow rate of vehicle arrival as λ. The departure flow rate from the first section into the second section in the red period is

$$\begin{gathered} f_{12r} = \min\{D_1(n_{1r} / L_1), S_2(n_2 / L_2)\} \times r / 3600, \\ n_{1r} = r\lambda \end{gathered} \tag{28}$$

$$f_{23r} = \min\{D_2((f_{12r} + n_2) / L_2), S_3(n_3 / L_3)\} \times r / 3600 \tag{29}$$

$$f_{m-1,mr} = \min\{D_{m-1}((f_{m-2,m-1r} + n_{m-1}) / L_{m-1}), S_m(n_m / L_m)\} \times r / 3600 \tag{30}$$

Where f_{ijr} denotes the flow rate from section i to section j in the red period.
So the queue number of vehicles at the m section is $N_{rm} = f_{m-1,mr} + n_m$.
In the green period, we have

$$\begin{aligned} f_{12g} &= \min\{D_1(((n_{1r}-f_{12r})+n_{1g})/L_1), \\ &S_2((n_2+f_{12r}-f_{23r})/L_2)\}\times g/3600 \end{aligned} \tag{31}$$

$$\begin{aligned} f_{23g} &= \min\{D_2((n_2+f_{12r}-f_{23r}+f_{12g})/L_2), \\ &S_3((n_3+f_{23r}-f_{34r})/L_3)\}\times g/3600 \end{aligned} \tag{32}$$

$$\begin{aligned} f_{m-1,mg} &= \min\{D_{m-1}((n_{m-1}+f_{m-2,m-1r}-f_{m-1,mr}+f_{m-2,m-1,g})/L_{m-1}), \\ &S_m((n_m+f_{m-1,mr})/L_m)\}\times g/3600 \end{aligned} \tag{33}$$

Where f_{ijg} denotes the flow rate from section i to section j in the green period.

So the number of vehicles at the m section is $N_{gm} = f_{m-1,mr} + n_m + f_{m-1,mg}$. The outflow number of vehicles in the green period is

$$f_{mg} = \min\{D_m(N_{gm}/L_m), S_{\text{intersection}}\}\times g/3600 \tag{34}$$

If $f_{mg} < N_{gm}$, then $N_i = f_{mg}$; if $f_{mg} > N_{gm}$, then $N_i = N_{gm}$.

The travel time of all dissolved vehicles in each cycle is

$$t_1 = [r+\Delta t_1]*\frac{N_1}{2}, \tag{35}$$

where $\Delta t_1 = C - \frac{N_1}{\lambda}$

$$t_2 = [(r+\Delta t_1)+(2C-\Delta t_2)]*\frac{N_2}{2}, \tag{36}$$

where $\Delta t_2 = \frac{N_1+N_2}{\lambda}$

$$t_n = [(r+\Delta t_{n-1})+(nC-\Delta t_n)]*\frac{N_n}{2}, \tag{37}$$

where $\Delta t_n = \frac{N_1+N_2+\cdots N_n}{\lambda}$

For the last cycle,

$$t_w = [(r+\Delta t_{w-1})+\Delta t_w]*\frac{N_w}{2}, \tag{38}$$

where $\Delta t_w = (w-1)\times Cy + r + g' - N/\lambda$ and $g' = \frac{N_w}{f_{mg}}*3600$.

4 Examples

For demonstrating the performance of the proposed macroscopic signal timing optimization model, we optimize a single intersection for the under-saturated and over-saturated situations respectively, and compare the results with those obtained with the Webster method and the Gazis method respectively. The demand function and supply function are fitted by detected data. The scattering of detected data is illustrated in Fig. 5.

The demand function is known as

$$Q_D = \begin{cases} 577.8 + 29.13k - 0.067k^2 & k < 217.585 veh/km \\ 3744.1 & k > 217.585 veh/km \end{cases} \tag{39}$$

And the supply function as

$$Q_S = \begin{cases} \text{-2211.82+52.60k-0.12k}^2 & k > 217.585 veh/km \\ 3744.1 & k < 217.585 veh/km \end{cases} \tag{40}$$

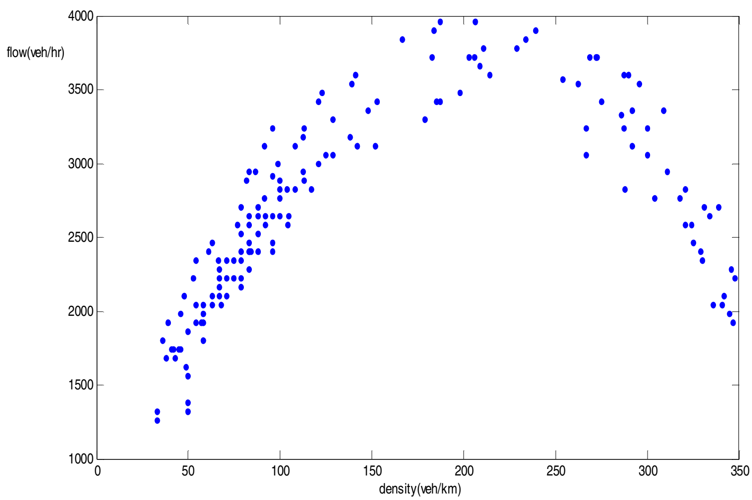


Fig. 5. Scattering of flow-density data

4.1 Signal Timing Optimization of Under-saturated Intersection

In the east-west direction, the key traffic flow rate is assumed to be 1872veh/hr, and the arrival density 79veh/km. In the south-north direction, the key traffic flow rate is assumed to be 1123veh/hr, and arrival density 43veh/km. The persistence time of arrival is both 280 seconds. The optimum cycle is calculated with the Webster method. The result is 70 seconds.

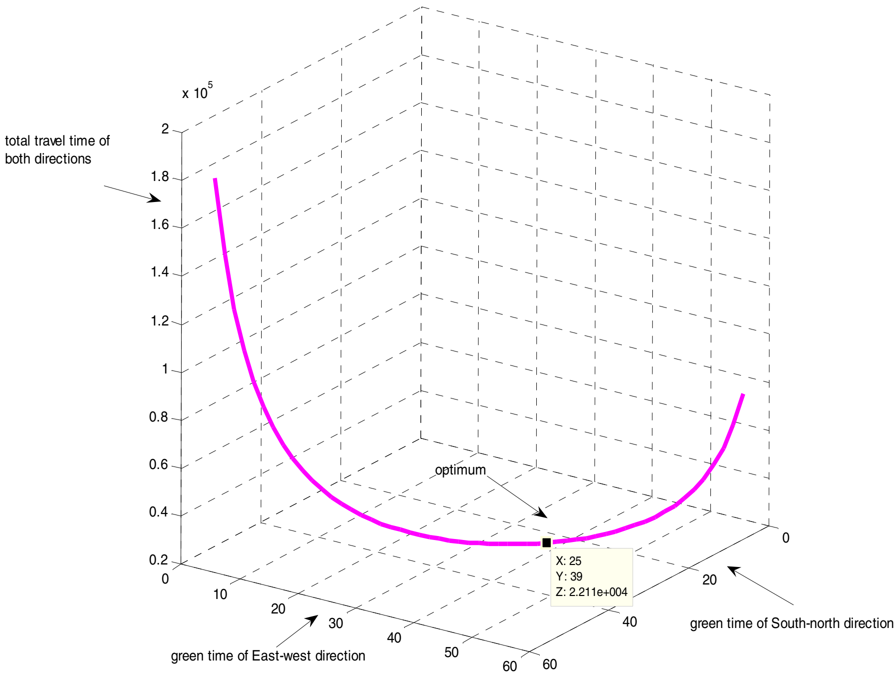


Fig. 6. Optimization results of signal timing

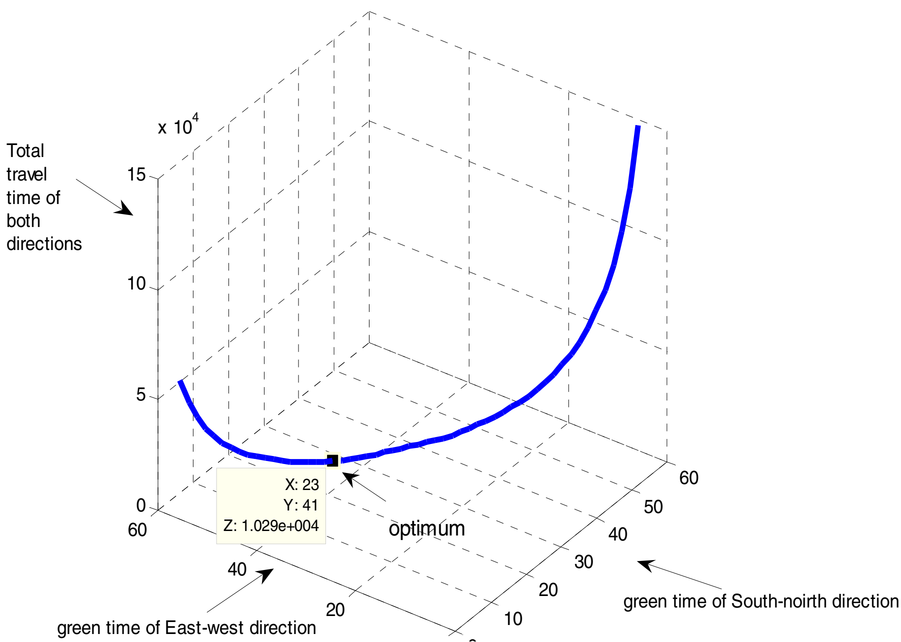


Fig. 7. Optimization results of the signal setting

Using the macroscopic signal timing optimization model with one road section, the optimized results are illustrated in Fig.6, which are respectively 39 and 25 seconds.

Using the macroscopic Signal Timing Optimization Model with two road sections, the optimized results are illustrated in Fig.7, which are respectively 41 seconds and 23 seconds.

The comparison is illustrated in Table 1. It shows that the proposed macroscopic signal intersection optimization has the slightly better performance than the Webster method.

Table 1. Comparison of signal timing plan and delay for three optimization methods

	Green period in the east-west direction(seconds)	Green period in the south-north direction(seconds)	Total delay (seconds)
Macroscopic signal timing optimization model with one section	39	25	800.7
Macroscopic signal timing optimization model with two sections	41	23	782.8
Webster method	40	24	791

4.2 Signal Timing Optimization of Over-Saturated Intersection

(1) Case of equal demand in both directions

Assume that the demand function is

$$Q_i(t) = \begin{cases} 50t & t \le 20 \\ 20t & t > 20 \end{cases} \quad \text{i=1,2} \tag{41}$$

The time unit is minute. The saturation flow rates in both directions are assumed to be 3744veh/hr. The cycle lasts 2 minutes. Total loss time is 6 seconds.

(2) Case of unequal demand in two directions

Assume that the demand functions are

$$Q_1(t) = \begin{cases} 50t & t \le 20 \\ 20t & t > 20 \end{cases} \tag{42}$$

$$Q_2(t) = \begin{cases} 40t & t \le 20 \\ 15t & t > 20 \end{cases} \tag{43}$$

The time unit is minute.

Based on the proposed macroscopic signal timing optimization model, the optimized result of equal demand in both directions is illustrated in Fig. 8, and the optimized result is illustrated in Fig.9.

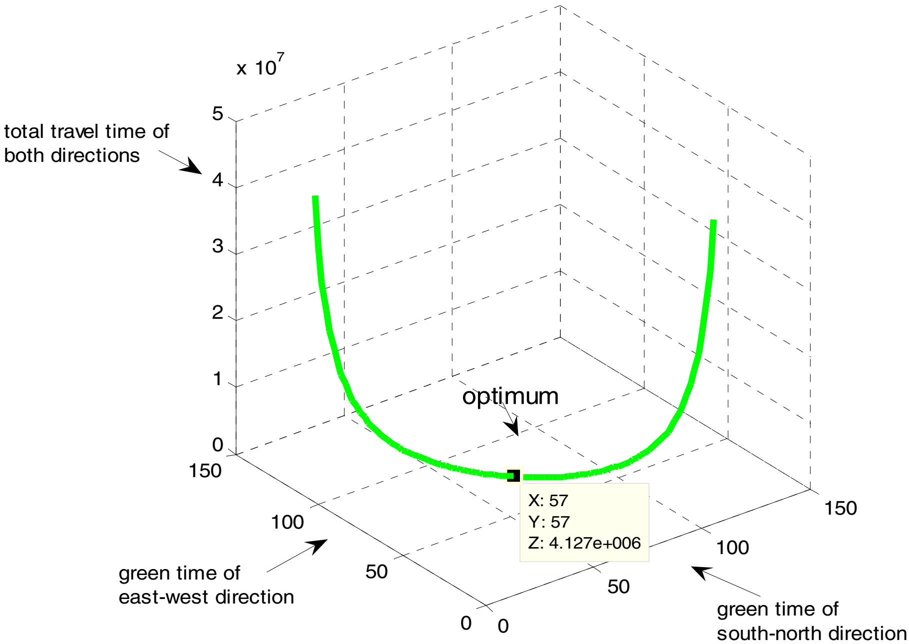


Fig. 8. Optimization results of equal demand

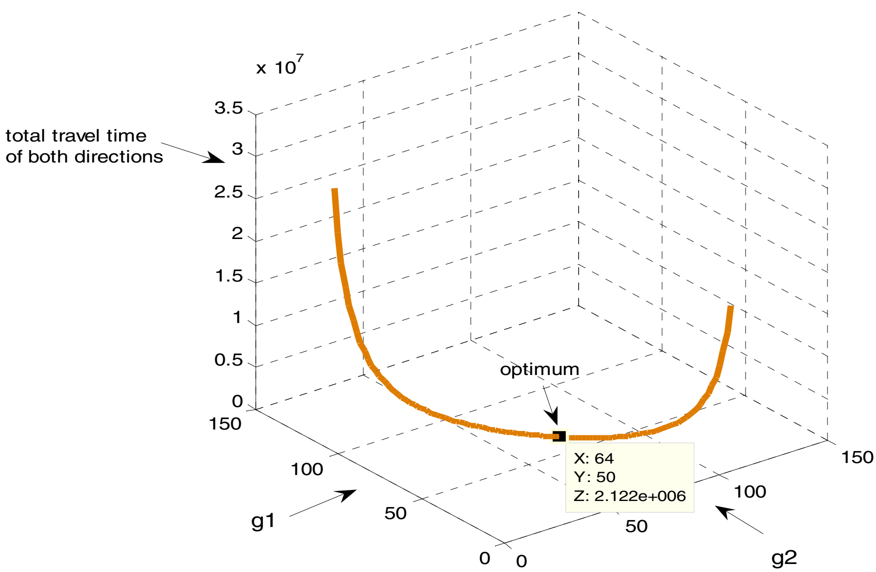


Fig. 9. Optimization results of unequal demand

Using Gazis and Potts's optimization method, the results of equal demand in both directions are $T = 62$ minutes, $g_1 = 0.95$ minutes, $g_2 = 0.95$ minutes. The results of unequal demand of both directions are $T = 45$ minutes, $g_1 = 1.07$ minutes, i.e., 64seconds, $g_2 = 0.83$ minutes, i.e., 50 seconds.

For over-saturated intersection, the proposed macroscopic signal timing optimization models have the same optimization results as those with Gazis and Potts's optimization model. In the above examples, the value of travel time is the relative one, which can be used to optimize the signal timing, but it is not the predicted real travel time.

5 Conclusions

We have extended Lebacque's internal state node model to the case of signalized intersections. The formulas for travel time are derived based on the cumulative arrival and departure curve. The extended internal state node model and travel time formulas are then integrated to resulting in a macroscopic timing optimization model for signalized intersections. Finally, some examples for under-saturated and over-saturated intersections are studied. The results show that the proposed macroscopic timing optimization model has better performance compared with the Webster model, and the same performance as Gazis and Potts's model. The advantage of the proposed model is that it is suitable for all the traffic situations from uncongested traffic to congested traffic.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 70701006, 71071024), the National Basic Research Program of China (Grant No.2006CB705500), China Postdoctoral Science Foundation Funded Project (200801195), the Key Project of Chinese Ministry of Education, the Scientific Research Fund of Hunan Provincial Education Department (09A003), and the Open Fund of the Key Laboratory of Highway Engineering (Changsha University of Science & Technology), Ministry of Education.

References

1. Quan, Y.: Urban Traffic Control. Renmin Communication Press, Beijing (1989)
2. Lo, H.: A Cell-based Traffic Control Formulation: Strategies and Benefits of Dynamic Timing Plans. Transportation Science 35(2), 148–164 (2001)
3. Chow, A., Lo, H.: Sensitivity Analysis of Signal Control with Physical Queuing: Delay Derivatives and an Application. Transportation Research Part B 41(4), 462–477 (2007)
4. Lebacque, J.P.: The Godunov Scheme and What it Means for First Order Traffic Flow Models. In: Proceedings of the 13th International Symposium on Transportation and Traffic Theory (ISTTT), pp. 647–677 (1996)

5. Lighthill, M.J., Whitham, G.B.: On Kinematic Waves. II. A theory of Traffic Flow on Long Crowded roads. Proc. Roy. Soc. London. Ser. A. 229, 317–345 (1955)
6. Richards, P.I.: Shock Waves on the Highway. Operations Research 4(1), 42–51 (1956)
7. Lebacque, J.P., Khoshyaran, M.M.: First Order Macroscopic Traffic Flow Models for Networks in the Context of Dynamic Assignment. In: Transportation Planning, Kluwer Academic Publishers, USA (2002)
8. Chevallier, E., Leclercq, L.: A Macroscopic Single-lane Roundabout Model to Account for Intersection Delays and O-D patterns. Computer-Aided Civil and Infrastructure Engineering 23(2), 104–115 (2008)
9. Khoshyaran, M.M., Lebacque, J.P.: Internal State Models for Intersections in Macroscopic Traffic Flow Models. Presented at the conference of Traffic and Granular Flow, Shanghai (2009)
10. Gazis, D.C., Potts, R.B.: The Over-saturated Intersection. In: Proceedings of the Second International Symposium on the Theory of Road Traffic Flow, O.E.C.D., France (1965)
11. Newell, G.F.: A Simplified Theory of Kinematic Waves in Highway Traffic, Part I: General Theory. Transportation Research Part B 27(4), 281–287 (1993)

A Research Methodology for Green IT Systems Based on WSR and Design Science: The Case of a Chinese Company

Yinghong Zhong* and Hongwei Liu

School of Management, Guangdong University of Technology,
Guangzhou, P.R. China
Tel.: +86-20-82568770
{Zhongyh,hwliu}@gdut.edu.cn

Abstract. Currently green IT has been a hotspot in both practice and research fields. Much progress has been made in the aspect of green technologies. However, researchers and designers could not simply build up a green IT system from technological aspect, which is normally considered as a wicked problem. This paper puts forward a research methodology for green IT systems by introducing WSR and design science. This methodology absorbs essence from soft systems methodology and action research. It considers the research, design and building of green IT systems from a systemic perspective which can be divided into as technological dimension, management dimension and human dimension. This methodology consists of 7 iterated stages. Each stage is presented and followed by a case study from a Chinese company.

Keywords: Green IT, Methodology, WSR, Design Science.

1 Introduction

The rapid development of economy in the past a few decades relied on large scale of energy consumption which brings environmental problems to the earth, such as pollution, greenhouse effect and global warming. The Information Communications Technology (ICT) industry plays a critical role in ensuring a reduction in energy consumption and greenhouse emissions. At the same time, the energy consumption of ICT, however, increases accordingly. In 2008, the energy cost of IT went up to the level of its purchase cost according to IDC. Currently ICT accounts for about 2% of total greenhouse emissions worldwide and this will increase substantially over the next 10-15 years as the adoption of ICT increases exponentially in developing countries, for example China. It is urgent to call for green information technology and green information system.

A survey of "Computer World" shows that one third of the CIOs (chief information officers) are working in progress or intending to implement green IT strategy to their organizations. A major problem is how to design and build an

* Corresponding author.

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 215–225, 2010.

effective green IT system, which is normally considered as wicked or ill-structured problems. As green IT is a new concern for many organizations, very few researchers have developed design methodologies or approaches for green IT systems. Most of the traditional methodologies are on the hypothesis of hard systems thinking, which could hardly tackle the wicked problems. Hence there is lack of a practicable research and design methodology for green IT systems.

Based on the above consideration, we propose a methodology for the research and design of green IT systems. This paper is organized as follows. In section 2, we put forward a model of green information system from the perspective of systems thinking. Section 3, which is the main body of the paper, is devoted to the introduction of a research methodology for green IT systems based on design science and the oriental systems methodology, Wuli-Shili-Renli (WSR) methodology. In section 4, a case study demonstrates the practicability of the methodology we put forward above. In section 5, we draw a conclusion and point out the future research work.

2 The Model of Green IT Systems

Currently, most researchers in the area of green IT focus on the technological aspect. The following definition of green IT proves this inclination. Green IT refers to environmentally sustainable computing of IT. It is the study and practice of designing, manufacturing, using, and disposing of computers, servers, and associated subsystems—such as monitors, printers, storage devices, and networking and communications systems—efficiently and effectively with minimal or no impact on the environment. Green IT also strives to achieve economic viability and improved

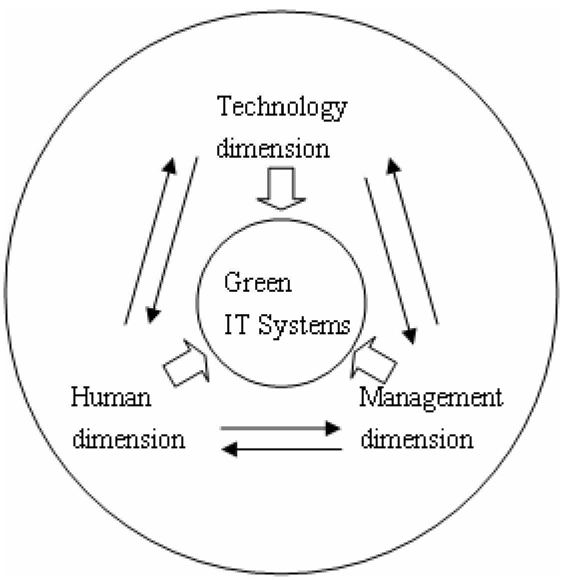


Fig. 1. The Three Aspects of Green IT Systems

system performance and use, while abiding by our social and ethical responsibilities. Thus, green IT includes the dimensions of environmental sustainability, the economics of energy efficiency, and the total cost of ownership, which includes the cost of disposal and recycling. It is the study and practice of using computing resources efficiently. The goals of green IT are similar to those of green chemistry, such as reducing the use of hazardous materials, maximizing energy efficiency during the product's lifetime, and promoting recyclability or biodegradability of defunct products and factory waste.

From the perspective of systems thinking, we argue that green IT systems should be studied, designed and built up with the consideration of three aspects or dimensions. They are technology, management and human dimensions. Figure 1 shows the three aspects of green IT systems and their relationship.

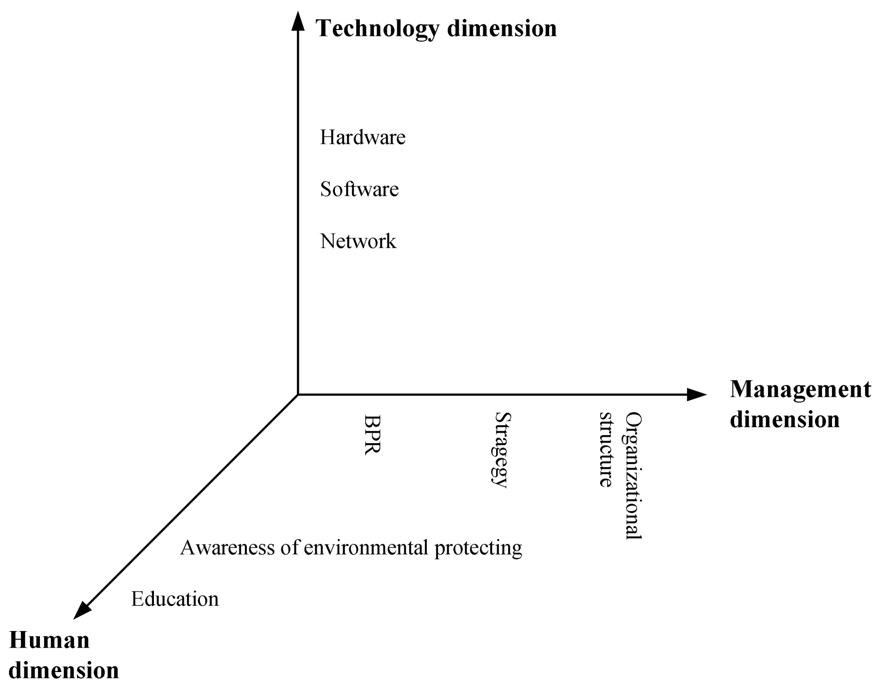


Fig. 2. The Model of Green IT Systems

The theoretical foundation of this model is Wuli-Shi-Renli (WSR), an oriental systemic approach. The Chinese term Wuli (theory of physics), Shili (theory of doing or managing) Renli (theory of humanity) are often used to reflect those laws or rules followed in practical activities as related to different situations. WSR is a conceptual framework developed from traditional Confucian teaching and contemporary management practice in China [1]. Wu in Chinese means objects, contents, 'the world' as distinct from the subject. Wus are objective in the sense that they are 'actual', existing, possessing certain properties.

Accordingly, wuli are principles and regulations, natural or man-made, that form and govern the relations among these wus. Wuli concerns the actual material regulative aspect. Wuli answers the question “what is the truth”. Shi in Chinese means (1)affairs, events; (2)troubles, problems; (3)work, business; (4)involvement, engagement, service. Shili emphasizes the virtual psycho-cognitive aspect. Shili are rules, principles and prescriptions that shape human action. Shili answers the question “how to work”. Ren in Chinese denotes human relations. Renli is concerned with the central normative question of how humans should live together, i.e., the ideals, norms and value orientations in and of social life. Renli constitute our “life world” which remains uncodified, but deeply ingrained and powerfully informing. Renli highlights the social normal-normative processes of human life. It answers the question “how man should work”.

Figure 2 shows the three-dimension model of green IT systems.

2.1 Technology Dimension

Literature review shows that most researches on green IT or information systems focus on technological dimension. In our model, technology is also an important aspect for a green design. In the level of company, green hardware and software should be considered when building an information system. Researchers and designers of information systems must take into account the following aspects.

- Data center and facilities. In 2006, US servers and data centers alone account for 1.5% of total US energy consumption. By 2011, US energy consumption by servers and data centers could nearly double again, representing$7.4 billion annual electricity cost. Up to 30% servers are “dead”, with peak and average utilization rates below 3%. In 2008, the IT facilities in China consumed 70 billion KWh, almost the annual total generating capacity of the Three Gorges Power Plant which is the largest power plant in China. Up to June 2008, there are 80 million personal computers and 2 million servers in China. Chinese enterprises show great enthusiasm for investing data centers due to the rapid economical development in China. The investment to data centers of Chinese enterprises is around 2.1 billion US dollars. For most designers of green information systems, greening data centers is the initial target. Some actions can be helpful to achieving green data centers, such as data center management, server and storage virtualization, building automation, etc. Some new technologies and products are available, for example blade server, HP Green Data Center Transformation Solution, cloud computing, IBM SAN Volume Controller (SVC) 4.3, Microsoft’s chargeback, Cisco’s TelePresence, etc. Fortunately, more and more Chinese enterprises have realized the importance of green IT and some of them have adopted green IT to their information systems.
- Distributed IT facilities. In 2008, the IT electricity use outside data center accounted for up to 55% of the total IT electricity consumption. Printers, mobile hard disks, network facilities, cooling systems, etc must be taken into account to reduce energy cost.
- Noise. Researchers and designers should develop quiet IT products and noise standards should be enacted as soon as possible.

- Pollution. The researchers and designers of IT systems should suggest the enterprises purchase those IT products that conform to RoHS, WEEE and other standards. Chinese government has also enacted a law "the Management Guideline for the Control of Electronic and IT products". All Chinese IT manufacturers must abide by this law. Some poisonous materials, such as Pb, Cd, glue, etc, are restricted to use in electronic and IT products.
- Radiation. Radiation pollution of IT systems is often ignored by researchers and designers. Radiation pollution is also a key index to evaluate IT products. The interaction among IT products may cause more radiation. Therefore, designers should pay more attention to the layout of the network and IT facilities.

2.2 Management Dimension

Reform of management can bring much contribution to a green IT system. Unfortunately, very few senior executives in China have realized the importance of it and think about green IT at strategic height. As we know, information system serves the strategy of a company. IT strategy is a part of company's business strategy. Therefore, green IT must be included in strategic management of enterprises. When CEOs make long term business strategic plans, they should ask themselves some questions. What's my budget for IT energy cost in the next 5 to 10 years? What is our responsibility for the environment? Are there any alternative green IT solutions to replace the current information system? Do we have a green index system to evaluate our suppliers and supply chain? Does the information system of our supply chain, for example ERP system, conform to green design? When planning to build a logistic center, how to design a green information system?

Business process reengineering/improvement (BPR/BPI) is a premise for green IT systems. Discussion between managers and designers of IT systems proves to be necessary to find out what the most effective business process is. In order to implement BPR/BPI, the following aspects should be taken into account, such as (1) support from senior managers; (2) customer oriented; (3) teamwork; (4) skillful employees and (5) change in organizational culture. As we have known, the basis of BPR/BPI is IT system. E-commerce and ERP systems have been widely used in Chinese companies. Only a few of them, however, have implemented BPR/BPI when building up their IT systems. Many ERP systems serve old business processes, which causes much waste in cost. Therefore, management reform is very important beyond technology itself to achieve green IT system.

Organizational structure must change according to corporate strategy and new business process.

2.3 Human Dimension

The aspect of human is also very important for a green IT system.

- Coordination and interaction. Research and design of green IT systems involves many people, such as researchers, designers, managers, employees, suppliers, customers, vendors, other stake holders, and so on. They may have different interests. To keep harmonious interpersonal relationship among them and coordinate their interests can pave the way to a green IT system. Dialogues, discussion and even debates are helpful to reach the balance point among stake holders.

- Awareness of environment protecting. To wake up the awareness of environment protecting in organizations is a hard task in China. It's a long way to go. Excessive waste of energy, paper, water, etc, can be found in many Chinese enterprises. Some employees think it's none of their business to save energy or paper. They don't shut down their computers or use power management mode when leave for lunch.
- Education. In order to wake up the awareness of environment protecting in organizations, senior managers should make green education plan or guideline and implement them. All green behaviors will be encouraged. Those employees who have green awareness can get rewards and those who don't have will receive punishment.

3 Research Methodology for Green IT Systems

There exist a large number of different approaches, methodologies and methods for the design and development of ICT-based information systems. As current management, organizational theories and context change rapidly, new approaches are needed.

For a long time, the research methodologies of information systems have been largely based on hard systems thinking, such as Hall's systems engineering and system analysis. Such methodologies, however, have encountered a series of difficulties, especially in the turbulent business environment and when dealing with wicked or ill-structured problems. It is a complex task to research, design and build up green information systems, which are normally considered as wicked problems. As we can predict, such hard systems methodologies will fail in the research and design of green information system. In order to research, design and build up good green IT systems, we need to reflect on the methodologies and make some improvement.

Recently, design science in information system (IS) cast light to IS research. IS research is largely characterized by two paradigms, namely behavioral science and design science. While behavioral science concentrates on the development and verification of theories, design science focuses on the development of solutions for practical problems and, thereby, on accomplishing utility. According to Hevner, March & Park (2004), the outcomes of a construction process under the design science paradigm are constructs, models, methods, and instantiations [2]. The development process of design science research artifacts commonly cover four phases 'problem identification', 'requirements specification', 'design', 'evaluation' and 'communication'. In order to actually achieve utility when developing an artifact, two fundamental requirements must be fulfilled: relevance and rigor. Relevance is achieved when the artifact actually addresses a real business need. Rigor, in contrast, is assured by appropriately applying the existing body of knowledge, i.e. theoretical foundations and methodologies.

Currently, more and more organizations, governments and people have realized the importance of environment protecting, which leads to great change in the context of enterprises. Entrepreneurs' thinking, organizational strategy, structure, business processes and culture, etc must be changed accordingly. Such change also influences the research and design of IT systems greatly. In this paper, we propose a methodology based on WSR and design science to address the problem of building a sound and practicable green IT system. The methodology is shown in figure 3. It has the following features.

- This methodology absorbs essence from interpretive methodologies, such as soft systems methodology (SSM) and action research. To design and build a green IT system is a complex task. Design is not a linear process in practice, but one that is confronted with uncertainty, uniqueness and conflict [3]. Designers often face wicked or ill-structured problems which could hardly be tackled with hard systems methodologies. Methodologies based on interpretive paradigm provide us with good solutions. The most widely used methodologies are SSM and action research. A widely spreaded view is that the goal of behavioral research is truth while the goal of design science is utility. We argue that truth and utility are inseparable. Truth informs design and utility informs theory. An artifact may have utility because of some as yet undiscovered truth. A theory may yet to be developed to the point where its truth can be incorporated into design. The design of IS artifact involves the interrelation with social practice in organizations. SSM is a methodology stemmed from interpretive paradigm in nature while incorporated with systems science. Just like action research, designers themselves are researchers, practitioners and actors. Just like SSM, the methodology we proposed here also emphasizes the importance of the iteration of some steps or phases.
- The methodology we proposed in figure 3 incorporates the essence of WSR. Design of green IT systems concerns with IT artifact and social practice. Traditional methodologies based on hard systems thinking could hardly tackle such problems which involve multiple aspects, such as technology, laws restriction, business strategy, organizational culture, stakeholders' interests, etc. WSR gives us inspiration to build up this methodology. Wuli, Shili and Renli are highlighted at different stages respectively.

The methodology consists of 7 stages. We explain the stages as follows.

At stage 1, executives feel some business problems arise. The designers try to understand the intention, demands or requirements, interests of customers and stakeholders. This stage emphasizes Renli. Coordination, the core of Renli, at this stage means inter-personal and inter-organizational considerations which includes emotional facet on behavioral patterns, knowledge combination, human relation and benefit negotiating. When dealing with the conflicts of different stakeholders, situations with uncertainty and social restriction, Mason and Mitroff's SAST methodology (strategic assumption surfacing testing) may be helpful to tackle some wicked problems. SAST involves 4 steps. They are (1) forming groups; (2) surfacing assumptions; (3) dialectic debates; (4) synthesis.

Stage 2 is investigation. This stage emphasizes Wuli. It includes (1) market investigation; (2) interview with those persons related, for example experts, suppliers, etc; (3) literature research; (4) technological aspects, for example technological trends, development in IT industry, etc; (5) laws and policies of governments.

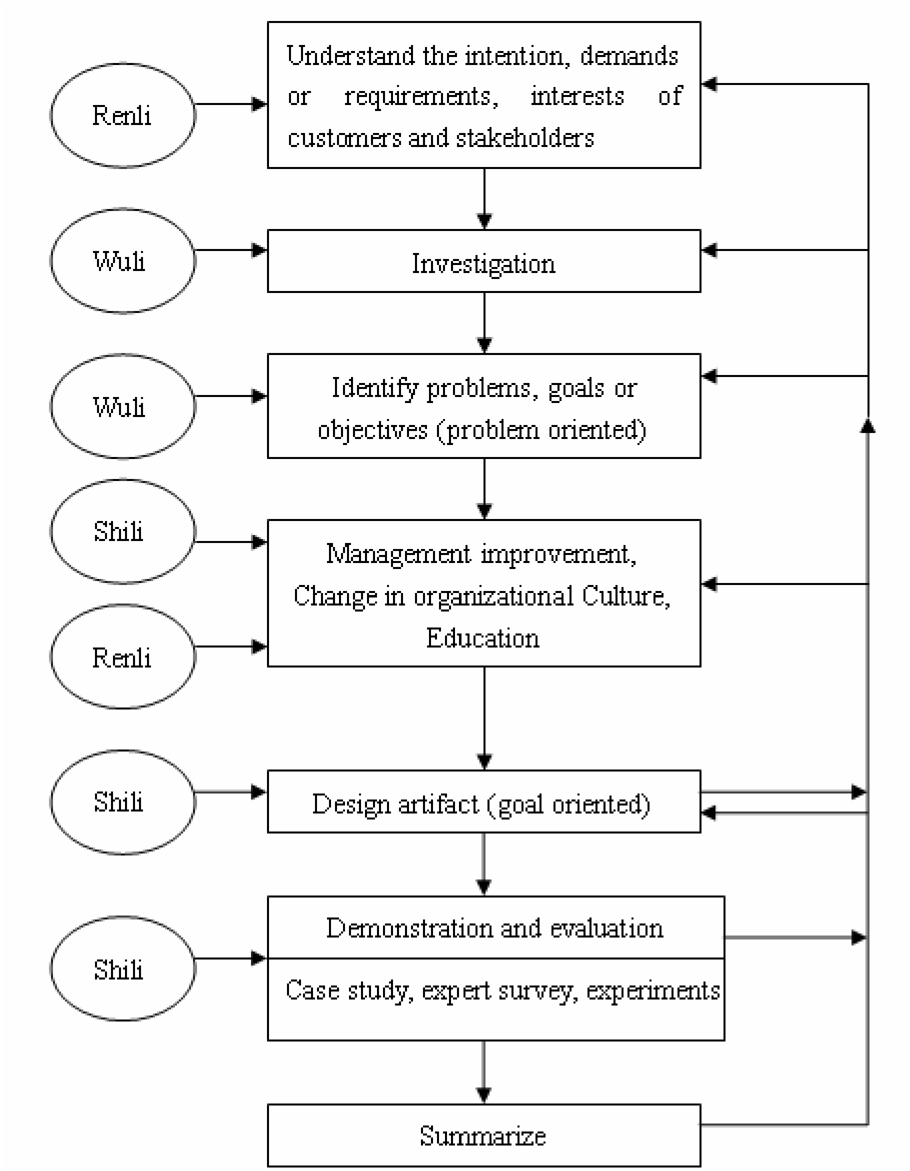


Fig. 3. Research Methodology for Green IT systems based on WSR and Design Science

Stage 3 requires designers to identify problems, goals or objectives. Designers may feel uncertain about what the problems are at first. After the investigation at stage 2, the designers can have a better understanding of the problems. The problems will be refined in the course of the current stage, to assure its practical relevance and understanding. The problems, goals or objectives should be of interest to more than one entity (e.g. company, government, stakeholders) and should be in the domain of information systems. This stage focuses on Wuli.

At stage 4, senior executives should think about some change in management, such as strategic change, BPR/BPI, reform of organizational structure, change in organizational culture, etc. This stage concentrates on Shili and Renli.

At stage 5, the artifact is designed. This stage emphasizes Shili. A solution has to be developed in the form of an artifact. At this stage, research rigor has to be ensured by using all related work available. Artifact design is a creative engineering process. Existing solutions and state-of-the-art have to be taken into account. For design research, there exist many tools and methods. The activities starting from "investigation" and "identify problems" can be iterated. This stage is goal oriented.

Stage 6 is demonstration and evaluation. Once the solution reaches a sufficient state, the designers need to demonstrate and evaluate its rigor and practical relevance. It is possible to iterate back to "design artifact" or even "identify problems" if necessary. There are some means for the goal of this stage. They are case study, action research, expert survey and experiments. Case study and/or action research should be performed on the general hypothesis. Because of the nature of action research, iterations back to "identify problems" or "design artifact" are necessary. At this stage, practical relevance and applicability is demonstrated and evaluated. As case studies and action research work on very limited samples, an additional expert survey is necessary to show that there is general interest in the solution. Besides, laboratory experiments or field experiments can be used to evaluate the refined hypothesis of the design. Statistical methods can be applied to the analysis. This stage ensures rigor. A green index system should be developed to evaluate the artifact.

Stage 7 is summarization. At the end of the process, the results should be summarized in the forms of report, papers of journal or conference, PhD thesis, etc.

4 Case Study

In 2008, we employ the methodology to research and design a green IT system for a company named GDV.

GDV is one of the largest power supply enterprises in Guangdong province of south China. With the rapid development of Chinese economy, power demands from enterprises and urban dwellers rockets up in recent years.

The information system of GDV undertakes very heavy computing tasks. It requires excellent computing groups. At the same time, the huge cost of energy could not be ignored. In the year of 2008, GDV began to rebuild its information system with the guidance of our methodology.

4.1 Understanding of Intention and Requirements

The project was launched by the senior executives of GDV and our team. The senior executives felt the problem of energy cost and planned to upgrade the information system for the future. Our team played the roles as researchers, designers and practitioners.

Based on our investigation and analysis, we identified stakeholders as follows: senior executives, managers, employees, government, power plants, enterprise power users. We employed SAST methodology to analyze their interests. After SAST, we synthesized all the assumptions and made some modification. Then we got a requirement list with consensus of all stakeholders. After that, a new IT strategy was made.

4.2 Investigation

We investigated the IT system GDV was using and interviewed several experts for suggestion. Then we studied some related laws, joint statements about low carbon economy among different countries, etc. Extensive literature review was conducted to find out what methods have already existed. We discovered that a considerable number of methods were available, but we should evaluate the quality and practicability of those methods and select some from them. Next, we survey the IT market for green technologies and products that might be used in this project.

4.3 Identifying Problems and Goals

After the investigation, we identify the goals of this project. A detailed objective list was given. The objective list consisted of energy, cost, technology, stakeholders' value, etc.

4.4 Management Improvement and Change in Organizational Culture

In accordance with the corporate strategy, the business process was improved after our investigation. Some key processes were found while some unimportant processes were deleted or merged with other processes. Education is needed to wake up the green awareness of employees. A mechanism of awards and punishment for rating employees' green behaviors was established.

4.5 Design of Green IT System

Some methods and tools were chosen to design the artifact. Before the design, we studied those methods and tools in depth and compared them to find out the strength and weakness. Besides, some good practices in method documentation were collected. In the aspect of technology, we made four improvements. (1) Storage technology, such as MAID (Massive Array of Idle Disk), thin provisioning technology, and data re-dupilcation. (2) Server improvement. Such as selecting the processors with low voltages, reduction of the energy cost of CPUs, automatic adjustment of the voltages of CPUs, and dynamically managing the servers. (3) Virtualization technology. Microsoft Virtual server technology was applied. (4) Re-design of the data center system. The layout and cooling system were improved.

4.6 Demonstration and Evaluation

In this project, we played the roles as researchers, designers and actors and action research methodology was employed. After the design was over, an expert survey was used to evaluate this newly-built green IT system. After running for one year, the performance of the green IT system proves to reduce 13% of energy cost than before. Management efficiency has been improved. Managers and most employees have the good habits to save energy and computer consumable materials, by which means saved 1.6 million US dollars per year.

5 Conclusion and Future Research Work

In this paper, a research methodology for green IT system was presented. It is divided into 7 stages that one can refer to. This methodology introduces WSR and design science into the research of green IT systems. This methodology brings systems thinking to green IT researchers by means of the combination of Wuli, Shili and Renli to the whole process. It overcomes the shortcoming of those methodologies that only focus on technological aspect. By absorbing the essence of design science, this methodology also overcomes the shortcoming of those hard systems methodologies that could hardly deal with wicked problems, such as the establishment of green IT systems. Theoretical foundation and meaning of all stages of this methodology were introduced. And then case study to illustrate the usage of the methodology was described.

In the future, we think that it's necessary to build up tool box or a collection of methods for each stage of the methodology we proposed. We are planning to apply this methodology to more projects of green IT systems. Some modifications will be made to this methodology if necessary.

Acknowledgments

This research is supported by the Natural Science Foundation of China (NSFC), NO. 70971027.

References

1. Zhu, Z.C.: WSR An Institutional Approach to Organization Studies. In: Proceedings of the 4th International Conference on Knowledge and Systems Sciences, Guangzhou, China (2003)
2. Hevner, A.R., March, S.T., Park, J.: Design Research in Information Systems Research. MIS Quarterly 28(1), 75–105 (2004)
3. Rohde, M., Stevens, G., Broedner, P., Wulf, V.: Towards a Paradigmatic Shift in IS: Designing for Social Practice. In: The 4th International Conference on Design Science Research in Information Systems and Technology, USA (2009)

Research on the Hotel Image Based on the Detail Service

Ban Li, Zheng Shenghua, and Yi He

College of Business Administration, Zhejiang University of Technology, China, 310023
liban@zstu.edu.cn

Abstract. Detail service management, initially developed as marketing programs to enhance customer loyalty, has now become an important part of customer relation strategy. This paper analyzes the critical factors of detail service and its influence on the hotel image. We establish the theoretical model of influencing factors on hotel image and propose corresponding hypotheses. We use applying statistical method to test and verify the above-mentioned hypotheses. This paper provides a foundation for further study of detail service design and planning issues.

Keywords: Detail Service, Hotel Image, Influencing Factor.

1 Introduction

Detail service is a kind of service philosophy and a service process which focuses on the needs of customers, and aims to utilize resources either inside or outside the hotel to a maximum extent by planning, implementing, controlling and innovating service in the hotel. The essence of detail service is to make service in hotel personalized, human-oriented and innovative, and create high-quality at low-cost. Furthermore, detail service is aimed at setting up a good image for a hotel.

In recent years, some scholars have achievements in the area of detail service and hotel image. Wang Bin [8] holds that detail service must be diversified from hotel to hotel, because different hotels have different kinds of customers, which directly determine that detail service must be provided in different forms. Zeng Qiang [9] points out that safety is the most essential need among hotel customers, therefore, hotels must enforce the security service-concerned management. Mu Jihong [10] refers that detail service brings remarkable effect in hotels.

However, there are also many deficiencies in this field, especially the rudimental composition of detail service and its intrinsic mechanism of promoting hotel image cannot be substantially unveiled. In order to understand the essence of detail service and its influence on hotel image, this paper establishes and tests a clear theoretical model from the viewpoint of "the influencing factors of hotel image based on detail service".

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 226–233, 2010.

2 Theoretical Background and Hypothesis Develop

2.1 Variable Definition

2.1.1 Hotel Image

Hotel image is a public emotional inclination towards the hotel. Zeithmal [7] points out that purchase is the behavior of customers based on their subjective cognition of value, so if this is applied to hotel business, it means that reputation of a hotel and satisfaction of its customers lead to the behavior of purchase. Sweat [12] thinks that measuring indexes of the purchase behavior includes: the reputation of purchased products, the purchased amount and the cost and so on. By combining those researches, this paper brings out three indexes for the measurement of hotel image: reputation of the hotel, satisfaction towards the hotel and loyalty of its customers.

2.1.2 Detail Service

Started with the image of service among hotel image, this paper chooses one item from three main influencing factors on hotel image as independent variables in the research. They are detailed service, detailed facilities and detailed products. This paper holds that detailed service is a kind of intangible service that focus on the personalized needs of customers. It can be classified into four types: entrusted service, personal inclination service, emergency service and flexible service. Detailed facilities mean the hardware equipments supplied by hotels are aimed to satisfy the various needs of customers, including needs for social activities, entertainments, approving and achieving of respect and self-identity. Detailed facilities also focus on the importance of human-oriented facilities in hotels. It can be divided as: lobby detailed facilities, room detailed facilities, canteen detailed facilities and security-concerned detailed facilities. Detailed products are tangible material objects provided by hotels in order to keep a balance between the total customer value and the total customer cost and to improve the satisfaction among customers. Detailed products lay emphasis on innovation. It can be divided as follows: optional products, products with complete sets service, random products and diversified products.

2.2 Hypothesis

2.2.1 The Relation between Detailed Service and Hotel Image

Entrust service. Entrust service is a very important part among hotel service. Entrust service such as ticket agent service for customers will play a positive role in setting up a good hotel image.

Personal inclination service. Zhuang Zhimin [13] proposes that different customers have different inclinations, so hotel can satisfied their customers only by fulfilling their personal inclinations timely.

Emergency service. It means that hotels should provide their customers with some useful service at occasions when they loss some important personal possessions, or alike. Emergency service is the best way for hotels to improve the feeling of loyalty of their customers.

Flexible service. It is service that changes by the different needs of customers consistently, for instance, no charge will be added to a customer's bill if he had asked for delaying checking out. It is also designed to improve the loyalty of customers. Based on the above expatiation, this paper sets the following hypotheses:

H1: Detailed service has a positive effect on hotel image.

H11: Entrust service has a positive effect on hotel image.

H12: Personal inclination service has a positive effect on hotel image.

H13: Emergency service has a positive effect on hotel image.

H14: Flexible service has a positive effect on hotel image.

2.2.2 The Relation between Detailed Facilities and Hotel Image

Lobby detailed facilities. Yao Le [14] supports that detail service should be carried out in a perceptible way so that customers can feel it. In the process of producing a harmonious environment in the lobby, the lighting should be adjusted to be softer at the first place, and other lights of different luminance and in different angles must be applied so as to prevent customers from visual fatigue caused by too shinning light. Secondly, attention should be paid to the volume and type of background music to make sure that it is controlled in a most comfortable range for customers.

Room detailed facilities. Mu Jihong [10] thinks that there are still many hotels in China that didn't pay much attention to details. Regulations like customers can only dry their clothes in the bath room and deeds such as no changes are made to the beddings throughout the year still exist in most hotels. If room detailed facilities are intensified, a feeling of home will be created for customers.

Canteen detailed facilities. A reinforced management in the hotel canteen service will effectively improve satisfaction among customers. Actions such as making a good control of the temperature in canteen, providing comfortable rest places for customers are all included in this area.

Security-concerned detailed facilities. Zeng Qiang [9] once claimed that security is the utmost need among customers of a hotel, so hotels need to strengthen their security management. In respect to detailed facilities, this paper sets the following hypotheses:

H2: Detailed facilities service has a positive effect on hotel image.

H21: Lobby detailed facilities has a positive effect on hotel image.

H22: Room detailed facilities has a positive effect on hotel image.

H23: Canteen detailed facilities has a positive effect on hotel image.

H24: Security-concerned detailed facilities have a positive effect on hotel image.

2.2.3 The Relation between Detailed Products and Hotel Image

Optional products. The development trend of the hotel industry determines that detailed products will be more common in hotels, because customers will gradually become much more eager to choose products themselves in a hotel, such as facilities in their rooms, according to their own personalities.

Products with complete sets service. Apart from the accommodation service in hotels, another important function of hotel is to hold all kinds of celebrations and conferences. Thus, hotels will gain reputation and trust from their customers if they can excel at such kind of complete sets service.

Random products. Random products service is designed to satisfied random needs of customers by providing them relevant products, so it is unpredictable. For example, a waiter will take the initiative to service a drunken customer a cup of green tea. Generally, random products are also effective for hotel to improve their service and intensify the brand effectiveness.

Diversified products. Customers in a hotel might be very different from each other in many aspects, such as their gender, nationalities, religion or personal inclinations. Take the Moslem customers as an example, hotels should never provide them with food cooked with pork. Since this product varies to different customers, hotels call it diversified products, which is a main measurement for hotels to upgrade their core competence. To this part, hypotheses are given out as follows:

H3: Detailed products have a positive effect on hotel image.

H31: Optional products have a positive effect on hotel image.

H32: Products with complete sets service have a positive effect on hotel image.

H33: Random products have a positive effect on hotel image.

H34: Diversified products have a positive effect on hotel image.

2.3 Construction of the Model

Based on above hypothesis, we establish the hypothetical model (Figure 1).

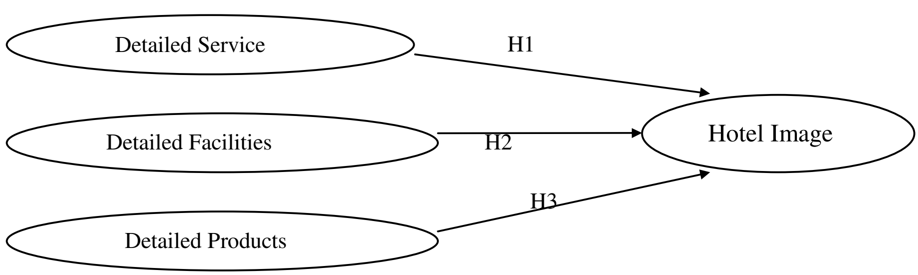


Fig. 1. Hypothetical Model of Influencing Factors of Hotel Image Based on Detail Service

2.4 Sample

Through the empirical and theoretical studies, we established a hypothetical model. Then we collect data by using questionnaire, we adopt factor and correlate analysis to validate the hypothesis by SPSS12.0 statistic software.

2.4.1 Sample Data

The sample data is collected through questionnaire. The questionnaire is designed based on interviews of multiple hotels in Zhejiang provinces and reference of past studies. In order to get more accurate data, we have surveyed customers and hotel managers at several hotels in Zhejiang province, eliminated questions with low responses or ambiguity and redesigned the questionnaire with improvement. It is designed using Likert's five points graph with "1" representing total disagreement and "5" representing total agreement. The number of total questionnaires distributed is 150 with 80.67% return rate (121 returned).Twenty of the returned questionnaires are not used in this study due to incompleteness.

Table 1. Questionnaire component list

Potential Variables	Sources
Detailed Service	Mu Jihong（2006）[10]
Detail facilities	Wang Dawu（2006）[2]
Detail Products	Yao Le（2003）[14]
The Reputation of Hotels	Boulding et al.(1993) [1]
The Loyalty of Customers	Sweat (2000) [12]
The Satisfaction of Customers	Xu Huimei（2000）[4]

2.4.2 Factor Analysis

Before factor analysis, the influencing factors of hotel image referred in this paper are inspected by Kaiser-Meyer-Olkin Measure of Sampling Adequacy and Bartlett test of Sphericity, as shown in Table 2.

Table 2. KMO and Bartlett's Test

KMO Measure of Sampling Adequacy.		.789
Bartlett's Test of Sphericity	Approx.Chi-Squ are	2750.354
	Sig.	.000

After the factor analysis, 15 influencing factors in detail service of hotel image are classified into three categories: detailed service factors, detailed facilities factors and detailed products factors. Among detailed service factors, entrust service（0.802）and flexible service (0.701)serve as the two most important factors. While among detailed facilities factors, room detailed facilities service (0.755) takes the most important place. And among detailed products service factors, the most important two are diversified products (0.737) and products with complete sets (0.722).

2.4.3 Correlate Analysis

In order to further the understanding of the relationship between dependent variable and independent variables, this paper carries out the correlate analysis on the dependent variable and independent variables, as shown in Table 4.

Table 3. Factor Analysis Result

Variable	Brief Description of questions	Factor F1	F2	F3	F4	Cronbach alph
Hotel Image	Knowing the hotel before check-in	.767				0.877
	Good reputation of the hotel	.627				
	Willing to check-in the same hotel	.579				
	Willing to recommend the hotel to friends	.545				
	Satisfied with service in the hotel	.512				
Detailed Service	Capable to provide buy-on-customers'-behalf service		.802			0.818
	Capable to solve questions from customers		.683			
	Capable to provide individualistic service		.582			
	Capable to provide service in time		.609			
	Capable to provide flexible service		.701			
Detailed Facilities	Soft, sufficient lighting and comfortable environment in the lobby			.635		0.778
	Available facilities including Shoe Polish Machine and Vending Machine for customers to use during their waiting for service in the lobby			.632		
	Capable to provide self-service furniture including dryer, electronic iron and microwave oven			.755		
	Pleasant background music in canteen			.588		
	Safeguards on patrol all-day			.679		
Detailed Products	Capable to provide personified food				.737	0.746
	Capable to adjust equipments according to needs of customers				.661	
	Excellent in holding activities including conferences and celebrations				.722	
	Capable to provide random products				.562	
	Capable to diversify products in accordance the change of customers				.609	
	Total Amount of Factors	5	5	5	5	

Table 4. The Correlation Coefficients of Detail Service and Hotel Image

Independent Variables	Relevant Coefficient	
Detailed Service Factors	Pearson Correlation	.520**
	Sig.(2-tailed)	.000
	N	101
Detailed Facilities Factors	Pearson Correlation	.487**
	Sig.(2-tailed)	.000
	N	101
Detailed Products Factors	Pearson Correlation	.379**
	Sig.(2-tailed)	.000
	N	101

Note:**Correlation is significant at the 0.01 level (2-tailed).

From Table 4, we can see that the correlation coefficients of each influencing factors of hotel image are significant at 0.01 level. It means that all the influencing factors have a positive effect on hotel image. As fig. 2, the following conclusions can be made: firstly, the influencing factors of hotel image based upon detail service fall into three major categories: detailed service, detailed facilities and detailed products. Secondly, detailed service has the greatest positive effect on hotel image, with the correlation coefficient 0.520. Thirdly, detailed facilities come the second in term of its positive effect on hotel image and its relevant coefficient is 0.487. Lastly, detailed products service has comparatively less effect on hotel image, with its relevant coefficient 0.379.

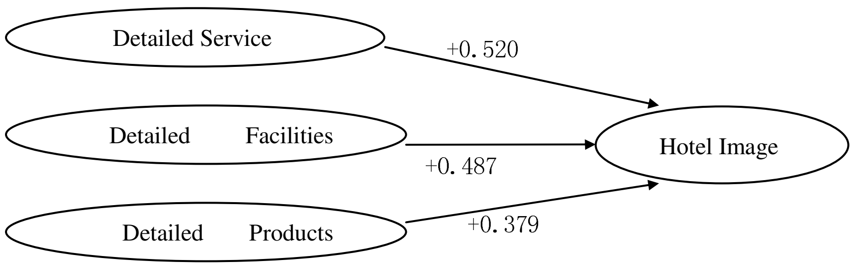


Fig. 2. The Model of Influencing Factors of Hotel Image based on detail service

3 Conclusions

Starting with the concept of detail service, this paper analyzes the critical factors of detail service and its influence on the hotel image. It establishes the theoretical model of influencing factors on hotel image and proposes corresponding hypotheses. And then this paper proves the above-mentioned hypotheses by applying statistical method. The empirical results agree with the paper's research design and hypotheses. Based on the results, this paper suggests hotels can effectively enlarge their service scale, improve their service quality and cultivate the royalty and satisfaction of customers by providing detail service.

Under current economic and situation, it is important to continue the study of detail service. This study offers a preliminary discussion on this topic and there are still areas which need further discussions: firstly, enlarge the volume of subjects in the experiment, so as to make the experiment cover more regions and the research results more universal. Second, the study on relations between demographic factors and hotel image can be furthered. Thirdly, the effect of detail service on hotel image can be analyzed on the basis of various perspectives.

References

1. Zhou, Y., Zheng, H.: The Crucial Standards to Value Hotel Service. China Travel 8(27) (2003)
2. Xue, Q., He, P.: Refined Service–36 Key Details for Successful Service. Beijing Scientific Publishing Company, Beijing (2006)

3. Wu, C.: A Research on International Travel Consumption in Taipei. Business Management Research Center of National Sun Yat-Sen University, Taiwan (1990)
4. Xu, H.: Evaluation on Images of Large International Tour Hotels by Tourism Experts. Business Research 2 (2000)
5. Li, L., Deng, Y., Li, M.: Researches on Satisfaction of Tourists Based on Market Segmentation-A Case Study on Dinggen Area. Tourism Management Research Center of National Kaohsiung University of Applied Sciences, Taiwan (2006)
6. Liu, W.: Detail Service-the Essential Way to Reestablish the Competence among Chinese Logistic Companies in the Age of Tiny Profit. Journal of Shanghai Maritime University 27(4), 1–8 (2006)
7. Payne, A.: Service Marketing. Zhong Xin Publishing Company, Beijing (1998)
8. Wang, B.: An Analysis on Personalized Service in Hotels. Journal of Hubei Commercial College 14(3) (2002)
9. Zeng, Q.: Innovation of Service Based on Clients Value. Zhejiang Gong Sang University, Hangzhou (2004)
10. Mu, J.: Detail Is the Magic Weapon to Success in Hotel Business. Market Recommendation 6 (2006)
11. Zaithmal, V.A.: Consumer Perception of Price, Quality and Value: A Means-end Model and Synthesis of Evidence. Journal of Markering 52(2), 2–22 (1988)
12. Jarvenpaa, S., Todd, P.: Consumer Reactions to Electronic Shopping on the World Wide Web. International Journal of Electronic Commerce 1(2), 59–88 (2000)
13. Zhuang, Z.: Eight Tendencies in Development of Tourism Culture. Nankai Business Review 1, 64–66 (1975)
14. Yao, L.: Analysis on the Specialization and Extension of Supermarket Service Connotation. Business Economics and Administration 12, 26–28 (2003)
15. Ma, Q.: Statistic and Management. Science Publishing Company, Beijing (2004)

Impact of Network Information on the Management of Library E-Resources

Zhixiang Yin and Hua Chen

School of Science, AnHui University of Science and Technology
zxyin66@163.com

Abstract. The Internet-based electronic resources (e-sources) are growing dramatically, however, there have not been enough studies evaluating the impact of e-sources in library source management. This study tries to investigate how much e-sources have been used in the area of library and information science (LIS) during the period 2001 to 2008. Characteristics of network information resources management are proposed and discussed in this study for the impact measurement. This paper develops a framework for categorizing the main management issue of e-resources in library.

Keywords: Network Information Resources Management, Internet, Library E-resources.

1 Introduction

Network provides a brand-new information resources environment, also brings about an unprecedented impact on the information resource management and pours into many new concepts and the means [1]. The network information resources refer to the digital recorded, multimedia expressed, magnetic, optical or other communicative medium stored, the computer network transferred information sets. Compared with the traditional information resources, the network information resources differ remarkably in quantity, structure, distribution, dissemination scope, type, carrier shape, connotation, control mechanism, transformation method and assumes many new characteristics.

2 Characteristics of Network Information E-Resources Management

From information resource viewpoint, Internet is undoubtedly an integrated data resource network of which information is collected from different departments and domains in order to be shared by on-line users [2]. On-line resources may be classified from the different aspects, however, no matter how these resources are to be classified, they severs to meet the need of better understanding, organization, retrieval, use and in short, management these resources. Therefore, understanding characteristics of network information resource and subsequent problems imposed on information management is perhaps more significant [3]. Characteristics of network information resources management can be described briefly as follows:

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 234–237, 2010.

1. Network information resources have large quantities, multimedia, non-standard, cross time, region and profession, multilingual and so on. Meanwhile, it also contains text, data, graph, audio and video information. The complexity and variety of information resources management increases unprecedently.

2. Internet's value comes from user's knowledge and the creativity, to a considerable degree, the growth of network and the fast dynamic delivery of information resources are directly driven by users. However, information security and the quality are far from satisfying due to lack of effective and consistent management mechanism.

3. Distribution and the constitution of information lack of proper structure and well organization, information source is scattered or disordered. Moreover, its update and die-out cannot be foreseen which burden information resources management.

4. The information is delivered randomly and arbitrarily. Private information enters the public information dissemination channel. Lack of necessary filter, the quality control and the management mechanism, academic, commercial, government and individual information mix together. Moreover, massive racial discrimination, the decadent information also gains the chance to disseminate which has aroused many problems.

5. The official publication and the unofficial information exchange interact, cause traditional the human information exchange chain pattern to break. All parts on the network may be the information producers, deliveries or may be disseminators or users, which has had the profound influence on the academic exchange environment.

6. The information exchange has gone beyond national boundary and border, promoted the sharing of mankind information resources and brought some unexpected problems, for instance, culture conflict, information aggression, information deterrent and so on. This certainly set the new request to humanity aspect of the information management.

Viewed from the information resource management perspective, the network information management has the following new characteristics in comparison to traditional information resource:

1. As far as information expression is concerned, traditional information resources are tend to be primarily of text or digital form, while network information resources are non-text or non-digital form instead of an integrated form of audio, video and text. Therefore, they differ in information processing, storing, collecting and transformation methods.

2. Their life cycles are also different from each other. Generally speaking, the accessible information on Internet is mainly of recent years. Although people started to devote to digitizing and making available on the network the scientific knowledge, the human civilization which has been accumulated for thousands of years, the completed transformation was just a tip on an iceberg thanks to limited material, manpower and financial resources. To these days, a vast majority of knowledge has been preserved in all kinds of information centers or libraries. They were undergone special process and treatment and has formed into system structure and could be accessed easily for three, five years even thousand years. But in network background, information resource undergoes nearly fast changes, sometimes even is “morning deliver evening alter”. Apparently, planning, processing, accessing, retrieving,

delivering and serving of information resource of this kind should be more reasonable and strict.

3. The carrier is different. At present the computer can accept the information carrier which may be summarized as five kinds: paper, microfilm, magnetic tape, compact disc and hard disk.

3 Impact of the Management of Library E-Resources

First, virtual information resource in network setting has greatly enriched the library collection, has enlarged the library resources space [4]. In an era of global networking and information, none of libraries can fully meet readers' need. But the network information resources have collected world wide information related to certain topic of all works of life. As a supporter, such rich, highly open and free, abundant resources information library, this brings hopes as well as challenges to the traditional library. Therefore, it fills in the library vacancy on one side and greatly reduced the library function and the status on the other side. Under this stern situation, will the library bravely stand up to meet the challenges or automatically flinch, or even disappear? As a disseminator of knowledge, library will undoubtedly last for ever since computer cannot substitute for book, book cannot substitute for language. Network information resource will not substitute for library for ever. Second, confronted with such abundant information resources, librarians are not manual laborer any more, they are equipped with computer and they can make fully use of it to collect, sort information resources of various types. Most important of all, knowledge structure of the librarian has changed. In the call of information and networking, it requests the librarian not only to furnish himself with specialized knowledge of book information but more importantly equip with modern high tech knowledge of electronic computer [5]. This imposes great pressure upon library, which determines utility efficiency as well as the future prosperity or failure of library. In view of the fact that abundance and dispersivity of information resources on Internet, some network resource research tools via World Wide Web were born. At present, there are many domestic and foreign search engines, but until now none of them are able to include all retrieval functions. This is mainly because these World Wide Web search tools are designed by computer and the network experts, thus have the certain limitations. After period of time of explorations and the practice, the experts reach a consensus that on-line information resource should be scientifically managed using "librarian's though". The effectiveness of network information resource for a period of time guarantee dynamic characteristic of on-line information. Not only kinds of information but the network, the website, the homepage that it links also undergoes constant change. In order to guarantee this effectiveness, many enterprises have downloaded related information resource to perform scientific organization and reorganization, set up the different databases according to different topics. At the same time, the contents of databases are also in unceasingly refreshment and the supplement along with the network information change, for example: technical report database, dissertation database, government document database, and so on.

Besides building up databases, another method that may lengthen the network information effectiveness for a period of time is to save text, graph and audio information in electronic periodic form and provide it to users after organizing and sorting using hypertext, hypermedia techniques. The most attractive features of resources shared network are the open access and shared resources. The orderly management of information resource is for better and fully resources sharing which is based on collaborative building and mutual benefit. At present, many libraries have resource sharing consciousness, but truly realization of resource sharing is actually difficult. The traditional resource sharing strategy is borrowing inter-library .But the efficiency is far from satisfied and difficult to scale up due to lack of human resource and insufficient found. Although network information resource sharing does not have technical problem, there are still many barriers in other aspects, causing truly resources sharing to be realized with difficulty. For example, the library manpower is so insufficient as to process the huge network information resource that the information is overdue before it is delivered to users and has lost the sharing value, and so on. What's more, there are still lack of an authoritative regulation body and related legal safeguard. Statistics shows that 80% information resource is possessed by government, leading to less than 20% can be public. Surely, without the participation and management of government, it is very difficult to guarantee information resource to be shared and developed for users. Of course, we can not require government to provide 100% of information resource to be shared by general public. Materials related to national sovereignty, national security and national secret should be preserved in pure text form, or else it will arouse peace problem. Actually, when speaks of so-called information resource sharing, it means resources to be shared to maximum degree. As a result, it is impossible and unrealistic to strive for 100% of information resource sharing.

References

1. Bi, Q.: Necessity and Urgency of Network Information Resource Management. Qin Bao Kexue Journal 18(3), 204–209 (2006)
2. Li, Y.: Characteristics of Management of Network Information Resources. Qin Bao Kexue Journal 18(4), 319–321 (2004)
3. Li, J.: Management of information resources in network. Journal of Academic Library and Information Science 18(3), 17–19 (2001)
4. Chen, N.Y.: Information Resource Management in the Network Environment. Library Tribune 21(2), 49–52 (2001)
5. Mao, X.W.: Development and Application of the Analysis and Assessment System for Utilization of Book Resources in University Libraries. Xian Dai Qinbao Journal 6, 44–47 (2008)
6. Wu, W.C., Zhang, J.Z.: Study of Network Information Standardization System. Qin Bao Kexue Journal 19(1), 40–45 (2008)

Technology Route and System Design of Green IT

Juchi Hou[1], Ying Liang[2], and Jianli Wei[3]

[1] Department of Students Affairs, Hebei University of Engineering, Handan P.R. China, 056038
[2] Department of social work and social policy in college of sociology in Nanjing University, Nanjing, China
[3] Department of Social Sciences, Hebei University of Engineering, Handan, P.R. China, 056038
hbgwy@163.com

Abstract. Green IT is new route of reducing energy consuming and lessening discharging of greenhouse gas. In this paper, we design a technology route for implementing of Green IT. The technology route depends on the coordination of two kinds of relation: firstly, IT manufacturers must adopt advanced technology route. IT manufacturers should use materials of environment protection to produce computers and developing IT technology of energy-saving to make computers be environmental friendly. Secondly, the related ones in the whole society should appeal for system design of Green IT. Governments and related sides should draw up evaluation indexes of Green IT products and raise the awareness of using Green IT products. A comprehensive framework for Green IT technology route is proposed, which provides a foundation for further study.

Keywords: Green IT, Technology Route, System Design.

1 Introduction

After the Copenhagen United Nations climatic change summit convenes, reducing the greenhouse gas emissions became a global-attention topic. Whole world information technology equipment centralism IT infrastructure mass consumption energy and so on use server, memory and switchboard, the direct emission's carbon dioxide has surpassed the global emissions total quantity 2%, every year must discharge about 35,000,000 tons waste gas to the atmosphere, the IT equipment had already surpassed many other professions to the environment pollution degree. At present the IT profession and the IT enterprise initiate Green IT, is must use the information and the communication reduction enterprise to the environment the influence, and must using the Green IT product technology, supply the chain and the service leads other enterprises to march into "the green management and the use", thus promotion environment and economical sustainable development.

2 The Definition of Green IT

2.1 Green IT

The definition of Green IT offered by IDC is that "to design, manufacture, operate and process IT products according to the demand of environment protection and other

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 238–244, 2010.

requirements.” However, such a definition is abstract, it only emphasizes that each chain of IT surviving period should be “Green”. To explain it in detail, the definition of Green IT should meet these demands: low energy consumption while IT products are operating, low pollution while IT products are being designed, manufactured and operating. Low radiation while IT products are being used; at the same time, IT products should save storing room and are easy to be recovered and re-utilized. The Green IT strategy locates such several aspects: First, in computer's manufacturing, must use some environmental protections the material, must consider including the entire computer's use material. In addition, regarding the reject recycling and processing, specially electronic trash's processing, must use the green technology to avoid the nonessential waste or the pollution as far as possible; Second, uses some IT aspect the technology to realize some production life and so on in industrial production equipment and electrical appliances equipment energy conservation questions; Third, through the green technology to IT environmental pollution's and so on product's electromagnetic wave disturbance either radiation governments, specially in the information society, each people possibly by each kind of electronic products surrounding, how the injury which will bring the electromagnetic wave disturbance or the radiation to reduce through each kind of green technology to is smallest.

2.2 Green IT Enterprises

The sustainable environment development has provided the long-term strategic benefit for IT manufacturer on the cost, the brand, the innovation potential and the market localization aspect in the whole. simultaneously also helped the enterprise to reduce from the government, the supervision organization and the money market and so on related benefit side pressure.

The production and the allocation as well as the recycling using the stage energy trail, the IT manufacturer takes again the energy efficiency in the life cycle the product and the service and an environment sustainable development appraisal key indicator. The development strategy appraisal and the localization Green IT manufacturer will release their innovation potential, causes its product to enter the market quickly, and when adaptation emerging and mature market different demand is more nimble. Product which and in supplier's brand, provides aspect and so on life cycle implementation sustainable development's levels will become one of most main measurable. The majority enterprises concentrate in are relatively easy, convenient and the familiar aspect enhance the energy efficiency, to reduce the greenhouse gas emissions, if in saves with the paper, reduces in the business travel and the IT infrastructure. Sometimes these actions can solve enterprise's substantive problem, sometimes actually cannot solve the substantive problem. Essentially, these actions and precedence has no relations. The IT enterprise must and reduces the greenhouse gas emissions aspect in the energy efficiency to adopt a more comprehensive action.

The enterprise first needs to draw up one Green IT the planning for action. This step is essential, because had determined in this stage must achieve goal and anticipated, when regarding can obtain any result as well as obtain such result to be possible to understand clearly in the chest. Various companies should also carry on a time thorough appraisal to the actual energy consumption, and analyzes the IT energy consumption rule, because is impossible to have Green IT universally applicable the solution. This step is helpful to the determination, needs to revise, which things the substitution or to develop, can by be less much energy consumptions to complete the similar work.

3 Technical Route of Green IT

As soon as many government apparatuses have formulated the standard and the laws and regulations encourage the green computer. The US environmental protection organization started in 1992, has contained the stricter computer equipment aspect efficiency demand and to promote the product the tendering system. Did the IT whether green, already become the enterprise to construct in the IT environment a very important festival, while pursues the IT construction performance, must pay attention to the reduced energy consumption, reduces enterprise's IT invested cost and so on. In the product life cycle, the different stage aims at "the green" the goal to have respectively the different request. For instance, in product design stage, this product designs whether already really meets the environmental protection requirement. In the product operational phase, its efficiency may recycle ability, whether to have satisfied the product sustainable use request. An essential spot is, the technology which to achieve Green IT the request uses in the product life cycle, how can also use by the terminal customer. Therefore from the supplier aspect, must stress that the product the lifeline, how does stress these products in the terminal customer the deployment and the use.

3.1 Connections of Hardware and Software

Green IT technology includes hardware, software and various technical contents. The hardware aspect, Green IT the technology already gradually entered the chip, the main engine, the server, the network equipment and so on IT equipment design and the manufacture domain. The Green of hardware places several demands on hardwires being designed and manufactured. Firstly, it must emphasize on reducing energy consumption and discharging of carbon dioxide, lowering harmful screen radiation and radiation of electric-magnetic wave, at the same time, it should save storing room and use data-condensing technology reasonably. The Green of IT hardware also requests to limit the use of raw materials that pose environment pollution and to use more materials that are easy to be recovered, re-utilized, and dissolved.

The software technique aspect, including the virtualization technology, the data optimization and the data reproduction deletion technology, the software dispatch and the managerial technique, as well as each kind of hardware Green IT the technical necessary software system and so on, may raise the existing IT equipment use efficiency through these technologies the goal, thus reduces the IT equipment purchase which and the use day by day effectively increases, then reduced related emissions and resources, energy consumption. Many scientific innovation achievement to raises the energy use efficiency, reduces the environmental pollution to have the important meaning, what now is most successful is the frequency conversion velocity modulation technology application. For instance, has sent a telegram in the production process, 60% electricity were consumed by electric motor, the air blower consumed 20% again, finally passed on the electricity also had 20%; Now, we may on the electric motor upper garment one kind of velocity modulation electronic installation, in long-distance carry on the algorithm control and the operation through the computer, may let the energy decrement which the electric motor consumes 30%. Home use electric appliance equipment and so on air conditioning, refrigerator already had also applied this kind of frequency conversion velocity modulation technology.

3.2 Development of Virtualization Software

To construct high-efficiency data-center and to reduce energy consumption through fabricating and combining of service-devices are values of Green IT.

- How to reduce costs for enterprises by relying on “Green IT”?
- How to reduce energy consumption?
- How to cool the data-center better while paying fewer costs?
- How to construct a better data-center?

All these issues above are fundamental elements of Green IT. Application of fabricated software can solve the problem of average utilizing efficiency of system service devices. Through combining service devices and increasing utilizing –rate of service devices, numbers of physical service devices can be reduced and less energy consumption can be attained. The following table shows electricity consumption before and after fabricating of service devices (assessed by online tools in a domestic computer company) through calculating, after service devices are fabricated, much electricity consumption can be saved.

Table 1. Electricity consumption before and after fabricating of service devices

	Before fabricating of service devices		**After fabricating of service devices**	
Sorts of service devices	Numbers of service devices	Power(W)	Numbers of service devices	Power(W)
Single CPU	20	475	0	550
Double CPU	50	550	10	675
Four CPU	10	950	2	1150

3.3 Correspondence Softwares become a Tendency

Carbon footprint has gradually entered into category which enterprises Green IT need pay attention to. The so-called carbon trail usually refers to the carbon dioxide quantity which a person moves discharges. Enterprise which travels on official business frequently regarding the staff, not only the travel expense cost is very high, moreover in the journey delivers the carbon dioxide emissions to is also not a small harm naturally.

For instance the use video frequency conference software, causes the different place the staff sits in a hypothesized conference room can carry on the very good exchange, not only saved the time, raised the working efficiency, because of the reduced carbon dioxide emissions, has made the contribution for Earth's green. The Green IT uncertain need complex method, may relaxed realize the enterprise through the software Green IT. For instance the Chinese ten thousand branch companies through reduce its correspondence hardware, reduced the carbon dioxide emissions. Wan Ke more than 20 Branch office has merged the correspondence infrastructure, concentrates the headquarters the core communications system, provides the unification correspondence function through a IP voice communication network to all Branch office. Thus, not only helped ten thousand branches to reduce the management and the maintenance demand, reduced the power loss, simultaneously realized has been quicker, a more consistent service, raised the cooperation efficiency and the service continuity.

4 Green IT System Design

4.1 To Realize “Green” Management and Production

Green IT is not only a technical question, but also a management question. In the modern management idea already had the management green concept, soon that kind the close relation consciousness which exists to natural environment's influence between is called the management to the organization decision-making and the activity and the organization the green. Green IT the technical product’s application must form the product and the management gradually in the enterprise “the green” the common values, then the performance to pursue the environmental protection the enterprise culture environment. Only then like this, Green IT the technical product and the application only then can more penetrate and continue to obtain the development. It is well known, what enterprise's community responsibility are more is realizes through leader’s decision-making and the enterprise all staff's combined action. Therefore, the enterprise Green IT the technical product's purchase and the application need the head of undertaking the correct decision-making, needs the enterprise staffs joint effort. The modern enterprise’s daily operation can apply each kind of IT equipment, “the green” as well as realizes “the green” these IT equipment the management is the essential requirements which organizes to the enterprise, in this most representative is the International Standardization Organization through ISO14001 environmental management system authentication. Now the ISO14001 authentication already accepted for the domestic many enterprises, and becomes manifests its community responsibility and the promotion enterprise image important means. The enterprise organizes the energy conservation which undertakes, to reduce the row of community responsibility is manifests through the enterprise operation process, in this process, the enterprise may through the consideration production operation environmental effect, take the measure, the application Green IT the technical product and the corresponding solution, realizes to the green management and the production pursue.

4.2 System Benefit

At present ,Green IT technology faces two main problems, First, generally speaking, cost of application of Green IT technical product is higher than ordinary products, causing final selling price to be also higher, this will cause the related product the attraction to be insufficient. Second, Green IT the technical product needs to form a complete set the use to be able to display the biggest benefit, because now massive must Green IT technical equipment also in use, causes necessary Green IT the technical equipment to be unable with it to display the biggest benefit. Third, the part Green IT the technology realizes the difficulty to be big, the reliability also waits for customer's further approval, therefore these Green IT the technology itself also waits for further consummating. Regarding this, on the one hand may enhance the public through the public opinion propaganda to be right Green IT the product understanding and the approval, expands its use the community. On the other hand must develop the correlation technique which and the equipment unceasingly and the core Green IT the technology forms a complete set, forms the system Green IT the solution. In addition, must enlarge Green IT unceasingly the technical investment, promotes its reliability compatibly unceasingly, and so on. Thus the country must support vigorously, must support the electric automobile to be the same likely, expands the investment and the support, such enterprise will also have the power. The enterprise is very often difficult alone to withstand the funds the pressure, needs the government, the scientific research courtyard, the university several aspect mutual unions, the present produces study grinds the cooperation as well as the industrial alliance is the quite good pattern, displays the respective superiority, strengthens some forward-looking research, masters the related core technologies. The user must transform the purchase product positively the idea, purchases some energy conservation product on own initiative, in the long run, the social populace also want to undertake some energy conservation to reduce the row of community responsibility.

4.3 “Green” IT Business

Environmental protection has already started to affect and dominate enterprises’ IT purchase, although the present field was also very difficult to discover one synthesis judgment standard to come the quantification to weigh the server and the memory product energy conservation situation, but many enterprises are trying to develop now in view of server’s energy efficiency target, started to begin to solve the data central memory energy consumption problem, and hoped that the development used in the data central saving hardware’s energy efficiency target. Through the third party organization and the manufacturers joint effort, can establish the science the judgment standard, helps the user when chooses and purchases the IT infrastructure has according to may depend on. In the government procurement aspect, China announced “Environment Symbol Product Government procurement Implementation Opinion” outward officially in November 22, 2006 and first batch “Environment Symbol Product Government procurement Detailed list”, the request all levels of the government institution, the Institution and the association organize when uses the financial fund carries on the purchase first to select and purchase the environment symbol product.

5 Conclusion

Green IT has brought an IT industry revolution, simultaneously through its products, the industrial chain and the related service. It has also brought a consumption revolution for its terminal customer. By using advanced technical route and entire society's Green IT system design and raising the public's awareness and using of Green IT products. It will offer a pointed solution for the worldwide difficult problem of reducing energy consumption and greenhouse gas emissions.

References

1. Johnson, M.B., Voas, R.B., Lacey, J.H., McKnight, A.S., Lange, J.E.: Living Dangerously: Driver Distraction at High Speed. Traffic Injury Prevention 5(1), 1–7 (2004)
2. Brown, I.D., Tickner, A.H., Simmonds, D.C.V.: Interference between Concurrent Tasks of Driving and Telephoning. Journal of Applied Psychology 53(5), 419–424 (1969)
3. Nunes, L., Recarte, M.A.: Cognitive Demands of Hands-free-phone Conversation while Driving. Transportation Research. Part F, Traffic Psychology and Behavior 5F(2), 133–144 (2002)
4. Recarte, Nunes: Effects of Verbal and Spatial-Imagery Task on Eye Fixations while Driving. Journal of Experimental Psychology Applied 6(1), 31–43 (2000)

The Tail of BPM

Steve Kruba and Jim Meyer

Northrop Grumman, 3975 Virginia Mallory Drive, Chantilly VA 20151, USA
steve.kruba@ngc.com, jim.meyer@ngc.com,
http://www.northropgrumman.com

Abstract. Business process management suites (BPMS's) represent one of the fastest growing segments in the software industry as organizations automate their key business processes. As this market matures, it is interesting to compare it to Chris Anderson's 'Long Tail.' Although the 2004 "Long Tail" article in *Wired* magazine was primarily about the media and entertainment industries, it has since been applied (and perhaps misapplied) to other markets. Analysts describe a "Tail of BPM" market that is, perhaps, several times larger than the traditional BPMS product market. This paper will draw comparisons between the concepts in Anderson's article (and subsequent book) and the BPM solutions market.

Keywords: BPM, BPMS, business process management, business process management suites, workflow, markets.

1 Introduction

In October 2004, Chris Anderson, editor of *Wired* magazine, wrote an article on what he describes as a "new economic model for the media and entertainment industries" called "The Long Tail." [1] In the article, he describes how companies like Amazon.com have capitalized on the internet to create niche markets that were previously too small to service by traditional retailers. The name is derived from mathematical power laws in which the two shaded areas in Figure 1 are equal. [2] The area to the left is referred to as the *head* while the area to the right is referred to as the *tail*. Since the tail is infinite in length, it is known as the *long tail*.

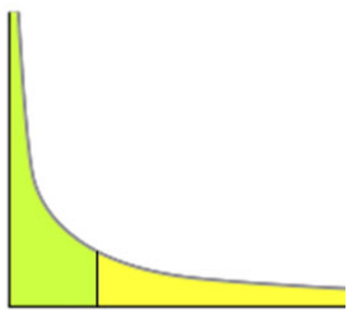


Fig. 1. Power Law

Anderson's seminal article generated such interest that in 2006, he extended his treatise by publishing *The Long Tail: Why the Future of Business is Selling Less of More* [3] based on considerable input from the global internet community. In the book, in addition to much more detailed research on Amazon, Netflix, and others, he extended the principles of the long tail to Kitchenaid (long tail

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 245–257, 2010.

of colors), Legos (long tail of created markets), Google (long tail of search) and others. Just how far does this concept extend?

Business process management, or BPM, is a management discipline that typically leverages technology to improve the processes that drive businesses. [4] A broad spectrum of technology vendors such as IBM Oracle, PegaSystems, and Northrop Grumman produce software products targeted at the broader BPM market. We were first introduced to the concept of the tail of BPM by Janelle Hill of Gartner.[1] The idea behind the BPM tail is that some vendors service a market, focused on *solutions* that is an extension of the traditional BPMS market. This BPM tail is *somehow* similar to the way in which the niche book market serviced by Amazon relates to the traditional book market. Some suspect that this BPM tail market is significantly larger than the traditional BPM market, and that it is driven by solution sales rather than product sales.

What validity is there to the tail of BPM market and how does it relate to the long tail phenomenon? That topic will be the focus of this article. We'll examine Chris Anderson's original thesis and explore the similarities and differences with the BPM market and examine why and how the long tail seems to apply. We will make a distinction between pure BPM product sales and solutions sold into this market.

2 The Long Tail

Brick and mortar bookstores have a simple economic constraint: a finite limit to the number of books that they can stock. At low sales volume, the profit generated from a particular title is offset by the ordering and warehousing costs.

Chris Anderson uses the power law graphic shown in Figure 2 to describe each of the markets he addresses.[2] Although the graph appears to be a continuous curve, the "curve" represents the peaks of a bar chart. In the case of book sales, the left-most bar would represent sales for, say, the *Bible*, while the adjacent bars would represent sales for *Quotations from Chairman Mao*, the *Qur'ān*, *A Tale of Two Cities*, etc.[3] As the curve moves to the

Fig. 2. The Long Tail

[1] Conversation between Janelle Hill and Steve Kruba at the Gartner BPM Summit in San Diego, March 2009.

[2] The Long Tail, inside back cover.

[3] Excerpted from a list of all-time best selling books at `http://en.wikipedia.org/wiki/List_of_best-selling_books`

right and down, it represents sales for less popular books until it finally reaches the point – identified by the shading change – where a particular bookstore can no longer economically carry a particular title.

The Products dimension (the x-axis) is simply the ordered set of bars, listed in decreasing order of book sales. The Popularity dimension in the case of book sales (the y-axis) could be either units sold or revenue. The area under the curve represents the revenue from all books sales. The area on the left (the head), representing the sale of popular or "hit" titles, is comparable to the area on the right (the tail), representing the sale of niche books. In Amazon's case, the tail represents about 25% of book sales.[4]

Amazon's innovations in addressing this long tail market have been numerous:

- Leveraging the internet to internationalize its market, greatly increasing its scale. Brick and mortar markets are local.
- Providing an electronic forum for sharing recommendations, increasing demand.
- Using internet forums to create niche markets for timeless classics that may have disappeared from the public consciousness.
- Creating a partner ecosystem that allows Amazon partners to sell through Amazon, driving Amazon's inventory costs for those titles to zero while creating a huge "Amazon" inventory that attracts customers to the site.

While the average Barnes and Noble store carries 100,000 titles, Amazon carries over 3.7 million[5] – comparable to large academic libraries. The original book market addressed by Amazon represented a market restricted by shelf-space and physical inventory limitations. Amazon's business model greatly reduces these restrictions.

Markets similar to the book market are the music CD and DVD video markets. In a similar fashion to the way that Amazon changed the book market, companies like Rhapsody and Netflix have leveraged the long tail in their respective markets. An interesting variant on this market is the iTunes market. iTunes allows their customers to download electronic versions of individual songs rather than complete albums – a practice that is complicated by the way that the music industry licenses these songs. iTunes has virtually no inventory since the cost of storing electronic bits is essentially zero. The principles that apply to power laws and market economics still apply. What is interesting in iTunes' case is their variation on the original theme,[6] indicating that a literal interpretation of the Amazon formula would miss out on great opportunities.

[4] The Long Tail, p. 23.
[5] The Long Tail, p. 23.
[6] Using long tail economics combined with a modified product offering.

3 Anderson's Innovation

Chris Anderson was not the innovator that created long tail economics. As pointed out in *The Long Tail*, retailers for more than a century have innovated on this theme. Richard Sears, in the 1880's, stumbled upon the concept of the mail order catalog (leveraging global warehousing) that culminated in Sears, Roebuck, and Company[7] – for decades, the world's largest retailer.

Anderson's innovation was his analysis and synthesis of the principles at work in the long tail.[8] He introduced two sets of principles that are germane to our BPM discussion and they will be briefly summarized here.

3.1 The Three Forces of the Long Tail

Anderson first recognized three elegant principles that drive much of the internet economy and referred to them as "the three forces."[9] Although these forces emerged because of the new capabilities introduced by the internet, their application is not restricted to internet commerce.

Force 1 *Democratize the tools of production.* A prime example of this new capability comes from the personal computer, allowing everyone and anyone to produce books, music, and videos. In essence, we are now all producers.

Force 2 *Democratize the tools of distribution.* The internet allows everyone to make their "productions" available to a wide audience, providing the critical mass to service niche interests that could not be served by geographically limited markets or traditional means of distribution.

Force 3 *Connect supply and demand.* Internet technologies such as web sites, web search, blogs, and message boards provide formats that allow suppliers and consumers to connect. Traditional methods such as advertising and word-of-mouth are still at work but on a limited scale.

3.2 Long Tail Rules

The last chapter of *The Long Tail* is titled "Long Tail Rules" and provides a list of rules that can be applied to take advantage of the forces shaping the new economy.[10] Table 1 describes these rules.

---

[7] The Long Tail, p. 42.

[8] Anderson also focused heavily on the social implications of the long tail. It's very instructive to review his thinking around the notion of our 'hit culture.' Our predisposition to the "hits" in a particular market is more a consequence of restrictions in the market's shelf-space, theater screens, or distribution channels than our herding instinct. Hits in the BPM space might be equated to "best practices." This topic is worthy of a separate discussion.

[9] The Long Tail, pp. 52-57.

[10] The Long Tail, pp. 217-224.

Table 1. Long Tail Rules

No.	Rule	Description
❶	Move inventory way in . . . or way out	Centralize warehousing or eliminate inventory entirely with electronic "bit" products (like iTunes) or through partnerships like Amazon's.
❷	Let customers do the work	Peer production, ala eBay, Wikipedia, and Craigslist.
❸	One distribution method doesn't fit all	Sometimes you want to buy clothing on-line and sometimes in a store.
❹	One product doesn't fit all	"Microchunking" products into component parts allows customers to recombine them to meet their needs.
❺	One price doesn't fit all	In markets with a wide variety of options, allow for creative pricing to address customer constraints.
❻	Share information	Transparency on product information breeds customer trust and facilitates the matching of supply and demand.
❼	Think *and*, not *or*	The new markets are not zero-sum games.
❽	Trust the market to do your job	Inefficient markets require you to guess what sells (in order to manage advertising costs) but efficient markets are self-leveling.
❾	Understand the power of free	"Free samples" and free content like television have low or zero cost but allow consumers to make important buying decisions.

4 The BPM Market

Just what is the "BPM market?" Traditionally, it has been tracked as the market for business process management products: software evolving out of the workflow software market. Analyst groups like Gartner and Forrester Research track sales of BPM product license and maintenance fees, advertising expenditures, and other measures to determine the "market leaders." But it has long been recognized that software sales are just the tip of the iceberg: complete solutions in the BPM space can be 4 to 6 times product revenues (or more) when services are factored in.

Business processes are the life-blood of all organizations. Whether the organization in question is a government entity or a commercial business, every organization is unique. In the federal space, many of these government entities are truly unique in that there are no other organizations quite like them. In the commercial space, companies serving the same market are similar, but each has its own "secret sauce" that differentiates it from its competitors, allowing it to put its unique brand on its segment of the market. That secret sauce is its set of business processes.

Automating a business's key processes is, perhaps, the most intrusive of any automation effort, fundamentally changing the way an organization does business. Because of the criticality of this intrusion, a solution that is compatible with its needs is important, as opposed to force-fitting the solution to the product's capabilities. For this reason we would suggest that creation of a solution in the BPM space is more extreme than in any other area of automation; e.g., selection of a relational database or even an enterprise resource planning (ERP) package.

Furthermore, BPM is hard. Solutions involve a blend of many technologies and key awareness of the business drivers of each process. BPM software tools offer a wide choice of options for solving particular problems, many of which might seem like good ideas, but often are not. Experience in fielding solutions in this space is critically important.

The BPM software product market grew out of a number of very different technologies.

- The so-called "pure-play" BPM vendors evolved out of the imaging and workflow market with strong capabilities in the human-centric aspects of the solution. The pure-play vendors have evolved into the BPMS vendors.
- Related product categories emerged in the areas of rules engines, simulation/optimization engines, business process analysis (BPA) tools, business activity monitoring (BAM) listeners, and correlation engines.
- Enterprise applications integration (EAI) merged into the BPM market by adding human-centric capabilities to their heritage of connecting machines to machines.
- ERP and customer resource management (CRM) vendors added workflow to their products.
- Document management and records management vendors added workflow to their repository capabilities to facilitate review of the documents that they support.
- Portal vendors are leveraging their control of web user interfaces by adding workflow to their products.
- And on and on ...

There is virtually no software product category in existence today that does not claim to have process management capabilities. But the focus and extent of these capabilities varies greatly, depending on their heritage and the aspect of the market that they serve. Since business process is such a broad term, it provides a wide basis for differentiation of the market or markets that have developed. Within this context, the notion of *a* BPM market is suspect.

4.1 BPM Solutions

A key concept in this paper is understanding the distinction between the BPM market and solutions created to service specific customer requirements in this market. The BPM software products identified in the list above exist in order to create applications that service customer needs in support of their business processes. Some of the products are designed to produce complete solutions while others service some aspect of a solution.

Although the end goal in all cases is to solve a problem or produce a solution, the products themselves can be sold without an associated solution; i.e., sold to an IT organization or a contractor in order to create a solution. Although this distinction may seem obvious and apply to all software categories, it is especially important for BPM solutions where the domain knowledge obtained through experience in what works and what doesn't work is hard-earned and especially important.

Let's revisit the notion of the tail of BPM. There are a number of levels on which relating the BPM market to the long tail adds insights – and some areas in which it does not seem to fit.

5 Dissimilarities: BPM vs. the Long Tail

First, let's look at the aspects of the BPM market that does not seem applicable to the premises of the long tail.[11]

- *Product vs. service.* As suggested earlier, the market comprising business process solutions has significantly larger services revenue than product revenue, due to the need for subject matter expertise or domain knowledge in two areas: (1) the customer's business; and (2) experience fielding solutions for a broad customer base. These services are not particularly related to tool-support, since the largely graphical tools are fairly intuitive to use, but to experience in understanding what works and what doesn't work within a particular customer's industry.
- *Product components.* BPM products have more complex delivery models than books, CDs or DVDs. These traditional products are discrete physical objects, even when those physical objects comprise a stream of computer bits. Selling an iTunes track is always the same for the same piece of music.[12] Two BPM solutions are never the same, even when they are created from the same base product. If you limit the discussion to the pure product components, the solutions are still so different that they are difficult to compare.

[11] It's important to note that the long tail does not apply universally. See Tim Wu's contrarian view in *Slate* magazine: "The Wrong Tail: How to turn a powerful idea into a dubious theory of everything." [5]

[12] Recognizing that there are alternate mixes for some music.

- *"Warehousing costs."* A major part of the innovation in long tail success stories was in the reduction in warehousing costs, which included distribution costs. BPM solutions involve virtually no warehousing costs: just the costs of storing the software on CD/DVDs for installation. The BIG cost for BPM solutions – the cost of creating the solution – could be viewed as "distribution costs." To some extent, these costs fall outside of the long tail concept, but as we'll see in the next section, have relevance to our discussion.
- *The power law graph.* Try to assign labels to the axes of the power law for BPM. The y-axis might be licenses sold or dollar sales, but this doesn't really add much value in the case of the BPM market. For books, the graph represents book sales for different books sold by the *same* bookseller. Every "book" sold (i.e., solution deployed) by a single BPM product vendor is unique, so the concept of a tail seems to lack relevance. If we graph the total product sales for *every* product vendor – *not* what we do for books and music – highest sales to the left and lowest to the right, we have something interesting, indicating who the largest vendors are, but the critical break point that is so significant in the traditional long tail markets is meaningless within this context.
- *Successful sales.* Discrete products like music, books, and DVD sales are binary – you are successful if you make a sale: there are no shades of gray. For software solutions, you can create a good solution, a bad one, or something in between. So there's more dimensionality to the BPM problem than in the traditional long tail markets.

6 Similarities: The Tail of BPM

But many who have studied Anderson's work come away with the feeling that there's a lot more to the comparison than shown in the previous section, whether it is for BPM software sales or sales in other markets. This section will provide some analysis to support the applicability of the long tail to the BPM market.

One interpretation of the Anderson power law graph to the BPM space is that the break point between the head and the tail relates to product sales to Information Technology (IT) organizations vs. direct sales to business customers. The trend toward having business customers lead the acquisition of BPM solutions has been gaining momentum for some time, with their IT organization playing a supporting role. In this comparison, vendors that traditionally sell to IT organizations (like IBM or Oracle) represent the head of the market while vendors that sell into the business organizations (like Northrop Grumman) represent the tail because domain knowledge dominates product content in solutions sales.

Another interpretation for the tail is the degree to which uniqueness is a key aspect of the solution. In this view of the market as illustrated in Figure 3, the x-axis represents solutions rather than products. Common, repeatable solutions using pure product with some configuration (as opposed to customization) comprise the head while solutions requiring significant services comprise the long tail. It's important to note that these services do not necessarily require custom software created from scratch, but could rely on a rich solution toolkit that requires a mix of custom glue code and product expertise.

The notion of a tail in this context is that solutions are like the rare book that is now being reprinted because the expanded market makes it economically feasible to produce it again. These "tail" BPM solutions are now enabled by more cost-effective technologies for producing them. We'll see examples of this later. It's also important to note that a particular vendor can participate in both the head and tail of this view of the market.

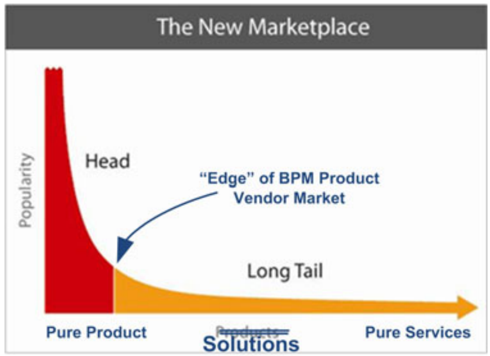


Fig. 3. The "Solutions" Tail

The long tail relates to scarcity, and within the BPM context, scarcity relates to cost. Historically, many solutions were technically feasible, but customers did not consider them cost-justified. New rapid solutions creation capabilities and the resulting agility discussed in the next section make solutions "less scarce."

Finally, another interpretation relates to Anderson's discussion of head vs. tail as hits vs. niche products (using Anderson's entertainment theme). Since BPM solutions are very individualized and unique, they are almost by definition a niche market. There are solutions in the BPM space that involve cross-organizational reuse, but they are the exception rather than the rule. They are referred to by various terms, including "templates" and "packaged business applications" and represent a true head to the market if reuse is the criteria.

7 BPM Agility

A key concept in the BPM market is *agility*. Agility relates to how readily a solution can be changed over time to reflect changing market conditions. Since the purpose of a BPM solution is to improve an organization's business processes, brittle automation approaches that freeze the operation of the business to unacceptable levels is counter-productive.

Modern, successful BPMS products provide their customers with agility. But how do they do so? These solutions are generally configuration-driven, with graphical designers used to define the business process and the user interfaces. By design, these solutions are agile.

But generic, successful BPM products are much more than that. BPM solutions rarely exist in an IT infrastructure vacuum and seldom provide the complete solution on their own. As a result, they must provide interface capabilities for pre-existing components of the overall solution. We refer to these related applications as "legacy" applications. These legacy applications can be modern solutions developed in a service oriented architecture (SOA) or decades-old, monolithic applications, developed before the dawn of the modern innovations used to create solutions today. The BPM toolset needs to offer a heterogeneous mix of capabilities in order to interact with this wide array of legacy technologies.

Examples of the types of integration that might be required for this complete solution include the following:

- Screen scraping technology to push and pull information into mainframe CICS systems that were designed long before integration needs were designed into solutions.
- Flat files used to push and pull information into legacy applications for which security restrictions preclude direct interaction with the information stores.
- Modern integrations based on flexible SOA architectures.
- Single sign-on solutions where legacy components require customization.
- Data exchanges where the information bases are incompatible and require translation.
- Interacting applications with dissimilar security models.
- And many more ...

Although integration middleware products exist for some of the above examples, there are no products that address all of the requirements that arise in automating complex business processes. This is where Anderson's concept of "microchunking" becomes interesting. Think of the long tail in this market as a set of not-often-used components that can be cobbled together to produce solutions. So in the BPM market space for a particular vendor, the head represents those components of their solution that represents their sweet spot targeted toward the mainstream market while the tail represents those innovations that allow them to address non-standard features. All competitive BPM product suites have customization capabilities, but those successful at addressing the long tail have a broad collection of tools and approaches to addressing those needs.

Some "microchunking" examples from the Northrop Grumman's e.POWER®[13] customer base might be helpful. These examples are in increasing order of reuse.

❶ *One and done.* A true one-off component built for one and only one customer. For example, reading a Defense Finance and Accounting (DFAS) print stream for the Defense Contract Management Agency (DCMA) contract management solution.

❷ *A template or programmable component.* For example, the e.POWER Quick Add feature for rapid document upload or the Active Directory look up component for single sign-on. Quick Add starts with a template, but needs to be configured for each document type, building specific objects with pre-set

[13] e.POWER is a commercial BPMS product and a registered trademark of the Northrop Grumman Corporation.

attributes, uploading and attaching documents. So Quick Add is a framework with many common re-usable elements, but the resulting objects are one-off's.

❸ *Reusable components* – two-off or more. In this example, a component can be used without alteration for more than one solution. For example, our DFAS interfaces are used in multiple solutions at the DCMA. Our PeopleSoft feed is a government off-the-shelf (GOTS) component of our eOPF product.

❹ *Configurable components* – no programming required, but technical skill or knowledge may be required. Examples include LDAPSynch (LDAP repository attribute integration), InfoSources (framework for reuse of relational data), and the XIA (XML Intelligent Agent – a web services client). The results produced are one-off configuration items, but not as restrictive as a template and use a COTS component instead of a programmable framework.

❺ *Configurable architecture* – for DBAs and Systems Engineers. Used to configure servers, databases, etc.

❻ *Configurable components* – no technical skill required: business analyst driven. The e.POWER Applications Builder for building end-user interfaces, the WorkManager Builder for graphically building workflows, and the WorkPlace Manager used for configuring user roles are designed for subject matter experts, not programmers.

NOTE: GOTS components or application specific components – like the DCMA example – are mostly at level 2 or 3 and occasionally 4. Services apply to all, but services at levels 4 to 6 are standard deployment services across solutions whereas services at levels 1 to 3 are "in the tail."[14]

8 Anderson's Principles Applied to the BPM Market

As noted earlier, Anderson established a set of principles that apply to the long tail and these principles have applicability to the BPM market. First there are the "market forces":

Force 1 *Democratize the tools of production.* Modern BPMS products field design tools that extend the development process to staff members outside of the traditional IT department, allowing business users to create aspects of the solution for which their expertise is critical, such as assisting in the design of the graphical workflow map. Graphic artists can participate in the process by designing user interfaces with products like e.POWER that have rich forms-design features that can be used by non-programmers.

Force 2 *Democratize the tools of distribution.* "Distribution" for BPM solutions is really the same as "production" since the cost of fielding the solution is minimal compared to the solution creation costs.

[14] The benefits of using COTS products as the basis for GOTS products is a subject worthy of separate treatment.

Force 3 *Connect supply and demand.* Connecting vendors with customers can be facilitated by the internet, but unlike the long tail market, traditional advertising and word-of-mouth are still dominant.

The long tail rules also make sense within the BPM context. Table 2 summarizes these rules.

Table 2. Long Tail Rules Applied to BPM

No.	Rule	Relevance to BPM
❶	Move inventory way in ... or way out	Consider pre-development of "microchunks" as "product inventory" as a facilitator for producing solutions.
❷	Let customers do the work	Democratized production methods, per **Force 1**.
❸	One distribution method doesn't fit all	If distribution relates to solution creation, having gradations of participation-levels for different customers is valuable, some preferring to participate in creating the solution and others preferring to manage the creation process.
❹	One product doesn't fit all	"Microchunking" improves the matching of needs with capabilities.
❺	One price doesn't fit all	This applies in particular to addressing the wide-range in participation levels among process workers, from heads-down users to on-the-fly occasional approvers.
❻	Share information	Transparency on product information breeds customer trust and assists the matching of suppliers and consumers.
❼	Think *and*, not *or*	BPM solutions require integration with a broad spectrum of related technologies, requiring partnerships with other technology vendors.
❽	Trust the market to do your job	Rapid solutions creation and agility sell themselves, allowing customers to quickly and easily understand the value proposition.
❾	Understand the power of free	Rapid solutions creation reduces solution creation costs, allowing customers to achieve initial benefits at greatly reduced cost.

9 Conclusions

Although the long tail is not a perfect fit for analysis of the BPM market, valuable insights can be gained into an examination of the so-called tail of BPM market that emphasizes BPM *solutions* over BPM product sales. Significant BPM solutions require customization, integration, and experience with many deployments; product sales that are not coupled with experienced service providers increase the risk of failure. Although there are a number of ways in which to relate the 'Tail of BPM' to the 'Long Tail,' we believe it is primarily a distinction between pure product sales to IT organizations and solutions sales to business organizations, where BPM rapid solutions creation capabilities overcome "scarcity" by enabling solutions that were previously not economically feasible.

References

1. Anderson, C.: The Long Tail. Wired (October 2004), http://www.wired.com/wired/archive/12.10/tail.html
2. Wikipedia, http://en.wikipedia.org/wiki/Long_Tail
3. Anderson, C.: The Long Tail: Why the future of Business is Selling Less of More, Hyperion, New York, NY (2006)
4. See, for example Cantara, Michelle: BPM Research Index: Business Process Management Technologies, Gartner (September 3, 2009)
5. Wu, T.: The Wrong Tail: How to turn a powerful idea into a dubious theory of everything. Slate (July 21, 2006), http://www.slate.com/id/2146225

The Aggregation-Diffusion Model of Conduction of Technology Value in the Large-Scale Sports Events

Chuansheng Dong

Sport Social Science Center of Shenyang Sport University, Shenyang China 110102
dongchuansheng@hotmail.com

Abstract. This paper studies the aggregation and diffusion of technology development in the large-scale sports events. We analyze the great aggregation and diffusion effects of the aggregation-diffusion conduction of technology value. We demonstrate the basic path of the conduction and build the aggregation-diffusion model based on the conduction in the large-scale sports events. We illustrate the intrinsic mechanism of the conduction operation with the value chain of technology as its basic structure, and the technology innovation as well as the value conduction as its fundamental driving force. We discuss the special functions of the conduction in the public technology service system. This paper briefly explores the feasibility, the implement path, and the government responsibilities in the large-scale sports events held in China by building the aggregation-diffusion model of technology development.

Keywords: Large-scale Sport Events, Technology Value Conduction, Technology Innovation, Aggregation-diffusion Model.

1 Introduction

Large-scale sports events play a more and more distinctive role as the special platform of technology development. Our research results [1] show that the technology aggregation and diffusion development model widely exists in the large-scale sports events, which has the remarkable effects and achieves the technology value-adding through the aggregation and diffusion of the technologies. Technology aggregation and diffusion is the basic process of developing technology aggregation and diffusion. Aggregation is the process of collecting, combining, and integrating technologies by which new products, technologies and techniques are created and innovated. Diffusion is the process by which the created products and innovated technologies and techniques are widely spread in every field of the society by way of technology aggregation and contributes to improving the technology level and innovative capacity.

Technology development which plays a fundamental part in the large-scale sports events with its growing influence has aroused more and more attention of the society. Hereby, it is very important and urgent to study the establishment of special model of technology self-development in large-scale sports event. However, scholars seldom focus on the aspect of technology in their research of large-scale sports event, which led

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 258–268, 2010.

to few research results. And in China, scholars have done their researches of the large-scale sports events from the perspective of social risk [2], economic and social benefits [3], and government responsibilities [4] [5]. In only some researches of technology, they start the study from the perspective of the investigation of technology application [6] with more descriptive while fewer theoretical researches. Therefore, it is urgent to increase the study of technology development in large-scale sports events.

The typical aggregation-diffusion effect is objectively formed by the intense aggregation of technology in the specific large-scale sports events which attract attention of the whole society. Technology aggregation-diffusion effect is about all the effects caused by which technology innovations with the products have been increasingly aggregated in the field of large-scale sports events and spread in the society through holding the events. So, it is a complex effect combining effects of aggregation and diffusion.

Viewing on the phenomenon, the technology aggregation effect is considered as the scale effect caused by the aggregation of technology achievements, that is, technology achievements from the whole society are aggregated in the large-scale sports events. The advanced technologies including construction, new material, information, and biometric technologies are aggregated in the stadium construction known as the “Bird’s Nest” in Beijing Olympic Games. The aggregation of these technologies in the field of construction has objectively catalyzed the innovations of technology, accelerated the pace of innovations, and enhanced the capacity of national technology innovations. The research results about the significant function of technology in 2006 Torino Winter Olympic Games done by Andrew McDowell [7] provide us with a better sense of the great and important impact of technology development on winter large-scale sports events which is more dependable on high technology.

Referring to the technology aggregation which has broke the barriers between industries, it is necessary to find the way of how to make the technologies aggregate with each other. Objectively speaking, there are two kinds of technology aggregation, which are the aggregation of different kinds of technologies and aggregation of the same kind of technologies aiming at solving new technological problems. The first kind of aggregation is about the innovative improvement between technologies, which focuses on how to match one certain technology with other kinds of technologies. While, the second one is the process of re-applying the technology innovations, which requires that current technology be constantly innovated to solve new technological problems so as to expand the technological functions and its applied areas, as well as innovate new technologies. It can be summarized that the focus of the first and second kinds of aggregation is absolutely about the innovative application of technologies, which will promote technology innovations and probably lead to the formation of technology cluster.

Diffusion effect is that the aggregated technology outcomes including the technology products, technology knowledge, and innovative spirit are spread in every field of the society owing to the promotion by the modern media through holding large-scale sports events.

The technology diffusion firstly has the effect on the development of people’s awareness. While people watching or getting to know about the large-scale sports events, they will get the information about the technology application by different kinds of media, which will help them develop and enhance their scientific and innovative awareness. Besides, it has the effect on overall enhancement of national technology innovation capability. For example, China has brought a large amount of manpower, material and financial resources, and all kinds of technology achievements together to make innovations in order to promote the High-tech Olympic Games strategy, which has not only expanded the innovative fields but also shortened the period of technology innovation, furthermore, enhanced China's innovation capability. Last but not the least, it has the enormous influence on the development of the technology industry: on one hand, the achievements aggregated through large-scale sports events will certainly and directly be known and understood with the heavy exposure of the media by the technology corporations, which is good to break the technology barriers between industries and immediately raise the overall technology level and increase the innovative capability of the technology industry; on the other hand, the aggregation of different kinds of technologies will objectively accelerate the aggregation between technologies, form a new model of technology innovation, and enhance the innovative capability in all the technology fields.

The essence of the technology aggregation in large-scale sports events is about the kind of special innovation model and the technology transfer process. According to the latest research results of technology innovation and transfer by Chinese and foreign scholars, there is an outstanding social effect in the process of transferring technology. Roger (1995) thought that [8] technology transfer is a general phenomenon of developing the technology, a process of developing and changing the social structure and function. Hereby, technology diffusion will effectively push the development of technology industry and speed up the development of economic society. The large-scale sports events have a more and more powerful influence on the society by the increasing magnification of the media, and even become one of the important social events which have a profound effect on the development of international societies at a certain stage. After doing the researches, scholars believe that every large-scale technology diffusion will bring the many economic and social benefits with the help of modern media [9], and has the great effect on the international society development [10] [11].

2 Technology Aggregation-Diffusion Model in Large-Scale Sports Events

2. 1 Technology Aggregation-Diffusion Path

Based on my researches on Beijing High-tech Olympic Games Model, the study of technology model in other large-scale international sports events is highlighted in this paper to prove the universality of the previous research results which indicate that the technology aggregation-diffusion path widely and objectively exist in large-scale sports events, in which there are two kinds of technology development path, that is, the aggregation and the diffusion path.

The first kind of path is the path of technology aggregation. In the large-scale sports events, technology aggregation path is a complex path which is made up of three sub-paths, which involves the technology aggregation through the market, the government, and the society. The aggregation through the market is an aggregating process of implementing the economic value of technology which is based on the demand and help of the market. Most of the technology corporations take an active part in the process of holding large-scale sports events and set up a specialized system of technology management for the events. The aggregation by the government is that government purposefully guides and promotes the aggregation of the technology power to the field of large-scale sports events in order to achieve the goals of enhancing the technology innovation capacity and provide better service to the society through technology. The High-tech Olympic Project actively pushed by Beijing Olympic Games is a typical example of the aggregation through the government, which is that China's government plays an important and representative role in holding the large-scale sports events and expanding the technology aggregation effect. The aggregation through the society is that some of the social groups promote the technology aggregation actively to implement their own purposes in the light of their respective value orientation. Some independent social organizations including international and domestic organizations carry out some activities to accelerate technology innovation and reasonably achieve technology value with the purpose of the organizations, which is the third sub-path of technology aggregation in large-scale sports events.

While, two technology aggregation path can be summarized from the function implement point of view, which are the path of technology innovation and technology research & development, and the path of technology application aggregation. Technology innovation and technology research & development indicate that the innovation and research projects are carried out to solve the problems based on the demands of large-scale sports events, the economic society development, and the environment value increment and protection. The main task of innovation and research & development is to rationally promote the conduction and increase the value of the technology in the three aspects discussed above. The aggregation path of technology application is rationally setting up the aggregation path of the technology for sport and the basic technology and environment technology in holding the large-scale sports events, and the strategic path of technology development based on the construction of the technology chain and the industry cluster, which provides the fundamental guarantee for completing the specific task of the technology.

The second kind of path is the path of technology diffusion. The essence of establishing technology diffusion path is focused on how to make technology achievements transformed to the society, that is, how to set up the path of technology industrialization and technology publicity.

Setting up the path of technology industrialization is trying to spread the research results and technology achievements in the market mechanism, that is to say, the market tries to promote the industrialization and international development of the technology with large-scale sports events. At present, initiated by holding Beijing Olympic Games, there are many research results from some Chinese scholars on how to accelerate China's technology industrialization with the help of hosting the Olympic Games, among which a project(2006) the National Natural Science Fundation of China "The Impact on Technology Industry Development of Hi-tech Olympics" hosted by Professor Huang Lu-cheng is a representative research with its essential study about the possibilities of advancing technology industry development based on the Olympic market rules.

Setting up the path of technology publicity is aiming at spreading the research achievements through the aggregation by the government among the social public area, that is, technology is used to solve the problems of the people's livelihood, raise the people's science quality, and provide the service to the public. In recent years, that technology provides service to the people's livelihood has gradually become one of the hot topics that some of the Chinese scholars have paid attention to. Some scholars have proposed the idea of technology application in the people's livelihood, and positively give the suggestions to build the technology service system [12]. As we all know, the essence of technology creation and development is to solve the problems of human beings' own existence and development, and to ensure the harmonious development among human beings, nature and environment. There are still much more debates on some academic topics such as the technology value in the academic field, and the topics about the basis of technology value production, the connotation of technology value, the paths of the technology value achievement, and the ontological and practical value of technology are still the hot topics in the academic discussions, however, the idea of the function of technology in promoting the harmonious development among human beings, nature and environment has been widely accepted by all the sectors of the society. Therefore, a great many technology achievements aggregated through the large-scale sports events must be encouraged to spread in order to increase the value of technology in the process of conduction in the society. It is the successful hosting of Beijing Olympic Games and the rapid spread of a large number of technology achievements aggregated through the Olympic Games that set up the persuasive example in this respect. With the opportunity of hosting the Olympic Games, Beijing has reached the goal of the development of technology publicity with five strategies including strategies of digital technology, new construction technology, new material technology, environment protection technology, and biomedicine technology.

2. 2 The Structure of Technology Aggregation-Diffusion Model

In this paper, it presents the technology aggregation-diffusion model based on the study of the effects and path of technology aggregation-diffusion in large-scale sports events.

From the figure, we can illustrate the technology aggregation-diffusion model in the following three perspectives: (1) the technology driven by social and its own demands are brought together in three path through the market, the government, and the society to the field of large-scale sports events, (2) the aggregated technologies which form an intense technology aggregation area have made a lot of technology innovations at high levels, got many technology achievements, and brought the aggregation effect, (3) the technology achievements are spread in all the sectors of the society through technology

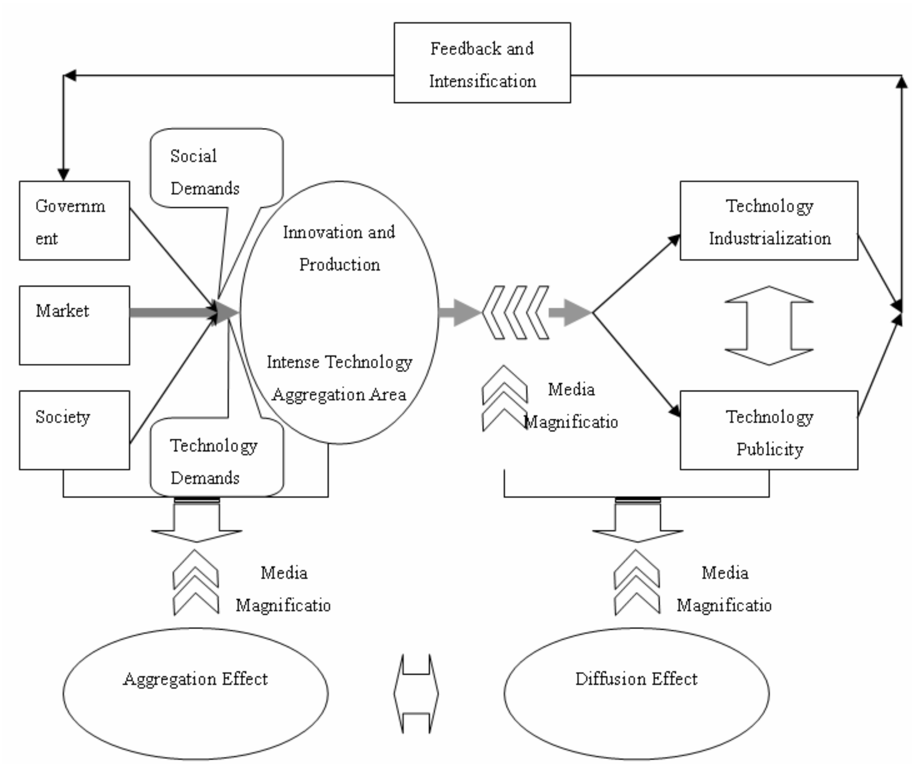


Fig. 1. Structure of Technology Aggregation-diffusion Model in Large-scale Sports Events

industrialization and technology publicity, which has brought an obvious diffusion effect. Generally speaking, the diffusion of technology achievements is still implemented through some more concrete path. For instance, the diffusion of industrialization of technology achievements in Beijing Olympic Games is carried out and spread in the path through industrialization of the technology in sports, industrialization of the technology infrastructure construction of the Olympic Games, industrialization of the environment technology applied in the Olympic Games, and the strategic technology industrialization formed by the technology industry chain and cluster [13]. While, the publicity diffusion of the technology achievements more reflects that technology provides the service to the social public benefits, which is proved by the " Technology achievements of the Olympic Games in the Field of Environment Protection and Food Security Seminar"[14].

It is obvious that modern media has played a significant role in the process of technology aggregation and diffusion. Since large-scale sports events have become one of the important social events deeply influencing the international society, the media's focus and reports on sports have objectively had the intense aggregation and diffusion effects which become more and more highlighted owing to the increasing magnification of the reports on every part and field of holding the sports events, and especially on the technologies applied in the events.

2.3 Technology Aggregation-Diffusion Mechanism

The study of technology development mechanism in large-scale sports events is prevalent in the studies of technology industrialization and international development caused by the process of China's promoting High-tech Olympic strategy [15]. For instance, Zhang Yong [16], Fang Fu-qian [17], as well as other scholars did the researches based on the life cycle theory, systematically analyzed the life cycle evolution of the technology cluster of the Olympic Games and its impact on the establishment of regional innovation system, and tried to explain the operation mechanism and fundamental rules of the development of technology industrial cluster in large-scale sports events.

These research results provide more information for our further study of the operation mechanism of technology aggregation-diffusion model. Considering the process of technology aggregation and diffusion as a process of the flow of value, it is easy for us to draw a conclusion that it is a fundamental model of technology value aggregation-diffusion development with the conduction and increment of technology value as its driving forces. In other words, technology aggregation and diffusion is a special process of conducting and increasing the technology value, for it is built in the basic structure of the value chain.

At present, there are numerous research results of the value chain. The analysis and the study of the value chain has been the hot topic since first described by Michael Porter, Professor of Harvard Business School, in his 1985 best-seller, the competitive advantage of nations [18]. The traditional theories about the value chain mostly focused on the study of an enterprise's production process, even if on the international value chain which still pay much attention to a enterprise's economic production, but they still provide a new analysis paradigm for the study of technology value chain. When doing the researches on technology value chain, scholars start their study with the discussion of the technology innovation process, which is because that technology innovation is always developed in the field of corporations, but the more important is that it is the driving force of the development of modern corporations.

Some scholars do a special study on technology value chain and propose the concept of it. Science-Technology Value Chain is the aggregation of the innovation chain linked by a series of independent and related innovations which will increase the technology developing value in the whole process of developing from the technology innovation source to the industrialization of technology achievements [19]. This concept is made from the industrial value creation point of view, but it is still suggestive owing to its analysis of the technology value chain as well as its creation process.

The implement of technology aggregation-diffusion development in large-scale sports events depends on the objectively existing intrinsic and extrinsic driving forces of conducting and adding the technology value. Unquestionably, technology carries the intrinsic and extrinsic (the use value) values. The intrinsic value is about technology's own value, which is created to complete some social work by the increasingly enrichment, advancement, and perfection of the technology system. While, the extrinsic value of technology is about the use value of the technology including its economic, social, and environmental values.

The two levels of value conduction are achieved in this mechanism, which are the increment of technology's own value and the use value carried by it. Based on its own demands and the social demands, technology is applied to provide relevant products to carry out the operation of the value chain, which will contribute to the realization of the two levels of conduction.

The adding of technology's own value is based on technology system's characteristic of self-reinforcement. After reviewing the technology system in history and present, we find that the fundamental reason why technology is the great economic driving force of the society is that it is always able to build a relevantly dependent and efficient self-operating system according to its tasks and goals to be completed and reached. Technology elements coordinate, support and depend on each other to promote the overall technology efficiency. Simply speaking, though the technology system is always initiated and maintained by human, an important element of keeping the system's self-operating efficiency is to eliminate the interference from the human factors as possible as we can, owing to which technology has gained the authority in almost the same model in all the sectors of the society. Technology also has the intrinsic demand of self value-adding in large-scale sports events. When we set up a technology system in the field of large-scale sports events, such as the judging system of sports by the optoelectronic technology, there is a continuous upgrading demand of the system generated by the technology system itself. As a result, the process of technology innovation has objectively become the fundamental measure of the system's self-improvement, by which technology value is increased.

The process of increasing the use value carried by technology shows that different social stakeholders have the right to ask for their own interest. (1) From the government point of view, we can see that the technology value must provide great support to the fast development of social health. In large-scale sports events, the government has fully realized that technology plays a significant role in raising the public service level, which is applicable in China as well as the countries with different political systems. So, the increment of technology value in the public has provided the guarantee to social development. (2) From the market point of view, we can know that the increasing technology innovations will bring enormous economic profits, ensure the successful operation of producing the value in the corporations and industry, and enhance the core competence of the corporations and industry, in which their interest plays an essential role. All the operation mechanisms are based on and worked by the interest control mechanism [20]. It is obvious that the technology value is finally achieved through technology industrialization, for the input of each node in the innovation value chain includes the financial, manpower, material, and information resources of technology. The output of the technology achievements in each node has different characteristics. If the output of the former node could not be used as the input of the coming one, the value of this node would not be achieved. From the value chain point of view, it shows that the value of innovation value chain will be achieved as long as the technology achievements of the end node are industrialized [21]. (3) From the society point of view, we can find that the demands of technology from the society is also based on the satisfaction of the increasing social development in order to ensure the sustainable development of human beings, which means that the increment of technology value contribute more to achieving the harmonious development among human being, nature and environment, which have more contributions on the value-adding of the humanistic value.

The demands of increment of the intrinsic value of the technology and the its carried use value become the intrinsic driving force to promote the technology aggregation and diffusion, and construct the intrinsic operation mechanism of the technology aggregation-diffusion model in large-scale sports events.

3 Government Responsibilities

Large-scale sports event is a complicated and systematical project involving all the sectors of the society, such as politics, society, economy, and culture. Meanwhile, there are the top demands of city infrastructure and other public products and services when holding the large-scale sports events [22]. Obviously, it is the government that has the capability to supply the public products. Therefore, government plays a leading role and has a dominant function in the management system of large-scale sports events.

The technology aggregation and diffusion in large-scale sports events are made up of a series of technology innovations which require the necessary situations of the society and the technology, and also have the vivid feature of the public products. Besides, it is the government that can complete the aggregation of technologies from every sector of the society. Thereby, the establishment and operation of technology aggregation-diffusion model in large-scale sports events need to be completed by the lead of the government as well as the active participation of all parts of the society.

(1) The government has the responsibilities to take some necessary measures to set up the technology aggregation-diffusion model in large-scale sports events and positively promote the operation of this model, which will guarantee the successful conduction and increment of the technology value in the events. It should actively integrate the technologies of all the sectors of the society and ensure the smooth operation of this model so as to achieve the maximum increment of the technology value, which is to reach the maximum increment of the values of the society, the economy, and the environment.

With the opportunity of holding large-scale sports events, the government should set up the fundamental mechanism of aggregating the technologies in such short time while making technology innovations in a large scale; meanwhile, it should build the fundamental model of promoting the industrialization and publicity of technology achievements, encourage the market and the society to spread these achievements and increase the technology value.

(2) The government should play a leading role in the aspects of coordination, policy making, and public services supplying in the light of polices required by the technology aggregation-diffusion model. It should build and improve the mechanism of establishing the large-scale sports events organizations, making polices, and constructing the service platform in the process of promoting the development of technology publicity.

(3) The establishment and operation of technology aggregation-diffusion model also have certain social risks which are caused by the operation of the technology system, such as the chaos of the game caused by the system faults. This kind of risk must be controlled by the government, for a huge and complex risk of social technology system cannot be controlled and solved by some social organizations and enterprise groups.

4 Conclusion

Since the technology aggregation and diffusion in large-scale sports events play a more and more important role in the development of the society, and large-scale sports events are more and more frequently held in China, we should pay more attention to the study of the feasibility, the implementing path, the government responsibilities and its policy protection to widely build the technology aggregation-diffusion model in the large-scale sports events held by the government, what's more, put forward the relevant suggestions of making decisions, which has the strategic significance of increasing the economic and social benefits for the countries through holding the large-scale sports events.

Acknowledgment

This project is support by National Natural Science Foundation of China (70679001).

References

1. Dong, C., Yu, X., Wang, J.: Inventory of Beijing Olympic Games: Value for the Hi-tech Olympics from National Strategy. Journal of Shenyang Sport University 28(1) (2009)
2. Dong, J., Liu, X., Song, L.: Risk Management of Emergency for Beijing 2008 Olympic Games. Sports & Science 26(1) (2005)
3. Tang, X.: The Spread Effect of Large Scale International Sports Competition Events on social development. Journal of Guangzhou Sport University 27(1) (2007)
4. Yao, S.: Strategic Research on Shanghai Holding International Major Sports Competitions. Journal of Shanghai Physical Education Institute 28(5) (2004)
5. Huang, H., Zhang, L., Li, N.: Study on the Role of Shanghai Government in the Operation of Major Sports Competition Events. China Sport Science 27(2) (2007)
6. Du, L.: Hi-tech Olympics and Technological Needs. Science & Technology Industry of China 8 (2002)
7. McDowell, A.: The Technology Behind the Torino Olympics, `http://www.shortflip.com/article/The-Technology-behind-the-Torino-Olympics-160.html` (last accessed, March 2010)
8. Rogers, E.M.: Diffusion of Innovations. Free Press, New York (1996)
9. Metcalfe, J., Georghiou, L.: Equilibrium and Evolutionary Foundations of Technology Policy. CRIC Discussion Paper 3, 1–31 (1997)
10. Comin, D.A., Hobijn, B.: An Exploration of Technology Diffusion. Journal of Technology Transfer 33(2), 187–207 (2008)

11. Keller, W.: International Technology Diffusion. Journal of Economic Literature 42(3), 752–782 (2004)
12. Zhou, Y.: Enhance Development of Livelihood Technology in China. Forum on Science and Technology in China 1 (2008)
13. Dong, C.: Study on Beijing Hi-tech Olympics Pattern Based on STS Theory. China Sport Science 28(6) (2008)
14. Liu, W.: Protect to Environment of Hi-tech Olympics. Science Foundation of China (2007)
15. Dong, C., Xing, H., Wang, J.: Three Key Questions to Hi-tech Olympics Impel Our Country Science and Technology Industrial Production. Studies in Science of Science 2 (2007)
16. Zhang, Y.: Research on Olympic High Technology Cluster Innovation Dynamic Mechanism based on Ecology. Beijing University of Technology, Beijing (2008)
17. Fang, F., Li, X.: The Characteristics of Scientific and Technological Cluster of Olympics and the Systematic Circle. Social Science of Beijing 3 (2008)
18. Porter, M.E.: The Competitive Advantage of Nations. Free Press, USA (1985)
19. Huang, G., Xu, J., Li: Science-Technology Value Chains and the Linking Models among Multiple Innovation Entities. China Soft Science 6 (2006)
20. Luo, S., Zhang, S.: A Discussion about Technological Progress Operating Mechanism. Economic Survey 1 (2003)
21. Zhu, H., Mo, Y., Zhou, X.: Analysis on the Mechanism of Technology Intermediation to Enhance Efficiency of the Chain Value. Science & Technology Progress and Policy 9 (2006)

Specification Patent Management for Web Application Platform Ecosystem

Yoshiaki Fukami[1], Masao Isshiki[1], and Hideaki Takeda[2],
Ikki Ohmukai[2], and Jiro Kokuryo[3]

[1] Graduate School of Media and Governance, Keio University,
5322 Endo Fujisawa-city, Kanagawa, Japan,
yofukami@sfc.keio.ac.jp,
http://www.sfc.keio.ac.jp/en/top.html
[2] National Institute of Informatics, 2-1-2 Hitotsubashi Chiyoda-ward, Tokyo, Japan
[3] Faculty of Policy Management, Keio University,
5322 Endo Fujisawa-city, Kanagawa, Japan

Abstract. Diversified usage of web applications has encouraged disintegration of web platform into management of identification and applications. Users make use of various kinds of data linked to their identity with multiple applications on certain social web platforms such as Facebook or MySpace. There has emerged competition among web application platforms. Platformers can design relationship with developers by controlling patent of their own specification and adopt open technologies developed external organizations. Platformers choose a way to open according to feature of the specification and their position. Patent management of specification come to be a key success factor to build competitive web application platforms. Each way to attract external developers such as standardization, open source has not discussed and analyzed all together.

Keywords: Web Application Platform, Patent Management, Standardization, Open Source, Grass Roots Activity.

1 Background

1.1 Role of Platformers

Role of the web has changed from a medium for share documents to platform for running applications like webmail, scheduler and other services.

Web service providers have tried to supply many applications to meet various kinds of needs. There are not so many players that can develop many kinds of applications. Companies with great amount of resources cannot develop all applications meeting demand of all users due to rapid progress of technology.

Web applications are used from diversified environment and process various kinds of data. Web service providers are made to develop platform where 1) many attractive applications run, 2) multiple kinds of data are exchanged among applications and 3) massive social contents and data can be generated and processed under adequate access control.

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 269–280, 2010.

The simplest way of realizing competitive platforms is to do it by yourself completely. Ways to add application are developed by themselves or purchase services. Applications developed and managed by themselves can be run with highly integrated way. You can solve interoperability issue by determining proprietary format for exchanging resources among your own applications.

Nonetheless, service providers are facing competition of providing more applications. It is difficult to develop various kinds of applications by myself because each application gets to be complicated. Thus service providers have to collaborate third party application developers and to make platforms attractive.

At the same time, resources for developing applications are restricted within internal manpower and accumulated technology and know-how within an organization. Nowadays, various kinds of new applications have launched one after another. There has come to be needs for utilizing outside resources to broaden breadth of application assortment.

There are two alternatives to make use of outside resources. One is purchasing technology, source code or organization (M&A). The other is opening specification of platforms (API: application programming interface) to enable third-party developers access database and upload social contents. Facebook has opened APIs named Facebook Platform in 2007, and other platformers have taken measures to encourage third-party developers to supply applications compliant for specification of platforms.

Serving web applications has come to be disintegrated into two layers, one is platformers which provide identifier and access control. Another is application provider. Platformers are facing competitions to provide wide range of applications which meet users' need.

1.2 Relationship Design with Patent Management

In this paper, we would like to discuss about strategies of web application platforms, especially making proprietary specifications opened, standardized, and open source to attract third party developers and build competitive platforms.

Patent management is key factor for building strategy to create competitive web application platforms.

Platformers can design relationship with developers by controlling patent of their own specification and adopt open technologies developed external organizations. It is impossible to share revenue with developers because most applications are supplied at no charge and there are no complementary goods. All platformers can do are to endow legitimacy with their specifications to encourage developers decide to devote R&D resources to the one.

Platformers choose a way to open according to feature of the specificationand their position. Patent management of specification comes to be a key success factor to build competitive web application platforms. Each way to attract external developers such as standardization, open source has not discussed and analyzed all together.

Patent management is key factor for building strategy to create competitive web application platforms. Pattern of collaboration with application suppliers is determined by patent policy, especially degree of openness and utilizing standardization.

This paper is organized as follows. Section 2 reviews structure of web application platforms. Section 3 analyses cases of ecosystem development with patent management.

Section 4 emerging players of web application platform business. Section 5 reviews Section 6 concludes and gives a brief outlook on further research.

2 Web Application Platforms

2.1 Structure of Web Application Platforms

If organizations cannot choose stand-alone strategy, they have to develop ways to corporate with others and connect applications outside. As the characteristic to be developed for sharing data, platforms on the web must be designed to attract many developers for increasing application and supply great variety of utilities for end users. As there is more than one competitor, it needs to attract application developers to own platform. Design of connection determines competing power to attract developers and aggregate attractive applications.

The role of platform is to enable mutual exploitation of data resources among each application. You have to design interoperability of three elements, data resources, authentication and authorization for building connection with application developed by others. Combination of design for these three elements affects affinity of platforms.

Interoperability of Data Resources. Most users use several applications according to type of resources. When you compose short text, you might post it to twitter. When you take a picture, you might make it be shared on Flickr.

It is needed to select all resources you have saved on different applications for integrating all resources. Thus, to develop applications that makes use of whole of data resources of other distributed applications, it is critical that interoperability of data resources is realized.

To just sort and layout resources from various web applications, RSS/Atom feed is enough. RSS/ Atom is XML based format. Most blog services, photo/video sharing services and twitter publish RSS or Atom feeds and RSS reader, RSS/ Atom feed aggregator is widely used for checking new arrival blog entries and articles. You can aggregate activities of friends to create timeline streams with services like FriendFeed, which collect distributed RSS/ Atom feeds from various services.

However, RSS/ Atom has only limited vocabulary/ attributes/ elements such as time, language, pubdata (published date time) (1). It is impossible to handle various kinds of resources without specification that identify data type and supply context information. Last.fm recommends artists, albums and friends whose preference of music is similar to you on the basis of log of playing music files.

If you develop recommendation service of movie, you have to calculate history of watching movies and reviews among various kinds of web applications such as blog, review sites like IMDb and others. Each logdata must be attached with metadata representing type of data and context information. More over common specification of data format is needed to make use of distributed data resources.

Social function provider. Identification is important factor of web application. When you upload a memorandum about the day as a text file to a server you manage, nobody but who know the file is uploaded by you only recognize that the contents of file is about your life. Even there are full of data resources generated by activities in everyday life, they cannot be used collectively without connection to the individuals generating data themselves. Social contents must be connected to individuals.

Most web users tend to store data in several platforms according to the attribute of each data. One uses Typepad for storing sentences. She post photo data to Flickr and record track data of listening music on the server of Last.fm at the same time. It needs to integrate resources with common identification.

Even if common identifying data is attached to every data resource, no one can utilize resources stored in other applications without access and extract adequate data. Thus, there must be specification for transaction of data resources and access control among multiple applications.

Highly detailed access control and standardized machine-readable formats for resources are necessary to utilizing information of relationship among friends (in other words, social graph) and various types of data resources together.

Once contents are generated with linking to IDs and the web has come to be platform not only for sharing information but also collaboration with generating contents, there have emerged the importance of access control and indicating authors or owners of contents. Portals, search engine venders and other service providers on the web have followed SNSs to provide social functions.

Most services on the web come to provide two functions. One is identification and access control based on identities. Another is platform for applications. Most SNSs and portals had provided these services integratedly by themselves. DeNA and GREE, Japanese mobile SNS platformers, still develop games, avatars, and other contents by themselves.

On the other hand, some SNS begun to make use of third party contents and service providers. Facebook has opened APIs named Facebook Platform with introducing 65 prevailing developer partners including Amazon.com, Digg and Twitter (2). Plenty of applications supplied by many third parties have increased Facebook users. Diffusion of web applications has encouraged value chain of services on the web disintegrated.

2.2 No Physical Complementary Goods

Choice of patent policy for specification controls relationship between platformers and outside developers. Thus there are activities of developing new or extended standards of format for exchanging metadata of data resources. Competition among web application platforms has some distinctive features from other existing platform competition.

Secondly, opposite to recording media specification such as optical disc (Bluray/ HD-DVD) and magnetic tape (VHS/ Betamax), there is no complementary goods. Web application platforms and compliant applications can run on most browsers conforming to open standards. It is no need to purchase some goods or proprietary software to make use of such applications. More over, most web applications are served for free.

Thirdly, interoperability among applications is strictly needed. The web is used as lifelog platform and collaboration. Concept of lifelog encourages integrating whole data generated by various applications. Each application must be designed to provide data with standardized protocol and format.

There are many researches about competition of ITC market. However, most of them (e.g. (3)) is about physical goods. There are many differences between development of web application and hardware like optical disks. It is much easier for small organizations, even individuals to build platforms which take large share because there is little need to investment for production facilities and cost for expanding scalability is getting much smaller with usage of crowd computing infrastructure.

2.3 Two-Sided Market

If organizations cannot choose stand-alone strategy, they have to develop ways to corporate with others and connect applications outside. As the characteristic to be developed for sharing data, platforms on the web must be designed to attract many developers for increasing application and supply great variety of utilities for end users. As there is more than one competitor, it needs to attract application developers to own platform. Design of connection determines competing power to attract developers and aggregate attractive applications. Competition of Web application platform market is a typical two-sided market [1].

2.4 Building Magnetism with Outside Resources

If platformers do not succeed to dominate enough shares to attract third party developers by themselves, there are two alternatives. One is to design the platform for niche. Another is to adopt open specification. There are some open specifications and each specification is different in patent policy.

3 Case Analysis of Developing Ecosystem

3.1 Enclosure with Proprietary Specification

The simplest way is to accumulate many users. Platformers, which has dominant share, keep their specification proprietary. Facebook has begun opening her platform to third party earlier than competitors. However, Facebook platform is still proprietary.

One way of embarking for developing applications is opening APIs and sharing proprietary protocol for connecting platforms. Facebook has opened APIs named Facebook Platform with introducing 65 prevailing developer partners including Amazon.com, Digg and Twitter (2). Plenty of applications supplied by many third parties have increased Facebook users. Facebook, a player with high competency has succeeded to assemble variety of applications developed by not only e-commerce giant but also small and agile start-ups. Lineup of application and registration with autonym have come to be great appeal to end users, and amount of users have attracted developers. Facebook have developed circle of reinforcing their platform. Facebook also open Facebook connect, protocol for third party developers to utilize social graph of users.

More typical case of this strategy is Twitter. Stance of Twitter in platform development has been concentrated in sharing just 140 characters. However, you can

---

[1] One of analyses about two or multisided market is (4).

post not pictures with body text. Twitpic[2] and similar services offer users function to post pictures with text based tweets by storing picture files and posting URIs of them. Most Twitter client software enable users to check body texts and pictures posted with them at once. Service and software supplied by third parties add functions to Twitter, single-function platform, and popularity is build by resources of outside, in other words, adopting ”proudly found elsewhere” strategy (5).

3.2 Industrial Consortia

Nondominant organizations tend to choose strategy to adopt open specification and there are some ways to open. One is to make proprietary specification be standardized.

For more sensitive function such as authentication, there are some cases that specification developed by industrial consortia. OpenID (6) is decentralized standard for authentication. Users can log in multiple services with one identity of services on the web. OpenID makes use of URI as key to identify individuals.

Independent engineers who pursue realization of distributed social web launched OpenID project. Distributed Social web means web as a platform, which consists of applications provided multiple organizations and individuals and identity, identifier, and data, linked to identifiers are not aggregated at one certain service provider.

OpenID foundation, non-profit organization formed by individual developers welcomed companies such as Facebook, Google, Microsoft, PayPal and other big companies as corporate members later. They have adopted OpenID as authentication technology.

While Facebook continue to make effort for diffusion of proprietary APIs called Facebook connect, Facebook work together with competitors to diffusion of common protocol for sharing users’ private resources over multiple platforms. Facebook have developed Facebook connect protocol, and also adopt OAuth (7), protocol of authorization.

Same as Facebook connect; many competitors have developed proprietary protocol for authorization such as FriendConnect (Google), and MySpaceID. However, Google and MySpace have made their protocol support OAuth protocol later. Most platformers adopt common and open protocol of authentication and authorization. Thus, web application platformers may adopt different strategy for each function and layer.

Engineers who worked for different companies and tried to implement authorization function to each services launched OAuth project. The activities are not conveyed under the command of corporate will. OAuth working group was established in Internet Engineering Task Force two years later than the launch of the project (8).

To process whole distributed and various kinds of data resources, protocol for authentication and authorization is not enough. There must be implementation of common specification to define what each data is. Activity Streams is emergent specification to complement this layer. Activity Streams[3], data format and specification of extension from Atom. You can attach context information to resources with Activity

---

[2] Twitpic: `http://twitpic.com/`

[3] Activity Streams website: `http://activitystrea.ms/`

Streams and applications implemented recognizing the format can process them easily and accurately. Some large platformers including Facebook, My Space and Windows Live provide their originally formatted APIs in parallel.

OpenID, OAuth and Activity Streams are developed from activities of DiSo project (distributed social network project[4]). Activities of DiSo project are moved forward by communication on mailing list, wiki, Google groups and face-to-face meetings. Members live dispersed places and work for scattered organizations. Each member makes effort to sophisticate specification at each respective organizations and communities.

Another example of grass-rooted specification/ standard is Microformats (9), an implementation of semantic markup rule. You can export information about events (hCalender), location (geo or adr) from web sites implemented Microformats-compliant markup and import them to applications including one served via the Web. Resources that Microformats are embedded can be treated with recognizing attribution.

Microformats is utilized by not only services on the Web but browsers installed in client hardware. Operator is an add-on package for Firefox to extract data resources marked up with microformats and other lightweight semantic specifications. You can import from event schedule notified on a blog entries to scheduler software such as Thunderbird, Microsoft Outlook and iCal of MacOS.

Microformats is developed by activities of microformats.org[5] , a grass-roots project like DiSo project. Grass-roots activities play important roles of developing web application platforms.

3.3 Standardization

Another way to enhance legitimacy to adopting specification is standardization. Once a specification comes to be regarded as standard, most developers tend to attempt to adopt it.

HTML is changing from just a mark-up language for composing web sites to one of technological layers of web application platform. Development of HTML5, forthcoming version of HTML was lead by industrial consortia named WHATWG (The Web Hypertext Application Technology Working Group), which consists of major browser vendors like Apple, Mozilla and Opera at first. They founded WHATWG and protest to re-start development of new version HTML because they need to make HTML advanced enough to be open platform of web applications. World Wide Web Consortium (W3C), international consortia lead by academic persons, and pursuing development of xHTML accepted WHATWG's proposal and launch process for development of new version HTML.

Some specs of HTML5 are developed on the basis of proprietary specification of Google. Google announced stopping development of Google Gears, the environment to run web applications while offline, which are adopted as part of HTML5 specification named web sockets and web storage. Most web browser will implement web sockets and web storage to conform to HTML5 standard.

---

[4] DiSo project: `http://diso-project.org/`

[5] Microformat.org: `http://microformats.org/`

This means that offline function for Google's services, Gmail and Google Reader, is available on most machines that any web browsers conforming to the standard are installed.

3.4 Making Specification Open-Sourced

Facebook connect is specification which is opened but still be proprietary. On the other hand, some platformers have made their own specification be open source. To make source code open and free can attract developers and end users who tend to avoid being locked in proprietary specification. There are some examples that software is made to be open source to compete with category giant.

Android is OS for mobile phone developed by Google. Android used to be a proprietary OS of Android inc. Google acquired Android inc. and establish Open Handset Alliance (OHA) with hardware venders, mobile network operators and semiconductor venders. OHA released Android OS as open source software and HTC and other member of OHA have sold mobile phone handsets, which Android is installed. Google provide software developers kit without charge.

Open source has two advantages. One is that it is much easier to diffuse platform widely because devices can be distributed with lower cost. Another advantage is easy to attract application developers because they can implement various kinds of functions with usage of opened specification.

There are some cases that proprietary software is made to be open source. OpenOffice.org is office suite competing with Microsoft office. OpenOffice.org used to be proprietary software. Sun Microsystems purchased the software and open source code to compete with Microsoft. OpenOffice.org have been used by public sector, which want to avoid using proprietary software widely.

America online made source code of Netscape browser open. Mozilla Firefox, open source browser build with source code of opened Netscape achieves more share and compete with Microsoft's Internet Explorer. Making proprietary software be open source is effective way to increase users and suppliers in some cases.

Suppliers do not develop applications for platforms, which is merely open. Platforms cannot attract developers without sustainable support for development and realization of interoperability. The more platformers adopt the specification, the more users applications target with same costs. To build trust from developers, it is important to endorse specification to be opened and widely used. Thus, to make proprietary specification open is not enough. One way is adopt common specification developed by large but neutral company, which can endorse sustainable maintenances and promote with much costs.

OpenSocial is a specification developed by Google. Google manages her own SNS platform named Orkut, however Orkut do not succeed gathering many users in North America, Europe and East Asia. Some competitors of Facebook, including MySpace and Japanese SNS Mixi, have opened their own APIs. However, most of them have changed their strategy of opening proprietary APIs to adopt Google's OpenSocial. OpenSocial is now Open-sourced. but not just open source. Open source endorsed by giant company, Google.

MySpace, LinkedIn, Mixi, Cyworld and other platformers'[6] . While Facebook had gathered big-name partners, Google and adopters promote this common API set to individual developers and start-ups. Google and partners established OpenSocial Foundation and provide information of specification and techniques. Some evangelists from Google, Mixi, NTT Resonant and others established developers' community named OpenSocial Japan. Evangelists have given presentation at some developers 'conferences and presented programming competitions named hackathon (hack + marathon).

4 Emerging Players: Power of Community

Diffusion of ITC products has been accelerated by standardized specifications. Most actors who develop standard have been international standard organization, such as International Standard Organization (ISO) and Internationa Telecommunication Union (ITU), or commercial enterprises. Developing common infrastructure and produce high tech products cost so great deal of money that only national organization and industrial giants can fulfill a role to develop and diffuse platform with advanced technologies.

It takes some years to diffuse specifications for physical infrastructure because it takes at least a few years to develop instruments and devices, which consist of infrastructure, such as switching equipment and Blu-ray disk recorder. However, it takes much shorter time to develop and diffuse web applications because you can develop services without materials and many types of equipment. All you have to do for development is to code with personal computers and upload it to servers. Thus, more and more individuals and grass-rooted community contribute for development and diffusion of specification for web application platforms.

World famous engineers some of whom work for platformers developed OpenID and OAuth and others do not belong to any organizations. Engineers from distributed organizations work together to establish non-proprietary specifications and establish foundations. OpenID foundation consists of not only independent engineers but also major platformers like Google and Facebook. Some engineers frequently change their affiliations and build real social network among engineers that have worked as basement of open specifications. There are many cases of open specification developed by grass-roots organizations.

The role of international consortia such as World Wide Web Consortium (W3C) and Internet Engineering Taskforce (IETF) as standard development organization (SDO) has got to be more important especially in ITC technological development. These consortia play a role of round table of discussion among enterprise giants, public institutions and independent individuals.

Fig.1 and shows there were not so many players in ITC standardization process. However, there come to emerge much new type of stakeholders (Fig.2). Standard Development Organizations differentiate into public organization like ISO (International Standard Organization) and industrial consortia. You cannot ignore

[6] You can see list of OpenSocial adaptor in the official site of OpenSocial: `http: //wiki.opensocial.org/index.php?title=Main Page#Container Information`

activities of individuals and grass roots communities. It gets more and more important to develop good relationship with stakeholders to attract platforms.

5 Conclusions and Further Research

There comes to be more and more scenes to use the web in daily life. Various kinds of web applications are used and users integrate data from various kinds of applications. It is more practical to build platforms to integrate various kinds of data resources from web applications than to develop whole kinds of applications at their own. Competency of managing web application platform is determined by relationship with third-party developers of applications.

You can recognize attitude for building ecosystems with third-party developers with protocol, APIs and other specification for applications. There are two elements of defining relationship with third-party developers. One is how to collaborate with external players. Another is patent policy. To attract third-party developers, platformers have to characterize their specification as open and flexibility. It gets to be important to keep good relation with third party developers. It is effective to make use of open and specifications of grass-root community and standardization for sending messages that they continue to promise keeping the platforms open and offering interoperability.

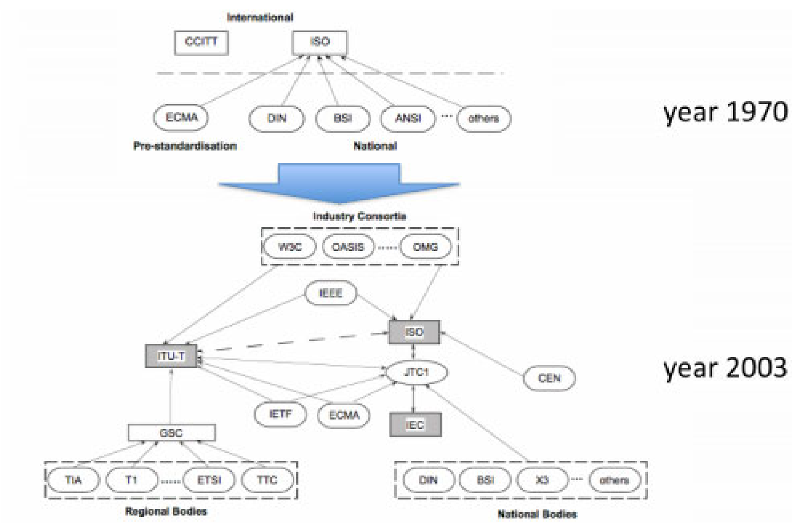


Fig. 1. The IT standardization universe in 1970 and 2003 (10)

Strategies of vertical integration with open API, open source community and standardization have still been analyzed repeatedly. Although most researches are limited in organization of large enterprises and public sector, many web applications have been developed and offered by individuals and small ventures. Platform developers must attract individuals and make them eagerly supporters of the their

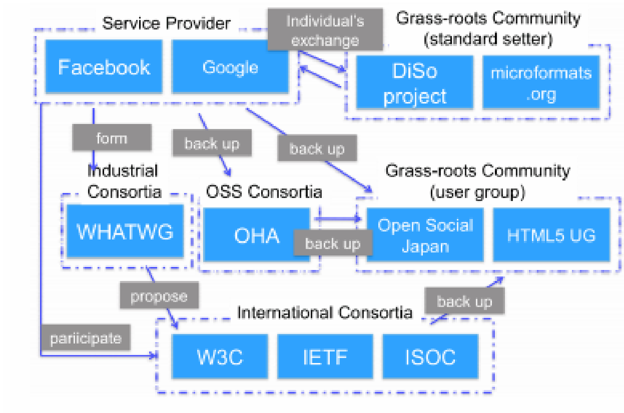


Fig. 2. Present stakeholders related to standardization

platform. What platform developers have to do are not only build high-performance infrastructure but also share profit. Strategy to develop attractive platform have been mixture of these three elements.

In this paper, we focused strategy of developing web application platforms. There are three strategies to develop web application platforms attracting application developers, 1) enclosure with proprietary specification, 2) building magnetism with out side resources with standardization and 3) open source activities. Activities of individual engineers and grass-roots organizations increase variety of types in process to develop and diffuse specifications.

Further research will be twofold. The first direction will be model construction of strategy combination. We would like to analyze the effect of these three ways precisely and develop the model for designing community for developing and enhancing worth of the platform. Especially we would like to build model of strategy of adopting appropriate ways in specific field of technology and put them together in the most suitable combination.

References

1. R.S.S.A. Board: RSS 2.0 Specification. RSS at Harvard Law 30 (2005)
2. Facebook: Facebook Platform Launches with 65 Developer Partners and Over 85 Applications for Facebook (2007), `http://www.facebook.com/press/releases.php?p=1319`
3. Riefler, B.: The Composition of Working Groups in Industry-Specific Standardization Organization. In: Discussion Paper Series from the Institute of Economics and Law, University of Stuttgart (2008)
4. Eisenmann, T., Parker, G., Van Alstyne, M.: Strategies for two-sided markets. Harvard Business Review 84(10), 92 (2006)
5. Huston, L., Sakkab, N.: Connect and develop. Harvard Business Review 84(3), 58–66 (2006)

6. Recordon, D., Reed, D.: OpenID 2.0: a platform for user-centric identity management. In: Proceedings of the Second ACM Workshop on Digital Identity Management, pp. 11–16. ACM, New York (2006)
7. OAuth Core Workgroup: OAuth Core 1.0 (2007), `http://oauth.net/core/1.0a`
8. Hammer-Lahav, E.: Explaining OAuth (2007), `http://hueniverse.com/2007/09/explaining-oauth/`
9. Khare, R., Celik, T.: Microformats: a pragmatic path to the semantic web. In: Proceedings of the 15th International Conference on World Wide Web, pp. 865–866. ACM, New York (2006)
10. Jakobs, K.: Information technology standards, standards setting and standards research. In: Proceedings of Cotswolds Conference, Stanhope Center for Communication Policy Research, pp. 8–9 (2003)

Evaluation of the Acceptance of Audience Response System by Corporations Using the Technology Acceptance Model

Hsing-Hui Chu, Ta-Jung Lu, and Jong-Wen Wann

Graduate Institute of Technology & Innovation Management
National Chung Hsing University, 250 Kuo Kuang Rd., Taichung 402, Taiwan R.O.C.

Abstract. The purpose of this research is to explore enterprises' acceptance of Audience Response System (ARS) using Technology Acceptance Model (TAM). The findings show that (1) IT characteristics and facilitating conditions could be external variables of TAM. (2) The degree of E-business has positive significant correlation with behavioral intention of employees. (3) TAM is a good model to predict and explain IT acceptance. (4) Demographic variables, industry and firm characteristics have no significant correlation with ARS acceptance. The results provide useful information to managers and ARS providers that (1) ARS providers should focus more on creating different usages to enhance interactivity and employees' using intention. (2) Managers should pay attention to build sound internal facilitating conditions for introducing IT. (3) According to the degree of E-business, managers should set up strategic stages of introducing IT. (4) Providers should increase product promotion and also leverage academic and government to promote ARS.

Keywords: Audience Response System (ARS), Technology Acceptance Model, User Acceptance.

1 Introduction

The rapid change of advancement in information technology (IT), globalization, innovation and knowledge-based economy in recent years has resulted in more harsh and intense challenges to enterprises. . Thus, enterprises have to be more innovative to overcome difficulties for sustainable growth and development. Today enterprises are facing one of the issues that their business thinking is too conservative, together with insufficient IT adoption, these are not conducive to business innovation and business model [1]. The integration of knowledge-based economy and IT has become indispensable condition for enterprises to maintain competitive advantages and also with the competitiveness of countries [2].

In pursuit of better learning outcomes, universities have invested tremendous resources to analyze the function of learning and methods to enhance the learning process. It is beneficial for industry trainers to study trends in higher education and employ strategies that have been proven effective. One method for improving learning outcomes that has been successfully employed in lecture halls is the use of

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 281–291, 2010.

audience response systems (ARS) [3]. ARS is one of the innovative IT which has been widely used in Europe and the United States by both educational and commercial organizations and has attracted broad academic discussion. Organizations that have adopted ARS technology include Boeing, Academy of the US Federal Bureau of Investigation, IBM, John Deere, McGraw-Hill, National Academy Foundation, Prentice-Hall, Raytheon, Toys ‘R’ Us, United States Army and Navy, Walt Disney World, and YMCA [4].

In Taiwan, ARS is used in more than 300 schools but is adopted by only a few private enterprises. The low usage rate of ARS by Taiwanese enterprises is markedly different from that of major organizations in other countries which prompted us to study this difference in the acceptance of ARS.

2 Objective

This study aims to explore the acceptance of ARS in Taiwanese enterprises. From literature review of ARS, related studies and theories of technology acceptance model (TAM), theoretical basis and factors which may influence user’s perception of usefulness and ease of use to understand the acceptance of ARS were chosen. The following research objectives will be achieved through data collection, questionnaire delivery and data analysis.

(1) What characteristics of ARS and facilitating conditions will affect the "perceived ease of use" and "perceived usefulness", and “user’s intention”?
(2) Does the degree of E-business will affect the "perceived ease of use" and "perceived usefulness", and “user’s intention”?
(3) To interpret Taiwanese enterprises’ acceptance of ARS.
(4) To find out what external factors will affect the acceptance of ARS.
(5) To promote and help enterprises to introduce ARS successfully.

3 Literature Review

3.1 Information Technology (IT)

IT is composed of several related parts. One includes techniques for processing large amount of information rapidly, and it is epitomized by the high-speed computer. A second part centers around the application of statistical and mathematical methods to decision-making which is represented by techniques like mathematical programming, and by methodologies like operations research. A third part is in the offing, though its applications have not yet emerged very clearly; it consists of the simulation of higher-order thinking through computer programs [5].

3.2 Audience Response System

ARS is an electronic device designed to allow immediate interaction between an individual presenter and a large audience. An ARS typically consists a remote control that audience members use to respond to questions and an electronic receiver that records, and optionally, displays individuals' responses. ARS allows for a large number of individuals to respond simultaneously. Each individual response is recorded by the hub and can be displayed via projector or exported as a data file for use in other software [4]. LaRose [3] divided ARS into four components: a handheld transmitter, a receiver (plugged into the USB port of the presentation computer), the software on the presentation computer and a projection system. For example, in a presentation, presenter poses a question which including four answers (A, B, C, D), audiences could use handheld transmitter to press button to submit individual response, via infrared (IF) or radio frequency (RF) to send individual response to signal receiver (presenter's computer) [6]. ARS software will count numbers of each answer and show results in formats like pie or bar charts. The name of this system differs in different countries and suppliers. Names other than ARS are shown below:

(1) Classroom Response System (CRS) [7]
(2) Classroom Communication System (CCS) [8]
(3) Interactive Response System (IRS) [9]
(4) Classroom Performance System (CPS) [10]
(5) Personal Response System (PRS) [11]
(6) Student Response System (SRS) [12]
(7) EduClick [13]
(8) Clicker [14]

Table 1. The difference between ARS and traditional teaching method

ARS	Traditional teaching method
Interact with large audience, to provide equal opportunities for interaction	Only interact with the minority
Understand audience learning performance immediately during instruction	Understand learning performance after instruction
Audience response can be further analysis in chart	Cannot show response through chart in real time
Built in multiple usage	Vary according to instructor
Anonymous functions can let audience feel easy to answer	Respondents may have pressure to answer when system doesn't have anonymous functions
Responses can save response directly through system	Responses need to be recorded manually

3.3 Theories of Technology Acceptance Model

(1) Theory of Reasoned Action (TRA)
TRA was very general model which is used to explain almost any human behavior. TRA proposes that intention solely and directly influences the adoption behavior and that intention is determined by factors of attitudes toward behavior and subjective norms toward behavior [15].
(2) Technology Acceptance Model (TAM)
TAM derived from TRA and developed by Davis. It is used to predict adoption and usage of technology in information systems and

organizational contexts. TAM supposed that perceived ease of use and usefulness are major factors that influence the acceptance of a technology [16].

(3) Theory of Planned Behavior (TPB)
TPB is an extension model of TRA, except original two factors (attitudes toward behavior, subjective norms) originated from TRA, the other is perceived behavioral control. TPB supposed that not all behavior may be under an individual's volitional control, so behavioral control as an important factor could be influential on behaviors [17].

(4) Technology Acceptance Model 2 (TAM2)
TAM2 is an extension model of TAM, which incorporates two additional theoretical constructs: cognitive instrumental processes and social influence processes. Four cognitive factors including job relevance, output quality, result demonstrability, and perceived ease of use influence perceived usefulness. Three social forces influence perceived usefulness: subjective norm, image, and voluntariness [18].

(5) Unified Theory of Acceptance and Use of Technology Model (UTAUT):
Venkatesh et al. combined eight models (Diffusion of Innovation, Theory of Reasoned Action, Theory of Planned Action, Technology Acceptance Model, Combined TAM and TPB, Motivational Model, Social Cognitive Theory, Model of PC Utilization) develop UTAUT which and proposed four factors (performance expectance, effort expectancy, social influence and facilitating conditions) and four moderating variables (gender, age, experience and voluntariness of use) [19].

Table 2. Comparison of TAM models

	TRA	TAM	TPB	TAM2	UTAUT
Author	Fishbein & Ajzen	Davis	Ajzen	Davis & Venkatesh	Venkatesh, Morris, Davis & Davis
Year	1975	1989	1991	2000	2003
Appropriate situation	Volitional action, use to explain human behavior	Not all control by volition, use to explain human behavior	IT acceptance behaviors	IT acceptance behaviors	IT acceptance behaviors
Factors affect behavioral intention	Attitude toward the behavior, Subjective norm	Perceived usefulness, Attitude toward the behavior	Attitude toward the behavior, Subjective norm, Perceived behavioral control	Subjective norm, Perceived usefulness, Perceived ease of use	Performance expectancy, Effort expectancy, Social influence

4 Methodology

This study, based on TAM, is focused to explain and forecast IT acceptance behavior, as well as to develop external variables which influence perceived usefulness, perceived ease of use and behavioral intention. Two of the three external variables, characteristics of IT and facilitating conditions, are selected from TAM2 and unified theory of acceptance and use of technology (UTAUT) respectively. The third external variable is the degree of E-business. The research structure is shown below.

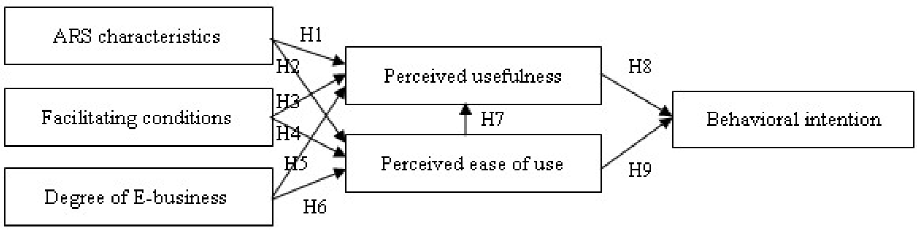


Fig. 1. Research structure

The questionnaire was developed from a critical review of literatures. The descriptive statistical analyses of quantitative data are performed using SPSS 15.0 for Windows statistical software. Statistical methods involve descriptive statistics, reliability and correlation analysis.

5 Results

5.1 Descriptive Analysis

214 questionnaires are distributed and collected, there were 26 questionnaires are invalid, and the rate of valid questionnaire is 88%. Top 3 industries of respondents are IT industry (31.9%), manufacturing (18.6%), and financial and insurance industry (10.1%). The numbers of male and female respondents were close, with 70% respondents' ages under 30 years old. Over 50% of respondents' working experiences are under 3 years, and most of respondents don't know about and never use ARS. About 30% of respondents' firm size are under 50 and 50% from firm size over 200.

5.2 Validity of Questionnaire

The initial questionnaire was developed from literature review and related research, and was pretested by 2 experts and 16 employees. Some problems such as the 3 industry categories (manufacturing, service, and financing) were found to be hard to select by respondents. Some question design is not easy to answer and the introduction of ARS is too wordy for them to have a quick understanding of ARS. The questionnaire was revised accordingly.

Table 3. Table 1 Characteristics of respondents

Industries	%	**ARS characteristics**	%
IT industry	31.9	Interactivity	37.2
Manufacturing	18.6	Demonstrate result Immediately	34.0
Financial and insurance industry	10.1	Anonymous response	12.8
Business consultancy	9.0	Allow large audience participation	12.2
Import and export trade	7.4	Auto-save result	2.7
Retail	6.9	Others	1.1
Education and publishing	6.9	**Have heard ARS**	%
Pharmaceutical and biochemical industry	4.3	Yes	5.3
Food, beverage and entertainment	2.7	No	94.7
Media and advertising	2.1	**Have used ARS**	%
Information professional	%	Yes	14.4
Yes	77.1	No	85.6
No	22.9	**Scores of intention**	%
Gender	%	Minimum	0
Male	46.3	Maximum	10
Female	53.7	Mean	6.46
Age	%	**Education**	%
Under 30	69.7	Graduate school and above	31.9
31-40	25.0	University and college	63.3
41-50	3.7	High school	4.8
Over 51	1.6	**Firm size**	%
Working experience (years)	%	Under 50	30.3
Under 1 year	19.1	51-100	5.9
1-3	39.4	101-150	10.1
4-6	17.6	151-200	3.7
7-9	11.2	Over 201	50.0
Over 10 years	12.8		

5.3 Reliability Analysis

The internal reliability or consistency of each aspect of TAM model is listed in Table 4. Cronbach's values are 0.905, 0.858, 0.880, 0.870, 0.852 and 0.920 respectively for ARS characteristics, Facilitating conditions, Degree of E-business, Perceived ease of use, Perceived usefulness, and Behavioral intention. A widely accepted rule of thumb

Table 4. Reliability of questionnaire

Factor	N of Items	Cronbach's Alpha
ARS characteristics	8	0.905
Facilitating conditions	6	0.858
Degree of E-business	6	0.880
Perceived ease of use	4	0.870
Perceived usefulness	4	0.852
Behavioral intention	6	0.920

of internal consistency for Cronbach's should be at least 0.70 [20]. In this study, all Cronbach's values are higher than 0.85, which shows the questionnaire has high reliability.

5.4 Correlations

The bivariate correlations among 6 factors are listed in Table 5. Most factors are significantly related to each other, only the degree of E-business does not show correlation with other factors.

Table 5. Bivariate correlations between variables

Variables	ARS characteristics	Facilitating conditions	Degree of E-business	Perceived ease of use	Perceived usefulness	Behavioral intention
ARS characteristics	1.000					
Facilitating conditions	0.717**	1.000				
Degree of E-business	0.019	0.123	1.000			
Perceived ease of use	0.703**	0.469**	0.089	1.000		
Perceived usefulness	0.845**	0.747**	0.076	0.597**	1.000	
Behavioral intention	0.631**	0.645**	0.153*	0.444**	0.716**	1.000

** Correlation is significant at the 0.05 level (2-tailed).
* Correlation is significant at the 0.1 level (2-tailed).

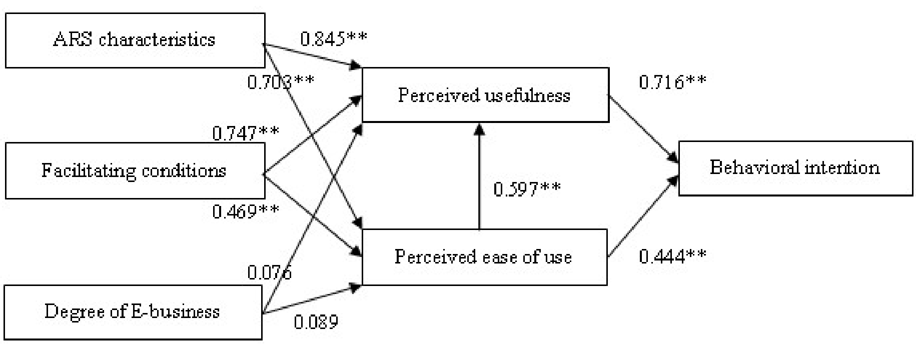


Fig. 2. Correlation coefficient

5.5 Analysis of Research Hypotheses

The purpose of this study was to understand relationships among external variables (ARS characteristics, facilitating conditions and degree of E-business), people perception and behavioral intention. From the statistic analysis, 9 hypotheses testing results were summarized in table 6. H1, H2, H3, H4, H7, H8, H9 were accepted, that is ARS characteristics and facilitating conditions have positive correlations with perceived usefulness and perceived ease of use. Between Perceived ease of use and perceived usefulness exist significant correlation and also with behavioral intention. H5 and H6 were rejected because Degree of E-business did not show significant correlation with perceived usefulness and perceived ease of use but with behavioral intention.

Table 6. Research hypotheses and statistical result

Number	Hypotheses	Result
H1	AC → PU	Accept
H2	AC → PE	Accept
H3	FC → PU	Accept
H4	FC → PE	Accept
H5	DE → PU	Reject
H6	DE → PE	Reject
H7	PE → PU	Accept
H8	PU → BI	Accept
H9	PE → BI	Accept
AC: ARS characteristics	FC: Facilitating conditions	DE: Degree of E-business
PE: Perceived ease of use	PU: Perceived usefulness	BI: Behavioral intention

6 Conclusions

6.1 IT Characteristics Could Be External Variables of TAM

Our results show that characteristics of ARS including interactivity, convenience, instantaneity, and anonymous have significant correlation with perceived ease of use and perceived usefulness. The relative importance of characteristics are interactivity (37.2%), instantaneity (34.0%), convenience (14.9%) and anonymous (14.9%). Activities like training and meetings are found to lack of interactivity and instant response in corporations. Compare with western countries, eastern countries are more conservative and introversive. People tend not to express opinion in the public, and therefore ARS can be a good tool for them to enhance interactivity and make activities more interesting.

6.2 Facilitating Conditions Could Be External Variables of TAM

Facilitating conditions are divided into internal (top manager support, education training, and organization arrange specific staff to assist using) and external (ARS providers, government, and schools) conditions. Our results show that the importance of facilitating conditions are top manager support (76.6%), committed helping staff (70.7%), providers' promotion (61.7%), schools' (56.9%) and government's (50.0%) support and endorsement. The importance of all variables was higher than 50% which demonstrated that these are important factors for introducing IT. Internal facilitating conditions are more significant than external ones because they are closer to employees to increase their behavioral intention of IT acceptance and adoption.

6.3 The Degree of E-Business Is Not a Factor of IT Behavioral Intention

The degree of E-business did not have significant correlation with perceived ease of use and perceived usefulness and hence H5 and H6 were rejected. We consider this factor unlike ARS characteristics and facilitating conditions have direct relationship with ARS, but it has significant correlation with behavioral intention. And we depend on result to revise ARS acceptance model to build linkage between degree of E-business and behavioral intention.

6.4 Preference Group of ARS

This research try to explore difference of ARS acceptance in demographic variables (gender, age, education, working experience), industries and firm sizes, but the result show that they didn't have. Since most of respondents did not know (85.6%) ARS and it's not popular in enterprises, not preference group can be found.

7 Suggestions

7.1 Creating Multiple Interaction Channels for Using ARS

Providers should emphasize interactivity and instantaneity when promoting or demonstrating ARS, and try to create multi-use or other interesting ways to enhance interactivity, real time response and innovative demonstration to reach the goal of popularizing ARS.

7.2 Improving Internal Facilitating Conditions

Enterprises' managers and ARS providers should understand that facilitating conditions are important factors for successful introduction of new IT tools. Managers are key persons to lead, require, and provide incentives to encourage employees using new IT. In addition, difficulties in using the system by the staff will reduce their intention of use. Staff assigned specifically for solving IT problems in time can effectively reduce resistance to new IT. Providers should focus on external facilitating conditions to promote ARS. In Taiwan, there are more than 800 teachers have been using ARS in teaching [21] which shows that it is a good venue for providers to promote ARS through schools and government.

7.3 Developing Strategic Stages of Introducing IT

Although the relationship between the degree of E-business and perception was not established, but it has significant correlation with behavioral intention. Hence, the higher degree of E-business the better acceptance of new IT and vice versa. The nature of company should be studied to understand whether it is suitable for introducing new IT or not because if employees are unfamiliar with IT may result in resisting behavior. We suggest low degree of E-business should have strategic stages of introducing IT and build sound internal facilitating conditions to increase employees' willingness.

7.4 Increasing the Promotion of ARS

The results show that 85% of respondents did not know ARS, 70% respondents are under 30 years old and we distribute and gather questionnaires through internet, the samples' characteristics of this study should have higher probability of knowing new IT, if samples from internet didn't know ARS, and we must can infer the population knowing ARS is lower than 15%, so providers should increase the promotion of ARS, and also through academic or government.

8 Future Research

8.1 Comparison between Eastern and Western Countries

The targets of this study were employees of Taiwan enterprises, but it's common that western people express their personal opinions, so they may not need ARS to enhance interactivity. Future research could have a comparison between eastern and western countries' enterprise on ARS characteristics and acceptance of ARS.

8.2 Define and Filter Samples

Introducing ARS does not have direct correlation with profit, and the mean of using intention is 6.46, it means respondents still wait to see. Our research did not choose specific groups to explore ARS acceptance, some firms' size are lower 50 people and some did not place importance on education training. Therefore, we suggest future study can define and filter samples before distribute questionnaire, such as firm size over 50 or 100 people and companies which pay much attention to education training.

References

1. Taiwan Research Institute (TRI), `http://www.tri.org.tw/ceo/page/a3.htm`
2. World Economic Forum (WEF), `http://www.weforum.org/en/index.htm`
3. LaRose, J.A.: Engage Your Audience. Prof. Saf. 54, 58–62 (2009)
4. McCarter, M.W., Caza, A.: Audience Response Systems as a Data Collection Method in Organizational Research. J. Health Organ. Manag. 15, 122–131 (2009)
5. Leavitt, H.J., Whisler, T.L.: Management in the 1980's. Harv. Bus. Rev., 41–48 (1958)
6. Adams, H., Howard, L.: Clever Clickers: Using Audience Response Systems in the Classroom. Library Media Connection 28, 54–56 (2009)
7. Loertscher, D.: Does Technology Really Make a Difference? Teacher Librarian 37, 48–50 (2009)
8. Beatty, I.: Transforming Student Learning with Classroom Communication Systems. Center for Applied Research (2004)
9. Liu, T.C., Lin, Y.C., Bhattacharya, M.: Introducing Learning Technologies into Classroom in Accordance with Teacher's Instructional Approach. In: 8th IEEE International Conference, pp. 1007–1008 (2008)
10. Smith, D.A., Maguire, K.A.: Effectiveness and Eeeiciency in the Optimization of Timing Structure of Assisted Recall Technology in Introductiory Accounting. In: Allied Academies International Conference, vol. 14, p. 33 (2009)

11. Strasser, N.: Who Wants to Pass Math? Using Clickers in Calculus. Journal of College Teaching and Learning 7, 4–52 (2010)
12. Lincoln, D.J.: Student Response Systems Adoption and Use in Marketing Education: A Status Report. Marketing Education Review 19, 25–40 (2009)
13. Liang, J.K., Liu, T.C., Wang, H.Y., Chang, B., Deng, Y.C., Yang, J.C., Chouz, C.Y., Ko, H.W., Yang, T.W.: A Few Design Perspectives on one-on-one Digital Classroom Environment. Journal of Computer Assisted Learning 21, 181–189 (2005)
14. Morse, J., Ruggieri, M., Whelan-Berry, K.: Clicking Our Way to Class Discussion. American Journal of Business Education 3, 99–108 (2010)
15. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading (1975)
16. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology. MIS Quarterly 13, 319–340 (1989)
17. Ajzen, I.: The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes 50, 179–211 (1991)
18. Venkatesh, V., Davis, F.D.: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Manage. Sci. 46, 186–204 (2000)
19. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 27, 425–478 (2003)
20. Nunnally, J.C.: Psychometric Theory. McGraw-Hill, New York (1978)
21. Liu, T.C., Liang, J.K., Wang, H.Y., Chan, T.W., Wei, L.H.: Embedding EduClick in Classroom to Enhance Interaction. In: International Conference on Computers in Education (ICCE), pp. 117–125 (2003)

Identification of Variables and Factors Impacting Consumer Behavior in On-line Shopping in India: An Empirical Study

Sudesh Chhikara

Department of Commerce, Kurukshetra University,
Haryana, India
sudesh_kuk@rediffmail.com

Abstract. On-line shopping is a recent phenomenon in the field of E-Business and is definitely going to be the future of shopping in the world. Most of the companies are running their on-line portals to sell their products/services. Though online shopping is very common outside India, its growth in Indian Market, which is a large and strategic consumer market, is still not in line with the global market. The potential growth of on-line shopping has triggered the idea of conducting a study on on-line shopping in India. The present research paper has used exploratory study to depict and highlight the various categories of factors and variables impacting the behavior of consumers towards on-line shopping in India. The data was collected through in-depth interviews on a sample of 41 respondents from Delhi, Mumbai, Chennai and Bangalore. The results of the study show that on-line shopping in India is basically impacted by five categories of factors like demographics factor, Psychographics factor, Online shopping feature and policies, Technological factor, Security factor. The results of the study are used to present a comprehensive model of on-line shopping which could be further used by the researchers and practitioners for conducting future studies in the similar area. A brief operational definition of all the factors and variables impacting on-line shopping in India is also described. And finally practical implications of the study are also elucidated.

Keywords: On-line Shopping, Psychographics Factor, Demographics Factor, Technological Factor, Security Factor.

1 Introduction

The Internet and worldwide web (www) and have dramatically changed the way consumers seek and use information. The Internet, which was earlier conceptualized as a tool for enchasing information, has become an important place of business these days. Internet as a business place provides enormous opportunities for everyone. Users on internet are no more limiting themselves to only mailing, chatting and information retrieval. Today through internet, they are shopping for information, shopping for goods and services, or using online services and many more such services. Internet has made almost everything just one click away from the customers.

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 292–304, 2010.

Due to exponentially rising business opportunities, there are a number of services being offered on the internet. On-line shopping has emerged as one of the most prominent services available through internet. It has enormous advantages for the customers as well as business houses. Through on-line shopping, business houses have been able to reach out to more customers at less cost. They have been able to reach out to customers living in remote areas. In-fact these are acting as stepping-stones to concept of global village. More over the inventory management overheads also decrease significantly through online shopping.

However, it is not that the online shopping has benefited only the business houses. As truly stated "Customer is the King today", the concept of on-line shopping will not materialize if the customers are not benefited. Online shopping brings many advantages to the customers. Customers can shop from any place and need not physically visit the shops/outlets for shopping purposes. Therefore, even if customer is staying in remote area, he/she can easily shop through internet. If a customer goes to any retail outlet or any other shop, the choice of products is normally very limited. However, here customers can visit any number of sites to reach at final choice. Hence, online shopping provides unlimited choices to the customers in nut shell. There is no fixed timing issue with the internet and availability is always there. The customer can shop any day of the year on any time of the day. This also helps in customers' time and energy saving. More over due to unlimited choice and less excess time, customers can easily search for the desire things and can easily compare the products/items.

Objectives of the research paper are based on the gaps identified in literature review. As the literature review shows more number of studies on online shopping pertains to foreign countries like U.S.A., U.K., and China, Japan etc. But in India on-line shopping is still in nascent stage. The results of the earlier studies in foreign context cannot be extrapolated in Indian context as the cultural, demographic and psychographics factor will creep in. Therefore, there is a need to conduct a study in the similar area in Indian context so that Indian retailers and customers could also take advantage of this new technology in the times to come. Based on the literature gaps the following are the objectives of the present study.

1. To identify the variables influencing behaviour of consumers towards on-line shopping in India.
2. To identify the factors influencing behaviour of consumers towards on-line shopping in India.

2 On-line Shopping Trends in India

India has not been far behind the online sales trends seen across the globe. The e-commerce sales were 2300 crores in 2006-07 as compared to 1200 crore in 2005-06 [1]. It shows a significant increase of around 100%. While significant, those sales numbers still represents less than 1% of total retail sales in India [2]. Online travel retail market in India is expected to get a boom over the next five years, with Internet-based travel retail transactions increasing by a whopping 271% between 2005 and 2010 [3].

Furthermore, a recent study by ACNielsen (2009), covering 38 markets and over 21,100 respondents across the globe has revealed that more Indians are taking to shopping online [4]. A significant observation of this study was that India beat the global counterparts in number of purchases per month, with a mean of 5.2 purchases against the global average of 4.9. In India, books followed airline reservations closely, with 35% of netizens buying them online. Nearly 24% have bought electronic items and more than 20% have purchased items such as apparel, music and electronic entertainment such as movies, DVDs and games. The most favored mode of payment for online purchases in India is the use of credit cards followed by cash-on-delivery. Online shopping in India is poised for greater acceleration as more manufacturers and providers integrate the Internet into their sales model. As PC and internet penetration grows, the key to increasing online purchases will remain in the hands of marketers in India.

3 Literature Review

Consumers and firms are conducting a substantial and rapidly increasing amount of business on the internet. According to Forrester research Inc., it was projected that the web will generate consumer and business-to-business sales in excess of US $ 1000 billion by 2010. This clearly shows the relevance of online shopping in future.

Identification of factors and variables impacting Consumer behaviour has been quiet new area of research throughout the globe. Initial research result in this area shows that willing ness to buy is most influenced by trustworthy ness of the suppliers [5]. Some studies have focused on the consumers' attitudes towards Internet shopping while others have focussed on the factors affecting those attitudes towards internet shopping. For example, Poel and Leunis (1999) suggested that the consumer's adoption of the Internet for retail purchases focused on three main attributes, money back guarantees, price reductions, and well-known brands [6]. Regan (2002) examined that the factors that would most strongly increase online shopping would be: (1) an increase in major catalogue retailers taking steps to convert customers into web buyers, and (2) overcoming the tactile needs of online shoppers to make them more comfortable in buying clothing without first touching or trying on the garment [7].

Research on the Internet shoppers has typically included demographic questions of age, education and household income. Over time the Internet buyers, once considered the innovator or early adopter, has changed. While once young, professional males with higher educational levels, incomes, tolerance for risk, social status and a lower dependence on the mass media or the need to patronize established retail channels, today's Internet buyers show a diversity of income and education [8-9].

For Internet buyers, gender, marital status, residential location, age, education, and household income were frequently found to be important predictors of Internet purchasing [10-11]. For example, Sultan & Henrichs (2000) reported that the consumer's related to income, household size, and innovativeness [12]. Akhter (2002) indicated that more willingness to and preference for adopting the Internet as his or her shopping medium was also positively educated, younger, males, and wealthier people in contrast to less educated, older, females, and less wealthier are more likely to use the Internet for purchasing [13].

From the customer's point of view, the Internet offered the potential advantages of reducing shopping time and money spent. It allowed twenty-four hours a day access, provided better service, and gave the consumers a perception of control over the shopping experience [14].

The consumer factor was suggested as an important factor for online shopping and items included under this were privacy, security, time saving, ease of use, convenience, enjoyment, previous experience, company reputation and tactility [15]. According to Komiak & Benbasat, (2004) Privacy in a communications system or network is defined as a protection given to information to conceal it from others' access by the system or network [16]. However, Privacy concern was the most frequent reason cited by consumers for not making online purchases.

Becker (1965) noted that the efficient use of time was a critical issue for the modern time-scarce consumer [17]. Internet shopping can be viewed as a time saver for the shopper and the buyer. As such, time positively influences Internet shopping as it can eliminate trips to the store and the long lines and delays when at the store [18]. Studies have found that more years of computer experience and use had a positive, direct effect on the user's acceptance of information technology This suggests that consumers with more years of computer use would be more likely to adopt the Internet for purchasing. Related technology variables identified by O'Keefe et al. (1998) included technology skill and the technology anxiety as significant elements that predicted online buying behavior [19].

Having a positive company reputation can reduce the consumer's perceived risk of trying a new means of distribution. Such a reputation is developed over time through long-term relationships with the consumer. A retailer's reputation is partially built on the customer's ability to have direct face-to-face contact with the store and its management [20]. Online stores, by not having direct contact with the consumer, may have a more difficult time of establishing a reputation, thus decreasing the likelihood of online buying [21]. The last consumer issue is the ability to test, in terms of touch and sight, a product before buying. Consumers express apprehension when buying a product without a tactile examination [22].

Product promotions attempt to influence the consumers' purchasing behavior [23]. Like other retail methods, online channels have various promotional tools such as corporate logos, banners, pop-up messages, e-mail messages, and text-based hyperlinks to web sites. These types of promotions have positively affected Internet buying [24].

The separation of buyer and seller noted above also plays a role in the consumer's level of comfort in regard to product returns. Today, businesses often respond to a customer's request to return a product by offering to repair, substitute, or refund the customer's money. In the case of online shopping, where the majority of products have been delivered through some third-party means, the customer is now faced with utilizing a similar service in the return process, an additional inconvenience and potential expense. These issues negatively affected online shopping behavior of consumers [25].

To a degree, online buying will depend on the efficiency and availability of the technology [22. In technological factors three main technological factors were suggested as important to online shopping: the availability of personal computers and Internet access, download time and representativeness of pictures and colours [27].

Summarizing the whole literature review, it can be said that over the years more and more factors have been studied in on-line shopping environment but no significant study has been done in Indian context. Companies have done few studies to target their customer through on-line retail store. The studies done abroad already make an assumption that internet is accessible everywhere and everyone uses it. But in India, internet diffusion factor has an important role to play. Therefore this research paper basically aims to identify and describe those factors and variables impacting on-line shopping in India.

4 Research Methodology

The choice of research design mainly depends on the type of research objectives addressed through the study. The phenomenon of online shopping has been quite a recent development in the context of India. The objective of the study was to explore the phenomenon of online shopping in Indian market perspective. Hence, the study used exploratory research design. The exploratory research design in the study was basically used to identify the various factors and variables impacting online shopping behavior in Indian context. In exploratory research, in-depth interview and review of existing literature were used to collect the relevant data from the respondents and to compile the final list of factors and variables. The in-depth interview used in the study was a structured direct, personal interview in which a single respondent was probed to uncover underlying motivations, beliefs, and feelings on a topic. Another method used in exploratory design was review of existing literature which is detailed in preceding part of paper. The review of literature on online shopping helped to identify the variables and factors relevant for the study.

The Primary data for the present study was collected through in-depth interviews. Whereas secondary data mainly includes literature review conducted through journals, national and international publications, internet, books and libraries. A brief discussion of the primary data collection method used in the study is given below.

In-depth interviews were used in the study to identify the relevant factors impacting online shopping in the Indian context. Another major objective was to validate the relevance of factors, identified from literature, in the Indian online shopping scenario. Respondents were mainly asked following questions in the in-depth interview:

Q1. What is your viewpoint regarding the growth of online shopping in India?
Q2. What is going to be the future of online shopping in Indian Market?
Q3. What are various types of variables and factors impacting online shopping in India?
Q4. How many categories of items are shopped online?
Q5. Which are the most critical factors and variables impacting online shopping in India?
Q6. Which are the most favorable (positive) variables for on-line shopping?
Q7. Which are the most unfavorable (negative) variables for on-line shopping?

The sample size for in-depth interview was 41. The sample for in-depth interviews was mainly collected from Delhi, Mumbai, Chennai and Bangalore. The sample was mainly collected from metros and big cities assuming the high Internet diffusion rate.

Table 1. Sample Size for In-depth Interviews

In-depth Interviews Respondents	
City	**No. of Respondents**
Delhi	12
Mumbai	11
Chennai	8
Bangalore	10
Total	41

Source: Compiled from the Exploratory Study

In the present study non-probabilistic sampling technique was used. Judgmental and snowball sampling technique were used in respondent selection for in-depth interviews. In the present study Judgement sampling was based on the following parameters:

- The sample comprised of people who have done online shopping.
- Only those people having credit cards were part of the sample.
- The sample comprised of people whose minimum qualification was at least Graduation.

The sample was taken from metropolitan cities assuming high internet diffusion rate.

While using snowball sampling, initial set of respondents was selected on the basis of judgmental sampling. Subsequently additional units were obtained on the basis of information given by initial sample units and then further referrals were taken from those selected in the sample. In this way sample was grown by adding more and more referral-based respondents until it reached the limiting number (15/city).

5 Analysis and Interpretation

A detailed literature review was conducted to identify the most critical factors impacting online shopping. However, the objective was to identify the various categories of factors impacting online shopping in the context of Indian Market. Therefore, detailed in-depth interviews were conducted on a sample of 41 respondents to ascertain the validity of critical factors (identified through literature review) in the context of Indian Market Scenarios. The respondents were asked to discuss the various categories of critical factors and variables, as perceived by them, during the in-depth interviews. The data analysis techniques used in this step were keyword search and pattern analysis. Keyword search technique was mainly used to identify the categories/variables and further to normalize the title and scope of the categories/variables. Pattern analysis was used to check the repeated patterns of categories/variables across the respondents and also within the overall response of individual respondents.

The data from interviews were analyzed through a process of analytic induction. For analyses notes and interview transcripts were scanned to identify initial set of coding categories that emerged from the qualitative data. Statements made by respondents were analyzed for their significance to answering the research questions. Though it is difficult to present entire data upon which conclusions are based, however, the variety of reasoning given by on-line shoppers and context in which they were given, necessitate a more qualitative presentation of the supporting statements. Therefore, selected excerpts of opinions are cited to give texture to the analytic conclusions and to provide samples of the kinds of evidence upon which those conclusions are based.

A synthesis of responses is presented here along with the category they represent; these categories are based on summarizing the patterns emerging from individual responses.

Following factors emerged as most important ones behind on-line shopping.

Major variables behind recall of Security

Many of the respondents recorded that they fear to indulge in on-line shopping because they don't want to share their financial information with a third party. And they have heard about the on-line scandals where the financial information was stolen by some malicious party and used illegally. But they were still optimistic about the future of on-line shopping in India.

"The most important variable I concern while shopping on-line is the privacy, as I am not comfortable in sharing my personal information with others".

(Sandeep Gupta, Male, 28 years, Software Engineer, Bangalore)

Major variables behind recall of Online shopping feature and policies

The most common variables identified through the people for this category were promotion policy, delivery policy, product return policy, detailed information of product, option of comparison of the products.

As said by a respondent during the interview: "I enjoy shopping on-line because it allows me to take advantage of the best deals available in the market. In addition I have the option to search out and compare several brands and products in on-line shopping mode at the same time".

(Mrs. Bharti Varshneya, Female, 26 years, housewife, Delhi)

Major variables behind recall of Psychographics factor

Most of the people identified Innovativeness, enjoyment, convenience, interaction with other people and touch and feel as the important variables for on-line shopping.

"I prefer on-line shopping over physical store shopping because of the ease and comforts in it. I can shop from anywhere in the world at any time of the day. It is hassle free. I need not wait in the long queues waiting for my turn. It saves my time, effort and energy".

(Vikas, Male, 30, Project Manager, IT, Mumbai)

Major variables behind recall of Technological factor

As cited by the respondents in in-depth interview the major categories of factors impacting on-line shopping in India are the Quality of product, representativeness of product, value for money in on-line shopping.

"I am not a frequent user of online shopping. Generally I use on-line shopping for buying the branded and quality products or international labels which are not much prevalent in India. Because I think that quality of the product is maintained in on-line stores".

(Tanmay Vyas, Male 24 years, Student Physiotherapy, Chennai)

The results of the exploratory study based on key word search and pattern analysis are given below in the form of a table including the different categories of factors and variables under each category of factor as studied in Indian context.

Table 2. Categories of Factors and Variables under Each Factor Category Identified from Exploratory Study

S#	Category of factor	Variables under factor category
1	Demographic factor	Age, Gender, marital status, income/salary, family size, education, ability to use internet
2	Psychographics factor	Innovative, enjoyment, convenience, interaction with people, touch and feel
3	Online shopping feature and policies	Promotion, delivery policy, product return policy, detailed information of product, option of comparison,
4	Technological factor	Quality, representativeness of product, value for money
5	Safety factor	Security, privacy

Source: compiled from the results of exploratory study and literature survey

A brief operational definition the various factors and variables identified in the study are given below:

1. Psychographics factor

Psychographics factors of consumers are those that delve deeper into peoples' lifestyle and behaviours, including their interest and values, their likings and disliking. A brief definition of psychographics variables identified in the study is as following:

- Innovative:

Being innovative means to be ahead of time, or a forward looking person, or being or producing something like nothing done or experienced before.

- Enjoyment:

Enjoyment is the condition of enjoying anything like pleasure or satisfaction, which arises due to possession or occupancy of something.

- Convenience:

The quality of being suitable to one's comforts, purposes or need, or anything that saves or simplifies work, adds to one's ease or comforts is called as convenience.

- Interaction with people:

It means taking an interest in others affairs, talking about their interest, likings and disliking, taking their suggestions etc.

- Touch and feel:

Touch and feel factors are related to a customer's experience of shopping by touching and feeling the product. In other words, it is an intangible feature in tangible shopping.

2. Online Shopping Features

Online shopping features mainly refer to all kinds of functionalities and features that are prevalent in the context of online shopping. Certain interactive features of e-commerce sites such as e-mail alerts of special offers, pop-up menu and "click able", product arrays stimulate e- shopping by consumers. Present study includes following most important features of online shopping:

- Promotion:

Promotion is the act devised to publicize or advertise a product, cause, institution, etc., as a brochure, free sample, poster, and television or radio commercial or personal appearance.

- Delivery policy:

Delivery policy in context of online shopping is the official pledge undertaken by retailers to deliver the goods in the stipulated time period to opposite party.

- Product return policy:

Product return policy is one of the features of on-line shopping where consumers can return a product back to the e-tailor, if the product seems to be defective or of substandard. And consumer either gets the new product or reimbursement of money from e-tailor for the same.

- Detailed information of product:

This is one of the facilities provided by online shopping where one can get accurate information on various websites about the specific product one wishes to buy.

- Option of comparison:

It means capable of being compared, having features in common with something else to permit or suggest comparison. Online shopping facilitates easy comparison of different products in terms of quality, features, price, discounts and availability of products with different web stores.

3. Demographic Factor

These are statistical socio-economic characteristics or variables of a population, like education, income and marital status, occupation, religion, birth rate, death rate, and family size etc. Following variables are identified in demographic factor for online shopping.

- Age:

Age is a period of human life, measured by years from birth usually marked by certain stage or degree of mental or physical development and involving legal responsibility and capability.

- Gender:

Gender denotes biologically determined, culturally and socially constructed difference between men and women.

- Marital status:

The condition of being married or unmarried is called as marital status.

- Income/Salary:

It is monetary remuneration computed on hourly, daily, weekly, or monthly basis. A fixed weekly or monthly wages is called as salary.

- Family size

A family consists of a domestic group of people (or a number of domestic groups), typically affiliated by birth or marriage, or by analogous or comparable relationships — including domestic partnership, cohabitation, adoption, surname etc. and family size represents the total number of people in a household.

- Education

The act or process of imparting or acquiring general knowledge, developing the powers of reasoning and judgment, and generally of preparing oneself or others intellectually for mature life. Education is a systematic instruction in a particular field of knowledge which will lead to development of character or mental power.

- Ability to use internet

Ability to use internet relates to how much a person is dexterous in using the new technology like internet.

4. Technological Factor

Technology is constantly changing and the pace of technological change is rapidly increasing. Technological development decides the growth of a nation. The online medium is a hybrid of multiple communication technologies, which offer a multitude of communication functions, and online shopping is one of such outcome of new technologies. Some of the variables in technological factor identified in the study are as following:

- Quality:

The totality of features and characteristics of a product or service that bears on its ability to satisfy stated or implied needs is called as quality. Quality of product in online shopping refers to the extent to which the consumer believes that the web vendor provides the quality of products and services he/she expected.

- Representativeness of product:

Representativeness of product means how closely the pictures and images shown in the website match with the actual product like colour, contrast, length, width, size, texture etc.

- Economic factor (value for money):

Value for money means whether or not a person obtained the maximum benefit from the goods and services he acquired from online shopping. It not only includes money value but an assortment of several values like quality, cost, fitness for purpose and convenience. In totality it includes Economy, Efficiency and Effectiveness.

5. Safety Factor

Perception of security refers to the extent to which the consumers believes that the web vendor does not use credit card information illicitly and has an ability to protect the information from hackers on the web. Security factor for the present study includes following variables:

- Security:

Online security means securing or safeguarding oneself from online fraudulent and mischief.

• Privacy:

Privacy literally means freedom from unauthorized intrusion. Perception of privacy refers to the extent to which the consumer believes that the web vendor protects the consumer's personal information by not selling it to third parties without the consumer's permission.

6 Conclusion

The final outcome of the study is a model that incorporates all the identified factors categories that impact On-line shopping in India.

Fig. 1. On-line shopping Model
Source: compiled from results of exploratory study and literature survey

The model given by the study shows that in India on-line shopping is basically impacted by Psychographics and Demographic variables of Consumers vis-a-vis On-line shopping Features, Technological, and Safety variables of On-line shopping.

The results of the study can be utilized by practitioners in re-looking or revamping their strategies for online shopping. And following are the practical implications of the study for online retailers, managers and practitioners:

• Online retailers should focus on better positioning and representation of their products in online shopping.

• Websites/e-tailors should also ensure that all the relevant details regarding the quality and features should be provided to the online shoppers on the website.

• Customer shop through online shopping because of time and effort saving, hence the online retailers should design their website in such a way that the customers can reach out to his/her desired product in the minimum time.

• To save time and effort of the customers the online retailers should also give an option to the customers to store and compare their choices before the actual purchase.

- Online website need to ensure that security and privacy policies are appropriately defined, implemented, and communicated to the relevant stakeholder.
- Online shopping companies could also look into the possibility of spreading and promoting security and privacy related information on the Internet.
- Delivery policy of the online product should also include a defined trial period to assure that the customer doesn't miss the touch and feel experience.
- Online retailers should also look into the possibility of running call centres which could ensure that the customer get a chance to formally interact with the other party before the actual purchase.

References

1. Kunz, M.B.: On-line customers: identifying store, product and consumer attributes which influences shopping on the Internet. Published doctoral dissertation. The University of Tennessee, Knoxville (1997)
2. Palmer, J.W., Markus, M.L.: The performance impacts of quick response and strategic alignment in specialty retailing. Information Systems Research 11(3), 12–23 (2006)
3. O'Keefe, R.M., O'Connor, G., Kung, H.J.: Early adopters of the Web as a retail medium: Small company winners and losers. European Journal of Marketing 25(1), 38–43 (1998)
4. Mahajan, V., Muller, E., Bass, F.M.: New product diffusion models in marketing: A review and directions for research. Journal of Marketing 54, 1–26 (1990)
5. Blattberg, R., Wisniewsk, K.: Price-induced patterns of competition. Marketing Science 8(4), 291–309 (1989)
6. Economic Times: E-tailing for success (2007), `http://www.economictimes.com/story/64859.html` (retrieved on October 4, 2007)
7. Citrin, A.V., Sprott, D.E., Silverman, S.N., Stem, D.E.: Adoption of Internet shopping: the role of consumer innovativeness. Industrial Management & Data Systems 100(7), 294–300 (2000)
8. Euromonitor International: A survey report on internet-based travel retail transactions (2005), `http://www.dancewithshadows.com/travel/india-online-travel-asia.asp` (retrieved on August 5, 2005)
9. Nielsen, A.C.: Indians beat world in cyber shopping (2009), `http://www.exchange4media.com/e4m/izone1/izone_fullstory.asp?section_id=4&news_id=19136&tag=13863`
10. Srinivasan, S.S., Anderson, R., Ponnavolu, K.: Customer loyalty in e-commerce: an exploration of its antecedents and consequences. Journal of Retailing 78, 41–50 (2002)
11. Mehta, R., Sivadas, E.: Direct marketing on the Internet: Ann empirical assessment of consumer attitudes. Journal of Direct Marketing 9(3), 21–32 (2004)
12. Komiak, S.X., Benbasat, I.: Understanding customer trust in agent-mediated electronic commerce, web mediated electronic commerce and traditional commerce. Information Technology and Management 5, 181–207 (2004)
13. Akhter, S.: Digital divide and purchase intention: Why demographic psychology matters. Journal of Economic Psychology 24, 321–327 (2002)
14. Becker, G.S.: A theory of allocation of time. The Economic Journal 229, 493–517 (1965)
15. Bell, G., Gemmell, J.: On ramp prospects for the information super highway dream. Communications of the ACM 39(6), 37 (1996)

16. Bhatnagar, A., Misra, S., Rao, H.R.: On risk, convenience, and Internet shopping behavior. Communications of the ACM 43(11), 98–105 (2000)
17. Eroglu, S., Machleit, K., Davis, L.: Empirical testing of a model of online store atmospherics and shopper responses. Psychology and Marketing 20(2), 139–150 (2003)
18. Ducoffe, R.H.: Advertising value and advertising on the web. Journal of Advertising Research 36(5), 21–35 (1996)
19. Jarvenpaa, S.L., Todd, P.A.: Consumer reactions to electronic sopping on the World Wide Web. International Journal of Electronic Commerce 1(2), 59–88 (2002)
20. Regan, K.: How bricks conquered the Net. E-Commerce Times (2002), `http://www.ecommercetimes.com/perl/story/16631.html`
21. Schiffman, L.G., Sherman, E., Long, M.M.: Toward a better understanding of the interplay of personal values and the Internet. Psychology & Marketing 20(2), 169–186 (2003)
22. Forrester Research Inc.: NRF/ Forrester online research index. Forrester Research, Inc., Cambridge, MA (2002), `http://www.forrester.com` (retrieved on July 22, 2008)
23. Poel, D.V., Leunis, J.: Consumer acceptance of the Internet as a channel of distribution. Journal of Business Research 45(3), 249–256 (1999)
24. Udo, G.: Privacy and security concerns as major barriers for ecommerce: A survey study. Information Management & Computer Security 9(4), 22–45 (2001)
25. Sultan, F., Henrichs, R.B.: Consumer preferences for Internet services over time: initial explorations. The Journal of Consumer Marketing 17(5), 23–46 (2000)

Knowledge Model of Emergency Decision-Making Based on Knowledge Supply and Demand

Bing Wu[1,2] and Lin-du Zhao[1]

[1] School of Economics and Management, Southeast University, Nanjing, China
[2] School of Economics and Management, Tongji University, Shanghai, China
Wubingsem@tongji.edu.cn

Abstract. In the process of rapid response to emergencies, to supply decision-makers the necessary knowledge is the key to improving the efficiency of decision-making. Starting from emergency decision-making process, this paper proposes emergency decision-making model based on knowledge supply and demand, and knowledge management process is designed to integrate into the emergency decision-making process to achieve an accurate auxiliary emergency decision-making. Modeling in knowledge management is very necessary to be able to clearly express the purpose of the creator. This model is to make an emergency decision-making through effective knowledge management, making full use of emergency management in all types of resources and optimizing resource allocation. And by the typhoon example illustrates the application of the model. Finally further research directions are noted.

Keywords: Knowledge of Supply and Demand, Matching, Knowledge Management, Emergency Response Decision-making, Knowledge Model.

1 Introduction

Currently the United States and Europe and other developed countries are on the forefront research of emergency response system, knowledge management research is one of the hot spot, stressing the use of knowledge management approach to decision-making support in the monitoring, data analysis, resource allocation and other issues, to support for the establishment of knowledge model and knowledge systems, at the same time domestic research has also begun off. Knowledge management support for the emergency is still in relatively early stage, but the more mature knowledge management technology is driving this process. To supply the necessary knowledge in rapid emergencies response to decision-makers is the key to improving the efficiency of decision-making.

This matching of knowledge supply and demand point of view put forward emergency decision-making model based on knowledge supply and demand, integrating knowledge management into the emergency decision-making processes to achieve an accurate auxiliary emergency decision-making. Model is a certain part of reality carried out in a purposeful abstraction, modeling thought in knowledge management is very

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 305–317, 2010.

necessary to be able to clearly express the purpose of creators. On the one hand from the destination point of view, the model is in line with knowledge management objectives, to make the right knowledge in the correct situations to reach the right people; the other hand, model provides an effective solution to the problems of knowledge management approach. Model can convert semi-structured and unstructured documents into a single standard format for storage, restructuring and integration of knowledge, so that decision-makers can quickly find the required knowledge, work together in sharing the knowledge to play a role in achieving the various regional organizations and individuals communicate better, faster and accurate decision-making. The purpose of the model determines the field of the model characteristics, emergency decision-making model based on knowledge supply and demand is to make an emergency decision-making through effective knowledge management, making full use of all types of resources and the rational optimizing resource allocation for scientific decision-making of emergency events.

This article is organized as follows, the second part of the overview of the relevant literature, the third section presents emergency decision-making model based on knowledge supply and demand, the fourth part of the application model instance of the typhoon were analyzed and summarized in Part V for further research directions

2 Review of the Literature

For the emergency management to follow the sudden and unprecedented features, the purpose of knowledge management is knowledge application, exception handling and business innovation. Another emergency management is a branch of management science to study the emergency management of information, organization, decision-making, plan preparation, emergency response system and other problems related to knowledge management issues.

Emergency management is a study of the phenomenon of an emergency incident and its laws of development disciplines, including management science, operations research, psychology and other knowledge, applied to natural disasters, accidents, disasters, social security incidents, and other areas of public health events, which itself is a multi - integrated body of knowledge categories. This requires integration of knowledge management concepts and methods, information on various aspects of information collection, integration, conversion, sharing, application and innovation in a series of activities to enable them to prevent disasters and reduce losses more effectively

Glorgos Papavassilio et al [1] proposed DECOR (Delivery of Context-Sensitive Organizational Knowledge) method, which defined required knowledge document for each task in the business process model, and described document content knowledge with operation task. P.W.H.Chung et al [2] put forward the process management framework based on task, directly linked task-related knowledge to the process of the task.

The U.S. government after the “911” incident, quickly set up a “national emergency response system” will integrate command, response, treatment technology system to the state comprehensive emergency response system, which focus on how to ensure that a variety of major accidents occurred, and efficient integration of various types of knowledge resources, quickly provide optimal decision-making and rapid relief of the knowledge management mechanism.

At the same time the United States, Britain and other research institutions have also invested more power, study of emergency response knowledge management, such as the emergency response decision-making system in the knowledge-based model, the contingency of knowledge of applied research, and the German Ministry of Research and Technology Development Organization based on inspection of the "toxic spill assessment system", through the application of knowledge management techniques, leakage of toxic substances that occur when the content and the proliferation of toxic gases to make rapid and accurate estimates to determine the protective measures.

Reorganization and integration knowledge by the scientific method is to use different sources, different levels, different structures, different knowledge of the contents in a comprehensive and integrated implementation of the re construction, so that a single knowledge, fragmented knowledge, old and new knowledge, explicit knowledge and tacit knowledge through the integration to enhance the formation of new knowledge. In accordance with the purpose of knowledge of a particular new combination of existing knowledge is a process from the component (Component Knowledge) into structural knowledge (Architecture Knowledge). To achieve new, ordering of knowledge, knowledge should be reorganized and integrated through effective mechanisms, methods, tools and technology. The main decision-making knowledge of the contingency model is shown as the following:

2.1 Structured Text-Based Knowledge Model

Its principles are to classify the text within area according to structural features of the text, and then establish corresponding data structure model for each feature of text. The general structure of the text are based on topic, title, paragraph formatting, etc., as separate unit break down, the smallest division also depend on the number of paragraphs in the text, the theme of completeness factors. In addition, the text should also be considered in some special data structure types, such as maps and tables, in short, with four kinds of data model to be structured decomposition: text data type, title, data model, the map data structure, table data structure.

2.2 Knowledge-Base Decision-Making Model

According to emergency response management in general processes, issues and away from the establishment of emergency management knowledge application process, that is, from needs of emergency incident to be done to determine what is an emergency, as well as the main object for which to take measures, and then determine the issue involved in the sudden emergency management department. After the emergency management department under the command center to handle emergency problems identified in the relevant knowledge, a series of arrangements to determine who is going to do, when to do, how to do, what resources, thus knowledge of emergency management issues document is acquired

2.3 Emergency Response Model for Emergency Events Based on Plans

With the contingency plans for the same kind of emergency incidents has feature of universal guiding significance, the reuse problem will be solved. The model is extracted from the contingency plan in emergency treatment process, according to subject, object, subject to resolve the object

Structured text-based knowledge model, based on a knowledge model and knowledge-based plans for emergency decision-making model is based on reorganization and integration of existing emergency response decision-making knowledge of the ideological model. The essence of the model lies in knowledge organization, focusing on emergency response from the perspective of the general process, following the logic of "issues discovery - problem analysis – problem solving" to reorganize knowledge, and linking into a process chain of decision-making based on the emergency response, in order to provide the decision-making knowledge for the purpose of emergency response management and know-how to play the effect.

However, this method for knowledge organization is based on epistemology level, neglecting that "create knowledge", "repositories of knowledge", "knowledge transfer" and "application of knowledge" would constitute a complete process to integrate into the emergency decision-making process, which can not really achieve the nature sense of synergy knowledge from the perspective of knowledge demand and knowledge supply. Therefore, this paper presents emergency decision-making model based on knowledge supply and demand from matching point of knowledge supply and demand, integrating knowledge management process into the emergency decision-making process, improving emergency response decision-making in knowledge cooperation, to achieve accurate and auxiliary emergency decision-making.

3 Emergency Decision-Making Model Based on Knowledge Supply and Demand

3.1 Knowledge in Emergency Decision-Making

3.1.1 Knowledge Characteristics in Emergency Decision-Making

Emergency event has characteristics of sudden, seriousness and urgency, which determines the knowledge of emergency management must be accurate, effectiveness and timeliness [3].

a) Accuracy: This is the primary features of emergency management knowledge. The main emergency command and dispatch process of a decision-making, a program, an action, may involve millions or even billions of huge economic losses and into the lives of millions of people, therefore, all knowledge, especially knowledge of the decision-making must be accurate and feasible.
b) Effectiveness: This is not only the knowledge feature of emergency management, but also one of the characteristics of a broad scope of knowledge. The ultimate goal of knowledge management is to apply knowledge to enable them more effective, so as to bring value to the organization.

c) Timeliness: Emergency incident response is a very urgent matter, policymakers need to make decisions and implement action programs in a very short period of time. Policy-makers for command and dispatch of knowledge must be timely, fast and be able to effectively to make emergency response, to minimize hazards and losses, as shown in Fig 1 for the rapid response of emergency schematic.

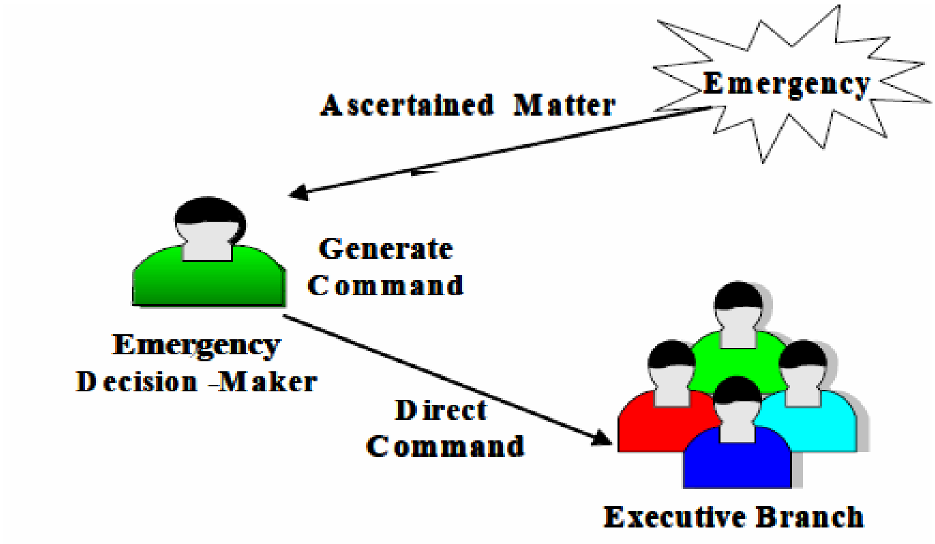


Fig. 1. Emergency Response Plan

In Fig 1, the ability of policy makers in emergencies after the incident, quickly and accurately to publish command and dispatch instructions, depends on two conditions: First of all, clearly what happened, and secondly, know how to resolve the problem has occurred, or solution. Three components constitute the time from problems to the executive branch of the rescue operation time: time from problem to decision-makers satisfying the circumstances is t1, time from identifying problems to decision-makers solving is t2, as well as time from program issued to the concrete implementation of sector departments is t3.

3.1.2 Classification of Knowledge in Emergency Decision-Making

According to OECD's view, the knowledge can be divided into four types:

a) Know what knowledge (Know-what): refers to the fact that knowledge about
b) Know why knowledge (Know-why): refers to the knowledge of principles and laws
c) Know how knowledge (Know-how): refers to something of skills, know-how and competence
d) Know who knowledge (Know-who): knows who has what knowledge and skills of knowledge, related to social relations.

Referencing to OECD classification of knowledge, combined with knowledge of emergency response decision-making content, adds aware of situational knowledge (Know-Contexts): refers to the knowledge in regard to situation; know the decision-making knowledge (Know-Decision): refers to knowledge of the decision-making principles and laws. Each type of knowledge has been given the characteristics of this area.

Decision-making knowledge is in the heart of emergency response decision-making, and only producing timely and accurate knowledge of effective decision-making, can decision-makers properly direct action to deal with emergency incidents. Emergency decision-making is directed against a variety of emergency incidents to make decision-making, and emphasis on more quickly and accurately compared to conventional decision-making,

3.2 Emergency Decision-Making Model Based on Knowledge Supply and Demand

In accordance with the contingency of knowledge in decision-making characteristics and classification, from a system point of view the decision-making under the emergency status of the external environment and internal change information, define problems, make decision-making objectives, evaluation, and options, and then a series of actions based on the selected program, and based on unexpected events at all stages of life cycle environment, the state feedback and constantly modify the plans, programs and actions, seeking the dynamic optimum decision-making.

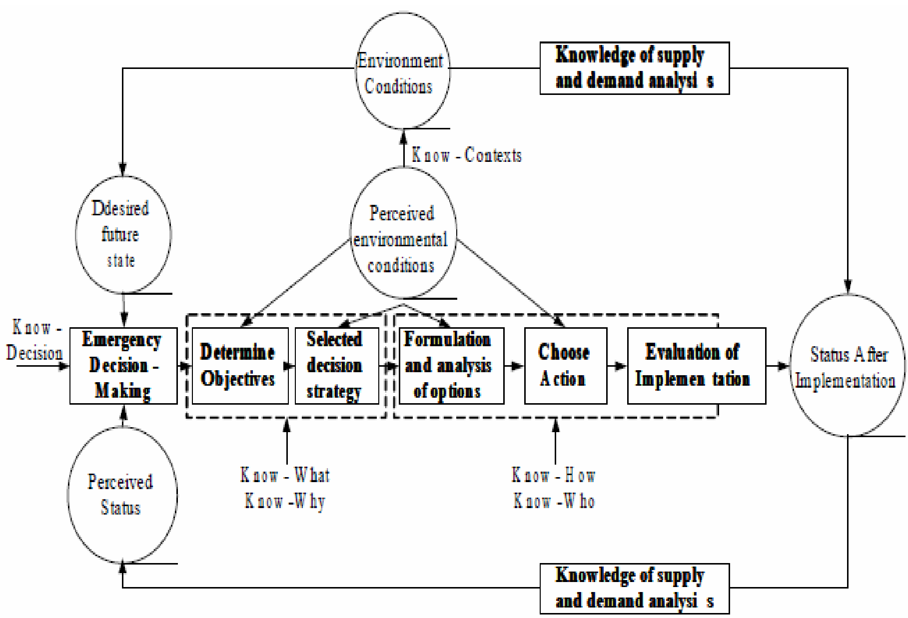


Fig. 2. Knowledge-Based Supply and Demand Model of Emergency Decision-Making

The model is divided into five processes, each process and its relationship shown as Fig 2. And emergency decision-making process of the knowledge input, knowledge processing and knowledge export processes, as shown in Fig. 3 and Fig. 4, have been reflected.

3.3 Decision-Making Process

Know-Contexts: Such knowledge mainly refers to the emergency decision-making process, according to specific situations of decision-making knowledge implementation [4], including: know when to do, what to do, what resources and protection. Context of knowledge throughout the decision-making process, each decision-making process needs to be complemented by the current situation of knowledge needed to support rapid and accurate decision-making and command and dispatch.

a) Propose a contingency decision-making problem
Know-Decision is core of proposing a contingency decision-making problem, which is shown in Table 1.

Table 1. Propose a Contingency Decision-making Problems with the Know-Decision

Items	Emergency Decision-Making Problem	Knowledge
1	To identify and find problems needed to be resolved is the starting point for emergency response decision-making process.	Know-Decision
2	The so-called problem is between expected and actual gap. Issue arising from emergency events is not only very complicated, but also required to identify and solve in a very short period of time.	According to experience and collective wisdom of policy makers on awareness and factual knowledge knowledge is reprocessing from the perspective of solving particular problem.. Response of decision-making on emergency incident is a manifestation of this knowledge.

b) Determine the Decision-Making Objectives
The decision-making objective, that is, decisions made through the implementation of the desired future state. Under normal circumstances, a decision-making to achieve the desired goal of the number of decision-making is often more than one. The same decision problem, if the desired goals are different, the final choice and implementation of the decision-making program will be different. To determine the emergency decision-making objectives, conditions existing to achieve this goal should also be considered. Therefore, conditions in different areas should be in comprehensive and detailed analysis. To achieve the goal of emergency response decision-making, controlled conditions must be met and non-controlled conditions should be estimated the trends and the impacts on emergency response decision-making goals in order to determine the appropriate emergency response decision-making goals.

c) Select Decision Strategy

Decision-makers in determining the decision-making goal, it must also be prepared to be taken to achieve the policy objectives of the strategy. The so-called decision-making strategy refers to the guiding ideology and action framework set by the decision-makers. Specifically, the decision-making strategies include: policy-makers want the decision-making problem to be resolved to what extent (i.e., expected how much can be achieved after the implementation of the decision-making) and how to achieve decision-making goals. Emergency decision-making should adopt a strategy in addition to its nature, scope and impact, the more depends on uncertainty of emergency decision-making environment.

Know-What and Know-Why, are mainly awareness and factual knowledge, it is selected to determine the objectives and decision-making strategy for conceptual, rational and regularity of knowledge, as well as the main organizing feature an emergency, responsibilities, status and emergency knowledge. Know-What and Know-Why in determining policy objectives, the selected decision-making strategies, as shown in table 2.

Table 2. Determine the Decision-Making Objectives, Selection of the Decision-Making Strategies and Know-What, Know-Why

Knowledge	Determine the decision-making objectives	Selection of decision-making strategies
Know-What	Such knowledge is ordinary rational knowledge required in emergency decision-making process. Emergency-related characteristics of the main body, emergency, and an overview of the general characteristics of the object and so on, which is what knowledge.	
Know-Why	Such knowledge is contingency events, development and evolution mechanism in emergency decision-making process, which means that an emergency incident knowledge of principles and laws	

d) Formulation and Analysis of Options

Decision-making process, select the program is the key component. In the emergency decision-making process, decision-makers are required around the goal to list many possible options for decision-making problems according to decision-making strategy adopted, according to the decision-making strategy adopted. Contingency options are able to quickly solve the problems caused by the emergency incident to ensure that decision-making objectives can be achieved with the feasibility of the implementation conditions of the program. Design options, the number of feasible options must be developed in order to evaluate and compare the final choice of a satisfactory solution.

Due to sudden emergencies, destructive disorder and uncertainty, emergency decision-making must be timely, fast, and accurate. In order to achieve this goal, the traditional decision-making patterns seem inefficient, and increase the psychological pressure on decision makers. At this time contingency plan is on the highlights of its importance. Contingency plan is the overall emergency response plans and procedures standards by the government manage, direct and coordinate the emergency response resources. Decision-making should be an emergency study to determine the decision-making goals of natural disasters, accidents, disasters, public health, social security incidents and other incidents of emergency management, command, and rescue plan, generally built on the basis of integrated planning. It describes the emergency decision-making disposal of domain knowledge, according to the state emergency response plans and events, adjusted immediately after the formation of an emergency plan approved. Plans are targeted, systemic, scientific, flexibility and operability characteristics. When they have induced signs of unexpected events, or emergency event has already broken, you can activate contingency plans, with the fastest speed control the spread of events so that the loss will be minimal. Plan is the core of decision-making, it has reduced the time the decision makers, reduced the stress of policy makers in emergency situations, and improved decision-making more scientific.

e) Selection Action

Program of action choice, the analysis estimated that the implementation of the various alternatives on the basis of the consequences. The methods used mainly for the advantages and disadvantages of the various alternatives were analyzed and compared the advantages and disadvantages of various programs, weigh the pros and cons, the final decision further. This phase is the design of the contingency options to conduct a comprehensive, detailed assessment, from which a satisfactory program can be selected. Identified goal is the standard to assess the strengths and weaknesses of contingency options, the most important goals should be as evaluation, selection criteria focus on emergency response programs. Due to the particularity of the object, emergency decision-making is not to choose the best programs, instead to choose the satisfied programs, which are so-called decision-making principles.

f) Implementation and Evaluation of Decision-Making

Whether decision is correct, effective and can achieve the desired objectives, should be tested through action program implementation. In implementation of the emergency decision-making process, decision-makers need to regularly evaluate the effect of the decision-making, including measuring the actual performance, and compare actual performance with expectations to assess the consequences of the implementation of the decision-making. If the performance is the lack of implementation due to improper decision-making, policy makers should take corrective action, such as re-coordinate the relevant departments to take remedial measures.

The result of decision-making process will be the decision-making knowledge; decision-making ultimately will be reflected in the implementation of a direct role of the emergency action. Know-Who and Know-How are the main part of an emergency response incident on the operational level of knowledge, that is what is at what time, what kinds of resources which protect the place, who should be how to do. Know-Who and Know-How and the formulation and analysis of options, select the programs of action and decision-making implementation of the evaluation, as shown in Table 3.

Table 3. Determine Decision-making Objectives, Strategies and Know-What, Know-Why

Knowledge	Formulation and Analysis of Options	Select Action	Implementation and Evaluation of Decision-Making
Know-What	This type of knowledge refers to the sudden emergency decision-making process, according to the decision-making knowledge to determine which departments, agencies, or to which an individual or organization responsible for implementation, namely, the knowledge to know who is doing		
Know-Why	Such knowledge is defined as the decision-making process of implementing an emergency department, the lead agency know how to respond to public emergencies, or emergency services to know how to do in this sector		

3.4 Knowledge-Driven Analysis of the Emergency Decision-Making Process

In the program implementation process, the decision-making environment is unpredictable due to the rapid changes, it will inevitably take place target deviation or new condition, and decision-making programs must be continually adjusting and improving to promptly correct any deficiencies in decision-making process. Application of enterprise-wide modeling IEM method, description, analysis and design of emergency decision-making processes, knowledge-oriented analysis of the emergency decision-making process is to make an emergency decision-making process more systematic application of knowledge and production activities in sync with the relevant personnel [5].

From a resources point of view, knowledge management, are "resources" sub-category; "knowledge" sub-category represents the knowledge needed to carry out its mandate the content; "knowledge holders" describes the possession of knowledge or management in their department experts; knowledge of explicit and tacit save as the electronic and paper form. Knowledge as "resources" of the object represents the supply, but knowledge as a "decision" of the object, on behalf of needs are. As a result, knowledge-oriented analysis of the emergency decision-making process, including the process inputs and outputs, the implementation of the tasks and the required critical knowledge areas, also describes the main carriers of knowledge.

Emergency decision-making process from the perspective of the knowledge requirements analysis, domain knowledge as input need knowledge processing, first of all, how to apply the knowledge; then, that knowledge where it comes from, how to come; third, the knowledge of how to save; fourth, the knowledge from where and how to generate. Emergency decision-making process from the perspective of the supply of knowledge [6-7] analysis, knowledge-processed output of the next state is tacit and explicit knowledge, First of all, resulting in what knowledge and the manner in which elections; Then, the knowledge how to save; Article Third, the manner in which the knowledge passed to where they are; Fourth, the knowledge is used where, for whom use.

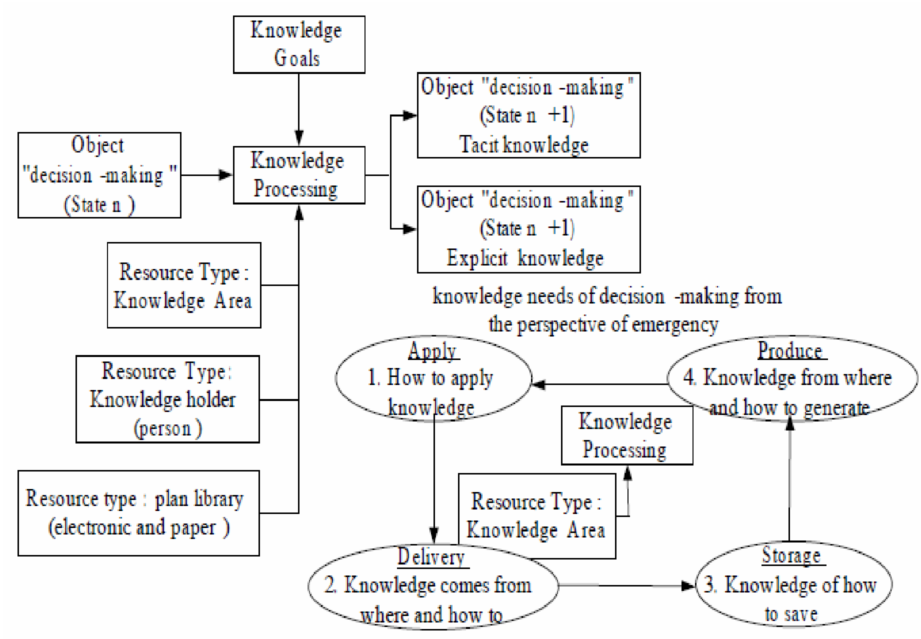


Fig. 3. Demand View analysis of Emergency Decision-Making Process

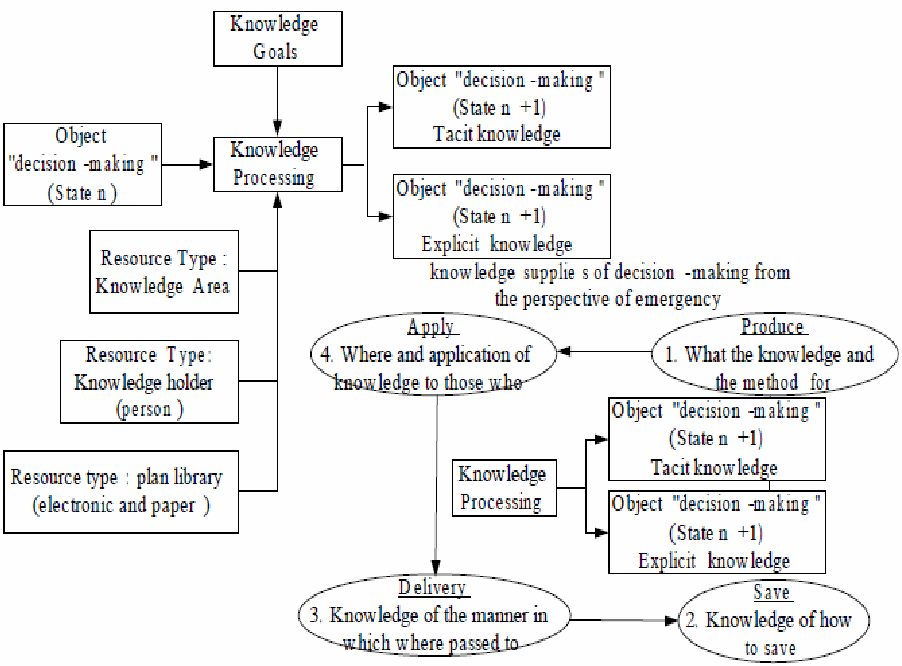


Fig. 4. Supply View Analysis of the Emergency Decision-Making Process

4 Typhoon Case Studies

Typhoon event of natural disasters, a typical class of unexpected incidents, and its incident impacts is extensive, evolution mechanism is relatively complex, triggered by the typhoon events and other unexpected events are rather broad, thus emergency decision-making model based on knowledge supply and demand is applied to knowledge-oriented analysis of the emergency decision-making process.

Know-Contexts: Typhoon events can trigger flooding events and heavy rainfall events, this event will lead to water logging, flash flood disaster events, landslide events, collapse events and debris flow events. Corresponding laws and regulations: Flood laws, flood control regulations, Dam Safety Management regulations, and weather management regulations, the armed forces to participate in rescue and disaster relief regulations. Appropriate measures: emergency assistance and resettlement measures, emergency safeguards and protection measures, emergency control and prohibition.

Know-Decision: Identify hazards, identify risk areas, identify the emergency event may trigger the types and levels; to determine the probability of an emergency incident, to study the consequences of an emergency nature and characteristics of the incident; identify early warning levels: red alert, orange alert, yellow and blue early-warning alert.

Know-What and Know-Why : Typhoon class, central pressure, spiral cloud range, landing sites, typhoon path, wind speed, intensity; disaster: The disaster-risk areas, direct economic losses, injuries, deaths, missing the number of collapsed houses, relocated the number of personnel, industrial The number of businesses damaged; warning Region: The early-warning district staff density and degree of early warning regional economic prosperity, social and public sense of crisis.

Know-Who and Know-How: The typhoon emergency response process, all participants in the typhoon prevention and control headquarters under the unified leadership, mutual communication, coordination and collaboration, the top emergency response participants to provide the necessary knowledge to support decision-making.

Knowledge-driven analysis of the emergency decision-making process: The flood control department, for example, according to the specific needs of emergency response it needs the financial sector for increased funding for flood control; the need to strengthen linkages with the health sector, exchange information epidemic information, and to prevent similar cases in from occurring; need to adjust to the public security departments do a good job blocking the work of affected areas to assist disaster-stricken areas for security and social security management; need to communicate with the transport sector, giving priority to emergency disaster supplies transported; needs and do a good job with the scientific knowledge of the propaganda departments propaganda. Thus, in a flood of reaction process, all participants should be in accordance with emergency incident situations and command requirements associated with specific emergency response departments to communicate and collaborate. Communication and collaboration in this process, an important element is to provide decision makers with the necessary decision-making knowledge. (Michael M. Richter 2009).

5 Summaries

Emergency management is the core of effective decision-making to emergencies, knowledge is making the most fundamental and important resource. Emergency decision-making and decision-making complexity of the object environment is not easy to control the rapidly changing, in order for the damage caused by unexpected events to minimize the decision-makers must be in a limited period of time to make major decision-making; the situation more serious, policy-making persons in the shortest possible time, the greater the need to control the spread of events; emergency incident sudden, destructive disorder and uncertainty in decision-making situation, a comprehensive knowledge, information and a wealth of experience, so an emergency decision-making needs of a knowledge-oriented, knowledge of supply and demand from the perspective of emergency decision-making process to be analyzed continuously adjust and improve decision-making plans, and promptly correct any deficiencies in decision-making process, decision-making fast and accurate completion of the accumulation of knowledge.

As the emergency management involves a multi-disciplinary, multi-field, and faced with a small amount of information and complex environment, useful information quickly and accurately grasp the crucial decision-making, emergency management information, organization, decision-making, plan preparation, emergency response system and other issues will be related to knowledge management issues, this paper presents the emergency supply and demand of knowledge-based decision-making knowledge model, from the knowledge of matching supply and demand point of view of knowledge management processes into emergency decision-making process aimed at improving the emergency response decision-making in the knowledge-oriented cooperation, to achieve accurate and auxiliary emergency decision-making. Proven model for further verification, as well as analysis of the activities of each process and evaluation criteria, further research directions

References

1. Huang, W.: Research Knowledge Demand of Emergency Management Based on the Environment Sensation. Journal of Sichuan University (Engineering Science Edition) 39(1), 182–186 (2007)
2. Chung, P.W.H., Cheung, L.: Knowledge-based process management-an approach to handling adaptive workflow. Knowledge-Based System 16(2), 149–163 (2003)
3. Wen, L.: The Study of Ontology for Knowledge Model Based on Emergency Decision-making. Master thesis, Dalian University of Technology (2008)
4. Raghu, T.S., Vinze, A.: A business process context for Knowledge Management. Decision Support Systems 43(1), 1062–1079 (2007)
5. Parsa, S., Parand, F.: Cooperative decision making in a knowledge grid environment. Future Generation Computer Systems 23(3), 932–938 (2007)
6. Zhen, L., Jiang, Z.: The knowledge Supply Model Based on Knowledge Grid. Journal of Shanghai Jiaotong University 41(1), 45–50 (2007)
7. Richter, M.M.: The search for knowledge, contexts, and Case-Based Reasoning. Engineering Applications of Artificial Intelligence 22(3), 3–9 (2009)

The Consistency and Ranking Method Based on Comparison Linguistic Variable

Qisheng Zhao[1], Fajie Wei[1], and Shenghan Zhou[2]

[1] Economics and Management School of Beihang University
[2] Dept. of Systems Engineering of Engineering Technology of Beihang University
100191, Beijing, China
Zhou.SH@buaa.edu.cn

Abstract. The study developed a consistency approximation and ranking method based on the comparison Linguistic variable. The method constructs the consistency fuzzy complementary judgment matrix by using the judgment matrix of linguistic variable. The judgment matrix is defined by the fuzzy set or vague set of comparison linguistic variable. The method obtains the VPIS and VNIS based on TOPSIS method. And the relative similar approach degrees with the distance between alternatives and VPIS or VNIS are defined. Then the study analyzes the impact on quality of evaluation which caused by evaluation method, index weight and appraiser. Finally, the improving methods were discussed, and an example is presented to illustrate the proposed method.

Keywords: comparison linguistic variable; vague set; relative similar approach degree; fuzzy judgment matrix.

1 Introduction

The linguistic variable concept as defined by Zadeh (1975) plays an important role in fuzzy theory[1]. The concept is widely used in evaluation theory.And the Linguistic Variable is defined by the form LV=<V,T,U,G,M>[2]. Hwang and Yoon (1981) developed the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) to solve the Fuzzy MADM problem [3]. As a decision tool combining quantity and quality, the methodology of the TOPSIS is basically to Integrate the different opinions of many valuators. It is difficult for a decision-maker to give two alternatives a definite comparison judgment value because of the fuzziness and personal limitation. However, it is general feasible to give the judgment scope. Atanassov(1986) developed the concept of Intuitionistic Fuzzy Sets (IFS) [4]. Then Gau and Buehrer(1993) developed the concept Vague set[5]. Bustince and Burillo (1996) believe Vague set be equivalent to IFS in fact[6]. And the vague set can be easily transform to Interval Number, and be the judgment matrix.

It is as common as transforming linguistic variable into triangular fuzzy number, use aberrance coefficient method and get evaluation index weight, calculate relative

M. Zaman et al. (Eds.): CETS 2010, CCIS 113, pp. 318–326, 2010.

approach degree and rank each alternatives. Zhou (2005) improved method by transforming linguistic variable into vague set, and develop the TOPSIS method based on vague set [7]. According to the element constitution of the judgment matrix, it is usually composed of two kinds: One kind is the AHP judgment matrix or reciprocal judgment matrix [8,9],another kind is the fuzzy judgment matrix or complementary judgment matrix[10,11,12]. At present, the reciprocal judgment matrix based on interval number has received the domestic and foreign scholar's attention and the related interval number reciprocal judgment matrix ordering theory have also made some progress[13,14], but the research about the interval number complementary judgment matrix does relative few.

The paper put the method which determining the ordering weight in vague set judgment matrix into interval number complementary judgment matrix, and proposes a method to determining the ordering weight in interval number complementary judgment matrix based on the error analysis theory. Then the relative similar approach degree is defined, and a method for ranking vague set judgment matrix is presented based on the degree. Finally, an example to apply our approach is presented.

2 Basic Definitions

2.1 Vague Set

Definition 1: Set U is the universe of discourse, the element x belong to U, the vague set in U is a couple membership function $t_A(x), f_A(x)$. The membership function meet the following requirements:

1) $t_A(x) \to [0,1]$, $f_A(x) \to [0,1]$;
2) $0 \le t_A(x) + f_A(x) \le 1$

The $t_A(x)$ is called true membership function, the $f_A(x)$ is called false membership function, the $\pi_A(x) = 1 - t_A(x) - f_A(x)$ means the hesitate level.

2.2 Comparison Linguistic Variable and Its Vague Set

Definition 2: Let Comparison linguistic variable P={xj}, xj=[ti, 1-fi], j is odd number and j>1, i=1,2,…,j. the Comparison linguistic variable follow the below rule:

1) If i≥k, then ti≥tk , it means that ti gets the higher degree of membership on xj than tk;
2) The comparison linguistic variable is transformed into triangular fuzzy number, and the interval number is defined as λ-cut sets of triangular fuzzy number. i.e., the triangular fuzzy number of 7 level linguistic variables is defined as Fig. 1.

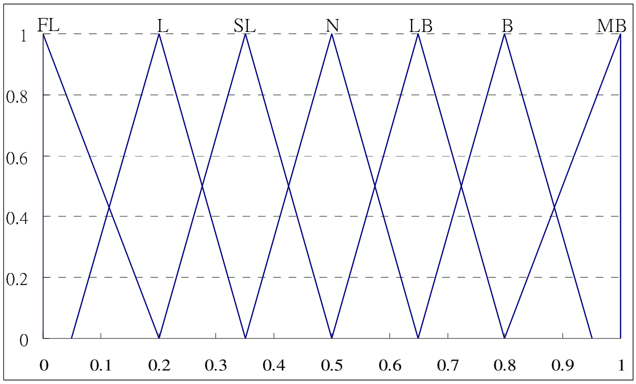


Fig. 1. Transform Linguistic Variable to Triangular Fuzzy Number

3) Abstention part π_{ij} of linguistic variable rise from both ends to midrange, but it is zero in the midpoint. i.e., the interval number of 7 level linguistic variables is given by Table 1(λ=0.5).

Table 1. Triangular Fuzzy Number of 7 Level Grade

Mark	Grade	Triangular Fuzzy Number	Intervalλ=0.5
FL	Far Lower	(0, 0, 0.2)	[0,0.1]
L	Lower	(0.05, 0.2, 0.35)	[0.125,0.275]
SL	Slightly Lower	(0.2, 0.35, 0.5)	[0.275,0.425]
N	Nearly	(0.35, 0.5, 0.65)	[0.425,0.575]
LB	Little Better	(0.5, 0.65, 0.8)	[0.575,0.725]
B	Better	(0.65, 0.8, 0.95)	[0.725,0.875]
MB	Much Better	(0.8,1, 1)	[0.9,1]

4) Set U is the universe of discourse, the linguistic variable x display a couple membership function t_{ij} , f_{ij} , let $a_{ij}=[t_{ij},1-f_{ij}]$, $t_{ij}\to[0,1]$, $f_{ij}\to[0,1]$,$0\le t_{ij}+f_{ij}\le 1$, then the vague set in U, the vague set of 7 level comparison linguistic variables is defined as Table 2.

5) Abstention part π_{ij} of linguistic variable rise from both ends to midrange, but it is zero in the midpoint. i.e., the abstention of vague set from 7 level comparison linguistic variables is shown in Fig.2.

Table 2. Triangular Fuzzy Number of 7 Level Grade

Mark	Linguistic variable	Vague Set	Abstention
FL	Far Lower	[0.1,0.2]	0.1
L	Lower	[0.2,0.4]	0.2
SL	Slightly Lower	[0.3,0.6]	0. 3
N	Nearly	[0.5,0.5]	0
LB	Little Better	[0.4,0.7]	0. 3
B	Better	[0.6,0.8]	0.2
MB	Much Better	[0.8,0.9]	0.1

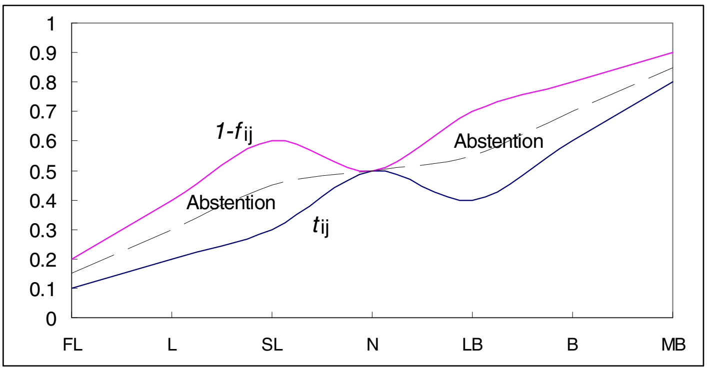


Fig. 2. Abstention of vague set from comparison Linguistic

Definition 3: Let $S=\{S_i | i\in I, i=1,2,\cdots n\}$ be a discrete set of alternatives and binary comparison matrix $A=(a_{ij})_{n\times n}$ be a fuzzy subset of $S\times S$, where a_{ij} denote the degree that S_i better than S_j, $0\le a_{ij}\le 1$, if a_{ij} meet the following conditions:

1) $a_{ii}=0.5$, for any $i\in I$;

2) $a_{ij}+a_{ji}=1$, for any i and $j\in I, i\neq j$ 。

Then matrix A is called fuzzy judgment matrix. Fuzzy judgment matrix $A=(a_{ij})_{n\times n}$. If $a_{ij}=a_{ik}-a_{jk}+0.5$, $i,j,k\in I$, then matrix A is called consistency complementary judgment matrix.

Miller (1956) believes the human beings have a limited capacity to work simultaneously with more than nine categories [15]. So the most common used linguistic variable is defined no more than 9 levels.

3 Consistency Approximation of Vague Set Judgment Matrix

Let A be complementary judgment matrix, $W=(w_1,w_2,\dots w_n)^T$ be ranking weight vector of A, where $\sum_{i=1}^{n} w_i=1$ .if $A=(a_{ij})_{n\times n}$ be a consistency complementary judgment matrix, then:

$$a_{ij}=a_{ik}-a_{jk}+0.5,\ \text{for any } i,j,k\in I \text{ and } a_{ij}=a(w_i-w_j)+0.5 \tag{1}$$

According to the theory, the study choose $a=(n-1)/2$, it means that decision maker pay higher attention to the differences between alternatives.

If A is a consistency complementary judgment matrix, then

$$w_i=\frac{1}{n}-\frac{1}{2a}+\frac{1}{na}\sum_{l=1}^{n}a_{il},\ \mathrm{i}\in I \tag{2}$$

And if A is inconsistent in the practical application, the degree that alternative S_i better than alternative S_j can be given as follow:

$$a_{ij}^{*}=\frac{1}{n}\sum_{k=1}^{n}(a_{ik}-a_{jk}+0.5),\forall i,j,\in I \tag{3}$$

Therefore, let $A=(a_{ij})_{n\times n}$ be vague set judgment matrix, where $a_{ij}=[t_{ij}^{-},1-f_{ij}^{+}]$, $p_{ij}=\frac{1}{n}\sum_{k=1}^{n}\left((t_{ik}^{-}\cdot(1-f_{ik}^{+}))^{\frac{1}{2}}-(t_{jk}^{-}\cdot(1-f_{jk}^{+}))^{\frac{1}{2}}+0.5\right)$, $\forall i,j\in I$,thus $P=(p_{ij})_{n\times n}$ be a consistency complementary judgment matrix, its ranking weight vector is $W=(w_1,w_2,\dots w_n)^T$, where:

$$w_i=\frac{1}{n}-\frac{1}{2a}+\frac{1}{n^2a}\sum_{j=1}^{n}\sum_{k=1}^{n}\left((t_{ik}^{-}\cdot(1-f_{ik}^{+}))^{\frac{1}{2}}-(t_{jk}^{-}\cdot(1-f_{jk}^{+}))^{\frac{1}{2}}+0.5\right),i\in I \tag{4}$$

Its ranking weight vector is given by formula (5)

$$w_i=\frac{1}{n}-\frac{1}{2a}+\frac{1}{na}\sum_{j=1}^{n}p_{ij}=\frac{1}{n}-\frac{1}{2a}+\frac{1}{n^2a}\sum_{j=1}^{n}\sum_{k=1}^{n}\left((t_{ik}^{-}\cdot(1-f_{ik}^{+}))^{\frac{1}{2}}-(t_{jk}^{-}\cdot(1-f_{jk}^{+}))^{\frac{1}{2}}+0.5\right),i\in I \tag{5}$$

4 Evaluation on Vague Set Judgment Matrix

Set VPIS is the virtual alternative which is given the best evaluation, and the VNIS is the virtual alternative which is given the worst evaluation. So the VPIS has the higher level linguistic variable than all alternative. The study maximizes any comparison linguistic variables by columns.

The distance between the alternatives and VPIS or VNIS is defined as [16]:

$$d^* = \sum_{j=1}^{n} w_j M(a_{ij}, VPIS), d^- = \sum_{j=1}^{n} w_j M(a_{ij}, VNIS) \tag{6}$$

Where

$$M(x,y) = 1 - \frac{|t_x - t_y - f_x + f_y|}{8} - \frac{|t_x - t_y + f_x - f_y|}{4} - |\frac{t_x - t_y| + |f_x - f_y|}{8} \tag{7}$$

The relative similar approach degree is defined as:

$$\sigma(Ai) = d^- / (d^* + d^-) \tag{8}$$

The procedure of algorithm is to:

Step1: the experts give the comparison linguistic variables to all alternatives under particular criteria;

Step2: transform the linguistic variables to vague set;

Step3: calculate the consistency approximation matrix of vague set Judgment Matrix;

Step4: find the VPIS and VNIS of all alternatives;

Step5: calculate relative similar approach degree with the distance between alternatives and VPIS or VNIS;

Step6: sort alternatives according to the relative similar approach degree.

5 Illustrative Example

Assuming that a decision-making problem, the experts give comparison to four alternatives under particular criteria, judgment information is as follows:

$$E = \begin{bmatrix} \mathrm{N} & \mathrm{SL} & \mathrm{B} & \mathrm{LB} \\ \mathrm{LB} & \mathrm{N} & \mathrm{LB} & \mathrm{SL} \\ \mathrm{L} & \mathrm{LB} & \mathrm{N} & \mathrm{L} \\ \mathrm{SL} & \mathrm{B} & \mathrm{B} & \mathrm{N} \end{bmatrix}$$

5.1 Ranking Method of Triangular Fuzzy Number

Firstly, the fuzzy Judgment Matrix A is:

$$A=\begin{bmatrix} [0.425,0.575] & [0.275,0.425] & [0.725,0.875] & [0.575,0.725] \\ [0.575,0.725] & [0.425,0.575] & [0.575,0.725] & [0.275,0.425] \\ [0.125,0.275] & [0.575,0.725] & [0.425,0.575] & [0.125,0.275] \\ [0.275,0.425] & [0.725,0.875] & [0.725,0.875] & [0.425,0.575] \end{bmatrix}$$

Then, the consistency approximation matrix of P is given by:

$$P=\begin{bmatrix} 0.500 & 0.557 & 0.603 & 0.443 \\ 0.443 & 0.500 & 0.546 & 0.386 \\ 0.397 & 0.454 & 0.500 & 0.339 \\ 0.557 & 0.614 & 0.661 & 0.500 \end{bmatrix}$$

The consistency weighted vector is given by formula (5):

$$W=(0.267,0.229,0.198,0305)$$

After that, the study gets the VPIS and VNIS. The best rank evaluation was selected as VPIS, and the worst rank evaluation selected as VNIS.

$$VPIS=([0.575,0.725],[0.725,0.875],[0.725,0.875],[0.575,0.725])$$

$$VNIS=([0.125,0.275],[0.275,0.425],[0.425,0.575],[0.125,0.275])$$

Finally, the relative similar approach degree is given by formula (7) and (8):

Table 3. Relative Similar Approach Degree (Triangular Fuzzy Number)

	A1	A2	A3	A4
d_i^*	0.8005	0.8077	0.7662	0.7926
d_i^-	0.7983	0.8138	0.8010	0.7817
$\sigma(Ai)$	0.4993	0.5019	0.5111	0.4965

The sort of four alternatives is $A3 \succ A2 \succ A1 \succ A4$ according to relative similar approach degree $\sigma(Ai)$.

5.2 Ranking Method of Vague Set

Firstly, the vague set Judgment Matrix A is:

$$A=\begin{bmatrix} [0.5,0.5] & [0.3,0.6] & [0.6,0.8] & [0.4,0.7] \\ [0.4,0.7] & [0.5,0.5] & [0.4,0.7] & [0.3,0.6] \\ [0.2,0.4] & [0.4,0.7] & [0.5,0.5] & [0.2,0.4] \\ [0.3,0.6] & [0.6,0.8] & [0.6,0.8] & [0.5,0.5] \end{bmatrix}$$

Then, the consistency approximation matrix of P is given by:

$$p = \begin{bmatrix} 0.500 & 0.538 & 0.563 & 0.462 \\ 0.462 & 0.500 & 0.525 & 0.424 \\ 0.438 & 0.476 & 0.500 & 0.400 \\ 0.538 & 0.576 & 0.601 & 0.500 \end{bmatrix}$$

The consistency weighted vector is given by formula (5):

$$W = (0.26, 0.235, 0.219, 0.286)$$

After that, the study gets the VPIS and VNIS. The best rank evaluation was selected as VPIS, and the worst rank evaluation selected as VNIS.

$$VPIS = ([0.4,0.7],[0.6,0.8],[0.6,0.8],[0.4,0.7])$$

$$VNIS = ([0.2,0.4],[0.3,0.6],[0.5,0.5],[0.2,0.4])$$

Finally, the relative similar approach degree is given by formula (7) and (8):

Table 4. Relative Similar Approach Degree (Triangular Fuzzy Number)

	A1	A2	A3	A4
d_i^*	0.7865	0.7911	0.8229	0.7897
d_i^-	0.8696	0.9088	0.9102	0.8571
$\sigma(Ai)$	0.5251	0.5346	0.5252	0.5205

The sort of four alternatives is $A2 \succ A3 \succ A1 \succ A4$ according to relative similar approach degree $\sigma(Ai)$.

6 Discussions

The paper gives a method for determining the ranking weight vector based on comparison linguistic variable and its judgment matrix. The illustrative example gives the comparison result of triangular fuzzy number and vague set based on linguistic variable. The result shows the difference between two methods. The rank order of four alternatives A2 and A3 has some difference. De Soto(1999) analyze the linguistic variable in different way[17], and give the axiomatization definitions of transforming linguistic variable to fuzzy sets. And they find the linguistic variable dependent on the weight assigned by supervisor [18]. It means that there are different results based on different transform weight.

To the example in this paper, vague set is more suitable for practical situation on empirical. The method takes the abstention into account, and gets more reliability. It obtains the consistency approaching matrix by making full use of the information in vague set. And the study provides a new way to find the truth in fuzzy judgment information. The method integrates the theory of vague set, Ranking research and the comprehensive application of some kind of the TOPSIS method. It means that there are many areas need further explore.

References

1. Zadeh, L.A.: The Concept of a Linguistic Variable and Its Application to Approximate Reasoning. Information Sciences 8(3), 199–249 (1975)
2. Zadeh, L.A.: The Concept of a Linguistic Variable and Its Application to Approximate Reasoning. part I, II and III. In: Zadeh, L.A., Yager, R., Ovchinnikov, S., Tong, R., Nguyen, H. (eds.) Fuzzy Sets and Applications: Selected Papers, pp. 219–366. John Wiley & Sons, Chichester (1987)
3. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making Methods and Applications, pp. 12–23. Springer, Heidelberg (1981)
4. Atanassov, K.: Intuitionist Fuzzy Sets. Fuzzy Sets and Systems 20, 87–96 (1986)
5. Gau, W.L., Buehrer, D.J.: Vague Sets. IEEE Transactions On Systems, Man, and Cybernetics 23, 610–614 (1993)
6. Bustince, H., Burillo, P.: Vague Sets are Intuitionistic Fuzzy Sets. Fuzzy Sets and Systems 23(2), 403–405 (1996)
7. Zhou, X.: Research on TOPSIS Methods based on Vague Set Theory. Systems Engineering 14(1), 537–541 (2005)
8. Hwang, C.L., Lin, M.J.: Group Decision-making under Multiple Criteria. Springer, Berlin (1987)
9. Ma, J., Fan, Z., Huang, L.: A Subjective and Objective Integrated Approach to Determine Attribute Weights. European Journal of Operational Research 112(2), 397–404 (1999)
10. Chen, S.J., Hwang, C.L.: Fuzzy Multiple Attribute Decision Making. Springer, Berlin (1992)
11. Tanino, T.: Fuzzy Preference Orderings in Group Decision Making. Fuzzy Sets and Systems 12, 117–131 (1984)
12. Orlorski, S.A.: Decision-making with a Fuzzy Preference Relation. Fuzzy Sets and Systems 1, 155–167 (1978)
13. Sajjad, M.: Incorporating the Uncertainty of Decision Judgments in the Analytic Hierarchy Process. European Journal of Operational Research 53(2), 206–216 (1991)
14. Lu, Y.: Weight Calculation Method of Fuzzy Analytical-hierarchy Process. Fuzzy Systems and Mathematics 16(2), 79–85 (2002)
15. Miller, G.A.: The Magical Number Seven or Minus Two: Some Limits on Our Capacity of Processing Information. Psychological Rev. 63, 81–87 (1956)
16. Zhou, Y., Wei, F.: A Consistency Approximation and Weight Calculation Method of the Interval Complementary Judgment Matrix. Operations Research and Management Sciences 15(2), 27–31 (2006)
17. De Soto, A.R., Trillas, E.: Second Thoughts on Linguistic Variables. Fuzzy Information Processing Society (1999)
18. De Soto, A.R.: Hierarchical Linguistic Variables. Fuzzy Information Processing Society (2009)

Author Index

Cao, Yuheng 159
Chang, Guiran 61
Chen, Chen 192
Chen, Gang 1
Cheng, Wei 61
Chen, Hua 234
Chhikara, Sudesh 292
Chu, Hsing-Hui 281

Dai, Shiqiang 200
Dong, Chuansheng 258

Esmaeilzadeh, Pouyan 23

Fukami, Yoshiaki 269

Gang, Chen 86

Hao, Xiaoling 180
He, Yi 226
Hou, Juchi 238
Hu, Yuanhua 171

Isshiki, Masao 269

Ji, Guojun 140
Jin, Xia 130

Kokuryo, Jiro 269
Kruba, Steve 245
Kumar, Naresh 23

Liang, Ying 238
Li, Ban 226
Liu, Heng 124
Liu, Hongwei 215
Liu, Ximin 200
Li, Wen 38
Li, Yancang 75, 105
Lu, Shoufeng 200
Lu, Ta-Jung 281

Ma, Jianhua 114
Mann, Inder Jit Singh 159
Meyer, Jim 245

Nan, Lin 46
Nsakanda, Aaron L. 159

Ohmukai, Ikki 269

Peng, Jiao 124

Qian, Wu 1
Qin, Zheng 192

Sambasivan, Murali 23
Shenghua, Zheng 226
Shi, Zhao 1
Shivanand M., Handigund 93
Shweta, Bhat 93
Suo, Juanjuan 75

Takeda, Hideaki 269
Tang, Bin 38
Tan, Zhenhua 61
Tian, Bo 46

Wang, Lidong 124
Wann, Jong-Wen 281
Wei, Fajie 318
Wei, Jianli 238
Wu, Bing 305

Xie, Kefan 1
Xing, Jian-kuan 192
Xue, Lili 171

Yang, Guangming 61
Yang, Lei 46
Yin, Zhixiang 234
Yuan, Jie 114

Zhao, Lina 105
Zhao, Lin-du 305
Zhao, Qisheng 318
Zheng, Qin 46
Zheng, Shi 130
Zheng, Wen 130
Zhong, Yinghong 215
Zhou, Shenghan 318
Zhou, Shujing 75, 105
Zhou, Shumin 38
Zhu, Zhiliang 61