Junzo Watada
Toyohide Watanabe
Gloria Phillips-Wren
Robert J. Howlett
Lakhmi C. Jain
Editors



Intelligent Decision Technologies

Proceedings of the 4th International Conference on Intelligent Decision Technologies (IDT´2012) - Volume 2

SMART INNOVATION, SYSTEMS AND TECHNOLOGIES 16

Springer

Smart Innovation, Systems and Technologies 16

Editors-in-Chief

For further volumes:
http://www.springer.com/series/8767

Junzo Watada, Toyohide Watanabe,
Gloria Phillips-Wren, Robert J. Howlett,
and Lakhmi C. Jain (Eds.)

Intelligent Decision Technologies

Proceedings of the 4th International Conference on Intelligent Decision Technologies (IDT'2012) – Volume 2

Springer

Editors
Prof. Junzo Watada
Waseda University
Kitakyushu
Japan

Prof. Toyohide Watanabe
Nagoya University
Japan

Prof. Gloria Phillips-Wren
Loyola University Maryland
Baltimore, MD
USA

Prof. Robert J. Howlett
KES International
Shoreham-by-sea
United Kingdom

Prof. Lakhmi C. Jain
University of South Australia
Adelaide
Australia

ISSN 2190-3018
e-ISSN 2190-3026
ISBN 978-3-642-29919-3
e-ISBN 978-3-642-29920-9
DOI 10.1007/978-3-642-29920-9
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012937732

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The Intelligent Decision Technologies (IDT) International Conference encourages an interchange of research on intelligent systems and intelligent technologies that enhance or improve decision making. The conference is organized by KES International, a research community consisting of several thousand research scientists. The focus of IDT is interdisciplinary and includes research on all aspects of intelligent decision technologies, from fundamental development to real applications.

Advances in Artificial Intelligence (AI) and computing environments that can deliver intelligent technologies effectively have enabled an explosion in intelligent applications. IDT have the potential to expand their support of decision making in such areas as finance, accounting, marketing, healthcare, medical and diagnostic systems, military decisions, production and operation, networks, traffic management, crisis response, human-machine interfaces, financial and stock market monitoring and prediction, and robotics. Intelligent decision systems implement advances in intelligent agents, fuzzy logic, multi-agent systems, artificial neural networks, and genetic algorithms, among others. Emerging areas of active research include virtual decision environments, social networking, 3D human-machine interfaces, cognitive interfaces, collaborative systems, intelligent web mining, e-commerce, e-learning, e-business, bioinformatics, evolvable systems, virtual humans, and designer drugs.

In this volume we publish research from the Fourth KES International Symposium on Intelligent Decision Technologies (KES IDT'12), hosted and organized by researchers in Japan. The conference was held in Gifu City located in the center of Japan. Gifu City is known for its traditions and rich history, including its 1300-year-old tradition of cormorant fishing on the Nagara River and Gifu Castle. This book contains chapters based on papers selected from a large number of submissions for consideration for the symposium from the international community. Each paper was double-blind, peer-reviewed by at least two independent referees. The best papers were accepted based on recommendations of the reviewers and after required revisions had been undertaken by the authors. The final publication represents the current leading thought in intelligent decision technologies.

We wish to express our sincere gratitude to the plenary speakers, invited session chairs, delegates from all over the world, authors and reviewers for their outstanding contributions. We express our sincere thanks to the Japanese organizers and Gifu City for their sponsorship and support of the symposium. We thank the International Programme Committee for their support and assistance. We would like to thank Peter Cushion of KES International for his help with organizational issues. Also our appreciation is extended to their contribution provided by the KES International Secretariat Team. We thank the editorial team of Springer-Verlag for their support in production of this volume. We sincerely thank the Local Organizing Committee, especially Professors K. Asakura in Daido University and T. Kojiri in Kansai University for their editing conference record and program, and their collaborators for their invaluable contributions.

We hope and believe that this volume will contribute to ideas for novel research and advancement in intelligent decision technologies for researchers, practitioners, professors and research students who are interested in knowledge-based and intelligent engineering systems. At the end we will express thanks for Dr. Jerzy Michnik, University of Economics in Katowice, Poland, for his help on Latex.

Gifu, Japan
May 22–25, 2012

Junzo Watada
Toyohide Watanabe
Gloria Phillips-Wren
Robert J. Howlet
Lakhmi C. Jain

Editors

Professor Junzo Watada received his B.Sc. and M.Sc. degrees in electrical engineering from Osaka City University, and Dr. of Eng. Degree from Osaka Prefecture University, Japan. He is currently a Professor of Knowledge Engineering, Soft Computing and Management Engineering at the Graduate School of Information, Production & Systems, Waseda University, after a professor of Human Informatics and Knowledge Engineering at the Osaka Institute of Technology, Japan and was with Ryukoku University, Kyoto for about 10 years each. Before moving to Academia, he was with Fujitsu Co. Ltd. for about 10 years. His research interests includes decision making tehnologies and management of technology. Dr. Watada is currently the President of International Society of Management Engineers. He was a co-chair of KES-IDT2011 at Pireus, Greek and plays an active role in editing International Journal of Intelligent Decision Technology (KES official journal) as a co-editor-in-chief.

Toyohide Watanabe has received B.S., M.E. and Dr.Eng. from Kyoto University, in 1972, 1974 and 1985, respectively. Since 1975, he has worked in Kyoto University and Nagoya University: first as a research associate in Data Processing Center, Kyoto University, from 1975 to 1987; second as an associate professor Faculty of Engineering, Nagoya University, from 1987 to 1994; third as a full professor in the same faculty, from 1994 to 1997; fourth as a professor in Graduate School of Engineering, Nagoya University from 1997 to 2003; and finally as a professor in Department of Systems and Social Informatics, Graduate School of Information Science, Nagoya University, since 2003. Additionally, from 2004 to 2008 he was the head director in Information Technology Center, Nagoya University. The current topics of his research interests are: Intelligent Tutoring System, Computer-supported collaborative learning, Knowledge management, Intelligent activity-support, etc. He is a member of ACM, IEEE-CS, AAAI, AACE, KES International, IEICE of Japan, IPS of Japan, IEE of Japan, Japan of SAI, Japan of SSST, JSiSE, etc. Also, currently he is an Editor-in-Chief of the International Journal of Knowledge and Web Intelligence. Moreover, he is a fellow on IEICE of Japan since 2004.

Dr. Gloria Phillips-Wren is Professor and Chair of Information Systems and Operations Management at Loyola University Maryland, and Academic Director of Executive Programs. She is co-founder and co-editor-in-chief of Intelligent Decision Technologies International Journal (IDT), Vice Chair and Chair-elect of SIGDSS under the auspices of the Association for Information Systems (AIS), Secretary of IFIP WG8.3 DSS, and leader of the focus group in Intelligent Decision Technologies (IDT) for KES International. She is a co-organizer of the IDT Conference Series under the auspices of KES International. She received a PhD from the University of Maryland Baltimore County and holds MS and MBA degrees. Her research interests and publications are in intelligent decision support systems, intelligent agents, decision making, analytics, business intelligence, data mining, and emerging technologies such as social media. She has published in the area of decision making and support in academic journals including the European Journal of Operational Research, Omega, Expert Systems with Applications, and the Journal of Network and Computer Applications. Her most recent book (co-edited) was published in 2011 and is entitled Intelligent Decision Technologies.

Robert Howlett is the Executive Chair and Chief Executive of the KES (Knowledge-Based and Intelligent Engineering Systems) International. KES is an associaton dedicated to knowledge exchange and the dissemination of research results in the areas of intelligent systems, sustainability in energy and buildings and innovation through knowledge transfer. He is a Director of the Institute of Knowledge Transfer, the body for knowledge professionals and researchers working in innovation, knowledge transfer and enterprise. He holds a Visiting Professorship in Enterprise at Bournemouth

University. He holds a PhD in Intelligent Systems from the University of Brighton, an MPhil from Sussex University and a BSc(Hons) from Portsmouth Polytechnic. He is a Chartered Engineer and a Chartered Information Technology Practitioner. He was at the University of Brighton for over 20 years where he headed a research group in intelligent systems and the was Director of the Knowledge Transfer Partnerships (KTP) Centre. For a number of years he chaired the KTP national Forum, representing KTP centres from all universities in the country, which he formed. He has been involved with innovation knowledge transfer, as a practitioner, facilitator and manager, for over 15 years. He has personally supervised many knowledge transfer projects and written and mentored many more.

Professor Lakhmi Jain is a Director/Founder of the Knowledge-Based Intelligent Engineering Systems Centre. His interests focus on the novel techniques such as knowledge-based intelligent machines, virtual systems, intelligent agents, and the application of these techniques in areas such as engineering, science, aviation, healthcare, defence and so on.

Contents

Part I: Applications of Intelligent Decision Technology

Part II: Other Related Topics

Part I

Applications of Intelligent Decision Technology

Evaluating Top Services-Prepackaged Software Firms in Standard and Poor's 500 Index by Using a Multiple Objective Programming Based Data Envelopment Analysis

Chi-Yo Huang, Po-Yen Wang, and Gwo-Hshiung Tzeng

Abstract. Services-prepackaged software firms are a field of Information technology (IT). IT is defined as the obtainment, procedure, storage and propagation of sounding, drawing, and textual information by combining microelectronics-based computing and telecommunications. Nowadays, IT has penetrated in daily life of human beings and become one part of the whole society. The importance of IT has become momentous. Therefore, to understand the performance of efficiency and productivity of the IT firms is critical for managers as well as for personal investors. Until now, there are very few researches tried to analyze final performance of the services-prepackaged software firms in IT sector. As a result, this research intends to use traditional Data Envelopment Analysis (DEA) CCR or BCC models to evaluate the performance of the services-prepackaged software firms. However, the traditional DEA models are not fair models from the aspect of improper weight derivations. Thus, this paper intends to analyze the efficiency of the services-prepackaged software firms by using multiple objective programming (MOP) based DEA. The Decision Making Units (DMUs) on this research are chosen from the services-prepackaged software firms in S&P 500 based on publicly available financial reports of the fiscal year 2010. In a MOP based DEA approach, DMUs will be evaluated based on an equal standard and the results will be evaluated more fairly. In the empirical study, the MOP based DEA demonstrated that Autodesk Inc.,

Chi-Yo Huang · Po-Yen Wang
Department of Industrial Education, National Taiwan Normal University
No. 162, Hoping East Road I, Taipei 106, Taiwan
e-mail: cyhuang66@ntnu.edu.tw

Gwo-Hshiung Tzeng
Department of Business and Entrepreneurial Administration, Kainan University
No. 1, Kainan Road, Luchu, Taoyuan County 338, Taiwan

Institute of Management of Technology, National Chiao Tung University
Ta-Hsuch Road, Hsinchu 300, Taiwan
e-mail: ghtzeng@mail.knu.edu.tw

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 3–12.
springerlink.com

BMC Software, and Citrix Systems should be the services-prepackaged software firms worthwhile to be invested. In the future, performance evaluation results can be served as foundations for investment strategies definition.

Keywords: Information Technology (IT), Standard and Poor's 500 index, Performance Evaluation, Data Envelopment Analysis (DEA), Multiple Objective Programming (MOP).

1 Introduction

In the twenty-first century, information technology (IT) has penetrated in daily life of human beings and become one part of the whole society. IT was first appeared by Harvard Business Review in 1958 since this new technology does not have a unique established name. IT depends on its reshaping by the basics of business and customer service, operations, product and marketing strategies, and distribution entirely (Keen, 1991). Also, IT plays an important role in knowledge-based economy around the world nowadays. The appearance of information technologies, especially in internet fields, has changed everything in whole world's business; it has, especially, affected the developing, marketing, and distributing procedures of products. Thus, obtaining a competitive advantage over competitors is the required precondition to survive in rival global market nowadays. The field of IT industry is broad. By using Standard Industrial Classification (SIC) code which is a United States government system for classifying industries by a four-digit code to choose IT firms, it shows that every company have different major products in its field. The author selects the services-prepackaged software firms to evaluate the performance. The reasons for selecting this classification of IT firms are mentioned in the following. Jorgenson (2001) also indicated that the investment in software is growing fast and much more critical than computer hardware investment. Besides, Parker & Grimm (2000) described the new estimation of investment in software.

Top managers in most organizations always want to know whether investment into IT has any linkage to the overall performance of the firm. There have been some researches concerning the performance of IT by using different methods. Therefore, performance evaluation plays an essential role in management process. It not only provides critical information for decision-making, but also gives a foreseeable advantage for following operations (Phillips, 1999). Hence, how to use performance evaluation to measure organizational performance in a multidimensional construct is important to determine a whole management procedure (Lewin and Minton, 1986).The data envelopment analysis (DEA) is a flexible tool and can be molded with other analytical methods. DEA is a mathematical programming approach to evaluate the relative efficiency of peer units of multiple performance measures (Charnes et al., 1978; Cooper et al., 2004).Thus, DEA methods have become popular tools, which were widely adopted on national, industrial as well as firm level performance evaluations.

Although DEA approaches were widely adopted on performance evaluations of nations, industries as well as firms, such performance evaluation results were derived based on different bases of comparisons of DMUs. Nevertheless, traditional

DEA approaches which were usually leveraged on firms' performances are based on the unfair weights' problems as mentioned by Fare and Hunsaker (1986). Apparently, to resolve the disadvantages being introduced by traditional DEA approaches, an appropriate measure of the IT by a suitable method is required. To resolve the above mentioned disadvantages being introduced by traditional DEA approaches, this research aims to introduce a Multiple Objectives Programming (MOP) based DEA method being developed by Prof. Gwo-Hshiung Tzeng (Chiang and Tzeng, 200a, 2000b) to evaluate the performance of the top IT firms in S&P 500. In the novel MOP based DEA approach, DMUs will be evaluated based on an equal standard (Chiang and Tzeng, 2000a, 2000b; Liu, Lee and Tzeng, 2004). By this approach, the efficiency rating of each DMU can be evaluated more fairly than the traditional CCR approach being proposed by Charnes et al. (1978).

In this research, the definition of the background, motivations and problem findings of this whole research will be first introduced. Secondly, the related literature according to performance evaluation and efficiency and productivity will be summarized by literature review. Then, the top services-prepackaged software firms in S&P 500 will serve as DMUs. Following, selecting the same inputs and outputs to evaluate the performance by traditional CCR model and then by the MOP based on DEA model. Finally, the differences between two different models and managerial implications will be compared and discussed.

The remainder of this paper is organized as follows. The related literature regarding to performance evaluation and efficiency and productivity will be reviewed in Section 2. The analytic framework based on CCR, BCC, and MOP based DEA methods will be introduced in Section 3. Then, in Section 4, the results of evaluating by CCR and MOP based DEA will be shown as an empirical study. Managerial implications as well as discussion will be presented in Section 5. Finally, the whole article will be concluded in Section 6.

2 Literature Review

In this Section, the author will review some literature with regard to the productivity and efficiency as well as performance evaluation. In the following Sub-sections, the definitions and literature of productivity and efficiency will be reviewed first. Then, the performance evaluation will be reviewed in the following part of this Section.

2.1 Productivity and Efficiency

The definition of efficiency is that the minimum resource level which is abstractly needed to conduct the desired operations in a given system when contrasted to the actual resource has been used (Sumanth, 1994; Tangen, 2005). Based on Coelli et al. (2005), the normal measurement of efficiency is used to measure a firm as the ratio of the outputs, and measure produces as the ration of the inputs. The existence of multiple inputs and outputs in regard to distinct resources, actions and environmental factors (Emrouznejad, 1995) cause the formula is usually insufficient (Coelli et al., 2005). The term "efficiency" concerns about the relationship between the amounts of required input to the amount of produced output.

Productivity is one of our most basic and intuitive measures of performance. At the firm level, productivity is a component of profit growth along with price changes, and at the aggregate level. Productivity is a fundamental part of economic growth and welfare. To determine the productivity of any organization, it is required to have a tool to measure it. Over the time period, measuring productivity would be helpful for the organization to compare the performance towards the industry of similar firms or similar service providers, and to compare the productivity of a certain department (Helms, 2006).

2.2 Performance Evaluation

Performance evaluation is an important part of the procedure of management. Performance evaluation not only provides required information for decision-making, but also gives a rival advantage for following operations (Phillips, 1999). As a result, it is critical to decide how to measure performance more organized in a multidimensional construct (Lewin and Minton, 1986). Among the methods of evaluating organizational performance, the technique of data envelopment analysis (DEA) proposed by Charnes et al. (1978) will be the most representative method of performance evaluation.

The performance evaluation based on DEA method has been widely used in a variety of fields, including airlines, banking, insurance, life insurance companies, telecommunications, transportation companies, textile companies, hotel industry, supplier selection, and high-tech companies.

Skinner (1966) indicated some increases of competitive pressure on manufacturers. Profit, lower production costs, quality design, volume flexibility; and delivery speed are included in these pressures. Therefore, to evaluate the performance of manufactures may be based on the factors above. Wheelwright (1978) pointed out four criteria, including efficiency, dependability, quality and flexibility as the important factors for performance evaluation of manufacturing. Flynn et al. (1994) introduced some factors for evaluating the performance of manufacturing, including manufacturing cost, empowerment in employee, flexibility, and speed. Kasul & Motwani (1995) also introduced several criteria for performance measurements such as, efficiency, productivity, quality, management of material, improved technology, skill control, flexibility, price leadership, and global competitiveness.

3 Research Methods

DEA has been developed for 30 years since Charnes et al. (1978) developed the CCR model. Up to now the DEA methods have been used in various applications including education, health care (hospitals, clinics), agricultural production, banking, armed forces, sports, etc. (Emrouznejad, 1995). Further, at the same time, a lot of advanced models of DEA have been developed (e.g., cross-period data (Fare et al., 1985) and the multiple objective programming approach (Chiang and

Tzeng, 2000a, 2000b. 2003; Yu et al., 2004, 2007; Tsai et al., 2006). In this research, the traditional CCR, BCC DEA models as well as the MOP based on DEA model will be applied to aggregate the efficiency scores of the top information technology (IT) companies in S&P 500 based on the values as weights versus each input and output.

3.1 DEA

DEA is a non-parametric approach and doesn't need assumptions about the inputs and outputs. In 1957, Farrell first introduced how to deal with the problem of measuring the productive efficiency to both the economic theorist and the economic policy-maker (Farrel, 1957). The first DEA model, a mathematical programming model by Charnes, Cooper, and Rhodes in 1978, was built to discuss the efficiency frontier by Farrell (Charnes et al., 1978). The CCR model assumes that production exhibits constant returns to scale. Then, in 1984, Banker, Charnes, and Cooper, extended the CCR model by assuming variable returns to scale and named the new model as the BCC model (Banker et al, 1984).

For company managers, controlling the range of inputs and decreasing inputs is easier than increasing the total sales. The CCR and BCC models of DEA are often used the input-oriented.

3.1.1 CCR

CCR–DEA model computes relative efficiency scores h_k of k^{th} DMU $(k \in \{1,2,...,n\})$ based on selected s outputs $(r=1,...,s)$ and m inputs $(i=1,...,m)$ using the following linear programming expression (Charnes et al., 1978; Charnes et al., 1985; Chiang and Tzeng, 2000b):

$$\text{Max } h_k = \sum_{j=1}^{s} u_j y_{jk} \Big/ \sum_{i=1}^{m} v_i x_{ik} \tag{1}$$

Subject to

$$\sum_{j=1}^{s} u_j y_{jr} \Big/ \sum_{i=1}^{m} v_i x_{ir} \le 1, \quad r=1,2,...,n$$

$$v_i, u_j \ge \varepsilon > 0;\ i=1,2,...,m;\ j=1,2,...,s;\ k,r \in \{1,2,...,n\}$$

In Eq. (1), it assumes the DMU (Decision Making Unit) has s outputs and m inputs, and there are n DMUs. The definition let x_{ik} be the i^{th} input (i =1,2,…,m) and y_{jk} be the j^{th} output (j =1,2,…,s) in kth DMU; the v_i and u_j are not zero, calculating as $v_{i,} u_j \ge \varepsilon > 0$, ε is non-Archimedean number and is 10^{-6} in this paper.

3.1.2 Correct CCR Based on BCC Model

Eq. (1) refers to maximize the ratio of weighted sum of output and input values. Then Charnes et al. (1987) proposed correct CCR model based on input-oriented BBC model (Banker et al., 1984) by Chiang and Tzeng (2000b):

$$\text{Max } h_k = \sum_{j=1}^{s} u_j y_{jk} \Big/ \sum_{i=1}^{m} v_i x_{ik} \tag{2}$$

Subject to

$$\sum_{j=1}^{s} u_j y_{jr} \Big/ \sum_{i=1}^{m} v_i x_{ir} \le 1, \quad r = 1,2,...,n$$

$$u_j \Big/ \sum_{i=1}^{m} v_i x_{ik} \ge \varepsilon > 0, \quad j = 1,2,...,s$$

$$v_j \Big/ \sum_{i=1}^{m} v_i x_{ik} \ge \varepsilon > 0, \quad i = 1,2,...,m$$

Since it is difficult to solve the fractional programming as Eq. (1) and Eq. (2), we transfer Eq. (1) and Eq. (2) to the linear programming by the following transformations:

Assuming $v_i^0 = t \cdot v_i$ (i.e., $v_i = v_i^0 / t$), $u_j^0 = t \cdot u_j$ (i.e., $u_j = u_j^0 / t$), $t^{-1} = \sum_{i=1}^{m} v_i x_{ik}$, then multiply the numerators and denominators in Eq. (1) by t, and add the consistency condition, $t\sum_{i=1}^{m} v_i x_{ik} = 1$. Thus Eq. (2) can transfer to Eq. (3).

$$\text{Max } h_k = \sum_{j=1}^{s} u_j^0 y_{jk} \tag{3}$$

Subject to

$$\sum_{j=1}^{s} u_j^0 y_{jr} - \sum_{i=1}^{m} v_i^0 x_{ir} \le 0, \quad r = 1,2,...,n$$

$$\sum_{i=1}^{m} v_i^0 x_{ik} = 1$$

$$u_j^0 \ge \varepsilon > 0, \quad j = 1,2,...,s$$

$$v_i^0 \ge \varepsilon > 0, \quad i = 1,2,...,m$$

$$v_i^0 = t \cdot v_i$$

$$u_j^0 = t \cdot u_j$$

$$t^{-1} = \sum_{i=1}^{m} v_i x_{ik}$$

The dual problem of Eq. (4) can be written as follows:

$$\text{Min } h_k = \left[\theta_k - \varepsilon\left(\sum_{i=1}^{m} S_{ik}^- + \sum_{j=1}^{s} S_{jk}^+\right)\right] \tag{4}$$

Subject to

$$\theta_k x_{ik} - \sum_{k=1}^{n} \lambda_k x_{ik} - S_{ik}^- = 0, \quad i = 1,2,...,m$$

$$y_{jk} - \sum_{k=1}^{n} \lambda_k y_{jk} + S_{jk}^+ = 0, \quad j = 1,2,...,s$$

$$S_{ik}^-, \ S_{jk}^+, \ \lambda_k \ge 0$$

where S_{ik}^- and S_{jk}^+ are slack variables.

The dual problem, presented by BBC (Banker et al., 1984), has two primary strengths, the reduction of calculation barriers and the provision of more helpful information for decision maker. S_{ik}^- and S_{jk}^+ are slack variables of input criteria and output criteria respectively. When $h_k^* = 1$ an individual DMU_k achieves

Pareto's optimality, where " $*$ " denotes for the optimal solution, for example, $\{(x_{ik}^*, y_{jk}^*) \mid S_{ik}^- = S_{jk}^+ = 0,\ \ i = 1,2,...,m;\ \ j = 1,2,...,s\}$, and $S_{ik}^- = S_{jk}^+ = 0$. If DMU has not achieve Pareto's optimality situation, its limited equation intrinsically includes $x_{ik}^* = \theta^* x_{ik} - S_{ik}^-$ and $y_{jk}^* = y_{jk} + S_{jk}^+$. In order to achieve its efficiency goal of optimality, this specific DMU may either reduce inputs $\Delta x_{ik} = x_{ik} - x_{ik}^*$ or increase output $\Delta y_{ik} = y_{ik}^* - y_{ik}$ to become efficient in the relative efficiency. Clearly, the slack variables analysis of DEA method provides DMU's with related information of range and direction for improvement in ideas. When a DMU does not achieve Pareto's optimality, we can make some improvement or innovation/creativity based on Eq. (4) in consideration. Then this results can help individual DMU how we can achieve Pareto's optimality in the relative efficiency, i.e.,

$$x_{ik}^* = \theta^* x_{ik} - S_{ik}^-,\quad i = 1,2,...,m$$
$$y_{jk}^* = y_{jk} + S_{jk}^+,\quad j = 1,2,...,s$$

3.2 Multiple Objective Programming Based DEA

Multiple objective programming (MOP) based DEA method provides a unitary weight (u^*, v^*) for all DMUs, which are evaluated by considering an equal standard measure (Chiang and Tzeng, 2000a, 2000b; Yu et al., 2004, 2007). By this approach, this research can obtain the efficiency rating of each DMU more fairly. Moreover, all DMUs can be treated simultaneously, which makes it effectiveness in handling large numbers of DMU.

Model 1

$$\left.\begin{aligned}
&\text{Max } z_1 = \sum\nolimits_{j=1}^{s} u_j y_{j1} \Big/ \sum\nolimits_{i=1}^{m} v_i x_{i1} \\
&\qquad\vdots \\
&\text{Max } z_k = \sum\nolimits_{j=1}^{s} u_j y_{jk} \Big/ \sum\nolimits_{i=1}^{m} v_i x_{ik} \\
&\qquad\vdots \\
&\text{Max } z_n = \sum\nolimits_{j=1}^{s} u_j y_{jn} \Big/ \sum\nolimits_{i=1}^{m} v_i x_{in}
\end{aligned}\right\} \tag{5}$$

Subject to

$$\sum\nolimits_{j=1}^{s} u_j y_{jk} \Big/ \sum\nolimits_{i=1}^{m} v_i x_{ik} \le 1,\quad k = 1,2,...,n$$
$$v_i, u_j \ge \varepsilon > 0,\quad i = 1,2,...,m;\quad j = 1,2,...,s$$

4 Empirical Study

By assuming the similarity of the firms may bias the results. When using the same standard to evaluate the performance of companies; however the difference including products, life cycle etc. of the companies being as DMUs is so critical, the equal standard seems to be a big problem on this research. Thus, the author tried to select the firms with equal character and field being as DMUs. The author chooses services-prepackaged software firms in S&P 500 index as DMUs to evaluate the performance. These companies include Adobe Systems Inc., Autodesk Inc., BMC Software, CA, Inc., Citrix Systems, Compuware Corp., Electronic Arts, Intuit Inc., Microsoft Corp., Oracle Corp., Salesforce.com, and Symantec Corp.

Cost of revenue and R&D expense have been selected as inputs while total revenue, Return on Investment (ROI) as well as net income growth have been selected as outputs. Based on the CCR DEA, Autodesk Inc, BMC Software, CA, Inc., Citrix Systems, as well as Compuware Corp. are 100% efficient. On the other hand, by using the novel MOP approach, Autodesk Inc., BMC Software, Citrix Systems still has achieved the optimum efficiency of 100%. Both results of CA, Inc. and Compuware Corp. are significant different form CCR DEA and MOP based DEA. (Please refer to Table 1 for the empirical study results.) This empirical study demonstrates that performance evaluation results based on CCR DEA or the novel MOP are different especially in some DMUs.

Table 1 Evaluation of Services-Prepackaged Software firms in S&P 500 index by CCR and MOP Based DEA

No.	DMU	CCR	Rank	MOP	Rank
1	Adobe Systems Inc.	0.977	6	0.977	4
2	Autodesk Inc.	1.000	1	1.000	1
3	BMC Software	1.000	1	1.000	1
4	CA, Inc.	1.000	1	0.975	5
5	Citrix Systems	1.000	1	1.000	1
6	Compuware Corp.	1.000	1	0.681	10
7	Electronic Arts	0.344	12	0.344	12
8	Intuit Inc.	0.798	9	0.775	8
9	Microsoft Corp.	0.796	10	0.770	9
10	Oracle Corp.	0.686	11	0.674	11
11	Salesforce.com	0.898	8	0.837	7
12	Symantec Corp.	0.943	7	0.890	6

5 Discussion

In the following section, both theoretical advances of the MOP based DEA approach as well as managerial implications will be discussed. At first, the novel MOP based DEA approach, can really provide a totally different result based on the assumptions of fair weights versus each input and output. For example, CA, Inc. and Compuware Corp. have been shown less efficient based on MOP based DEA; however, both firms have been indicated 100% efficient based on CCR DEA. The successful introduction of the min-max approach resolved the traditional DEA model successfully. Apparently, this novel MOP model can be considered as a better alternative to the traditional CCR DEA.

In the past, people seldom developed an evaluation model especially by using DEA for the services-prepackaged software firms. Thus, this research defined a feasible framework for evaluating the services-prepackaged software firms which can be leveraged as the basis for government policy and firm level investment, R&D and competitive strategy definitions. Further, the novel MOP based DEA approach demonstrated different results to the traditional CCR DEA. In the empirical study, the MOP based DEA demonstrated that Autodesk Inc., BMC Software, and Citrix Systems should be the services-prepackaged software firms to invest. Meanwhile, based on the MOP based DEA results, firms including Compuware Corp., Electronic Arts, Intuit Inc., Microsoft Corp., as well as Oracle Corp. are still far less than efficient. Appropriate adjustments of both inputs and outputs are required to enhance their competitiveness.

6 Conclusions

The information technology is one of the popular industries at the moment due to IT is starting to spread further from the conventional personal computer and network technologies to integrations of other fields of technology such as the use of cell phones, televisions, automobiles, etc. Albeit important, very few researches tried to define an evaluation framework for IT especially in services-prepackaged software firms. This research bridged the gap and introduced a novel MOP based DEA approach which overcame the shortage of the traditional CCR DEA model. Based on the evaluation results, Autodesk Inc., BMC Software, and Citrix Systems are the most efficient services-prepackaged software firms. The evaluation results can serve as the basis for investment strategy definitions.

References

Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30(9), 1078–1092 (1984)

Charnes, A., Clark, T., Cooper, W.W., Golany, B.: A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the US Air Force. Annals of Operations Research 2(1), 95–112 (1985)

Charnes, A., Cooper, W., Rhodes, E.: Measuring the efficiency of decision making units. European Journal of Operational Research 2(6), 429–444 (1978)

Chiang, C.I., Tzeng, G.H.: A new efficiency measure for DEA: efficiency achievement measure established on fuzzy multiple objectives programming. Journal of Management 17(2), 369–388 (2000a) (Chinese)

A New Efficiency Measure for DEA: Efficiency Achievement Measure Established on Fuzzy Multiple Objectives. Programming. Journal of Management 17(2), 369–388

Chiang, C.I., Tzeng, G.H.: A multiple objective programming approach to data envelopment analysis. In: Shi, Y., Zeleny, M. (eds.) New Frontiers of Decision Making for the Information Technology Era, pp. 270–285. World Science Publishing Company, Singapore (2000b)

Coelli, T., Prasada Rao, D.S., O'Donnell, C.J., Battese, G.E.: An introduction to efficiency and productivity analysis. Springer, New York (2005)

Cooper, W., Seiford, L.M., Zhu, J.: Data envelopment analysis, pp. 1–39. Kluwer Academic Publishers, Boston (2004)

Emrouznejad, A.: Data Envelopment Analysis Homepage (1995), http://www.DEAzone.com

Fare, R., Hunsaker, W.: Notions of efficiency and their reference sets. Management Science 32(2), 237–243 (1986)

Fare, R., Grabowaski, R., Grosskopf, S.: Technical efficiency of Philippine agriculture. Applied Economics 17(2), 205–214 (1985)

Farrell, M.J.: The measurement of productive efficiency. Journal of the Royal Statistical Society 120(3), 253–291 (1957)

Flynn, B., Schroeder, R., Sakakibara, S.: A framework for quality management research and an associated measurement instrument. Journal of Operations Management 11(4), 339–366 (1994)

Helms, M.M.: Encyclopaedia of Management, 5th edn. Thomson Gale (2006)

Jorgenson, D.W.: Information Technology and the U.S. Economy. The American Economic Review 91(1), 1–32 (2001)

Kasul, R.A., Motwani, J.G.: Performance measurements in worldclass operations: A strategic model. Benchmarking. An International Journal 2(2), 20–36 (1995)

Keen, P.C.W.: Shaping the future. Harvard Business School Press, Boston (1991)

Lewin, A.Y., Minton, J.W.: Determining organizational effectiveness: another look, and an agenda for research. Management Science 32(5), 514–538 (1986)

Liu, L.C., Lee, C., Tzeng, G.H.: DEA Approach of the Current-period and Cross-period Efficiency for Evaluating the Vocational Education. International Journal of Information Technology and Decision Making 3(2), 353–374 (2004)

Ohta, H., Yamaguchi, T.: Multi-goal programming including fractional goal in consideration of fuzzy solutions. Journal of Japan Society for Fuzzy Theory and System 7(6), 1221–1228 (1995)

Parker, R.P., Grimm, B.T.: Recognition of Business and Government Expenditures on Software as Investment: Methodology and Quantitative Impacts (2000)

Phillips, P.A.: Performance measurement systems and hotels: a new conceptual framework. International Journal of Hospitality Management 18(2), 171–182 (1999)

Sakawa, M., Yano, H.: Interactive decision making for multi-objective linear fractional programming problems with parameters. Cybernetics and Systems: An International Journal 16(4), 377–394 (1985)

Sakawa, M., Yumine, T.: Interactive fuzzy decision-making for multi-objective linear fractional programming problems. Large Scale Systems 5(2), 105–114 (1983)

Skinner, W.: Production under pressure. Harvard Business Review 44(6), 139–146 (1966)

Sumanth, D.J.: Productivity engineering and management. McGraw-Hill, New York (1994)

Tangen, S.: Demystifying productivity and performance. International Journal of Productivity and Performance Management 54(1), 34–46 (2005)

Tsai, H.C., Chen, C.M., Tzeng, G.H.: The comparative productivity efficiency for global telecoms. International Journal of Production Economics 103(2), 509–526 (2006)

Wheelwright, S.: Reflecting corporate strategy in manufacturing decisions. Business Horizons 21(1), 57–66 (1978)

Yu, J.R., Tzeng, Y.C., Tzeng, G.H., Yu, Z.Y., Sheu, H.J.: A fuzzy multiple objective programming to DEA with imprecise data. International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems 12(5), 591–600 (2004)

Yu, J.R., Tzeng, G.H., Chiang, C.I.: Raw material supplier rating in the semiconductor manufacturing industry through fuzzy multiple objective programming to DEA. International Journal of Quantitative Management 13(4), 101–111 (2007)

Zimmermann, H.J.: Fuzzy programming and linear programming with several objective functions. Fuzzy. Sets and Systems 1(1), 45–55 (1978)

Hot Fomentation of the Lower-Back for Stress Relief in Students Preparing for a National Examination of Clinical Medical Technologist

Hanachiyo Nagata, Junzo Watada, Ito Yushi, Takao Shindo, Masasuke Takefu, Masahiro Nakano, Kumiko Satou, Yoriko Hasimoto, Sadahiro Kumamoto, Sumiko Oki, Fusako Fujii, Yukmitu Satou, and Norio Akaike

Abstract. The objective of this study is to propose a method of hot fomentation for a lower-back care for stress relief. Experimental evidence statistically validated this approach for relieving stress. In the experiment, hot fomentation was applied to twenty senior college students (from 21 to 22 years old) who had complained of stiffness in their shoulder and lumbar region of the back during the month prior to a national examination for clinical Medical technologist. Before and after the 30-minute hot fomentation care with a hot-pack, we measured their psychological and physical reactions. The changes in their psychological reactions

Hanachiyo Nagata · Masahiro Nakano · Yukmitu Satou
Junshin Gakuen University, Faculty of Health Sciences

Junzo Watada
School of Information, Production and System, Waseda University

Ito Yushi · Fusako Fujii · Norio Akaike
Kumamoto Health Science University

Takao Shindo
Teacher of English at Fukuoka Senior High School

Masasuke Takefu
Center for Comprehensive Community Medicine, Saga University

Kumiko Satou
University of Human Arts and Sciences

Yoriko Hasimoto
ASOU Nursing and Medical College

Sadahiro Kumamoto
Kumamoto University

Sumiko Oki
The Former of Kumamoto Health Science University

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 13–20.
springerlink.com

were evaluated using the visual analog scale (VAS) to assess the degree of comfort. The physical reactions were evaluated using the cortisol level of blood serum as an index of stress.

(1) The experimentally determined VAS values indicated that hot fomentation significantly relieved stress.
(2) The cortisol level of blood serum (a proven neuroimmunological index of stress) decreased significantly following the hot fomentation care.
(3) However, both the changes of VAS values and the cortisol level changes of blood serum are not significantly correlated.

We found that lower-back hot fomentation by a hot-pack provided a significant stress relief effect, both mentally and physically, for the 20 students who planned to take a national examination of Clinical Medical technologist. However, both the psychological and physical outcomes were not significantly correlated. This lack of correlation indicates that the anxiety relief might exceed the physical stress relief.

Keywords: Hot fomentation, Hot-pack, Lower-back portion, Visual analog scale, VAS, the cortisol level of blood serum, National examination, laboratory technologist.

1 Introduction

It is currently well known that various types of stress produce great burdens on human lives and behaviors. The objective of this study is to propose lower-back hot fomentation care as a method for stress relief. Stress can be measured using both psychological and physical methods. The visual analog scale (VAS) of comfort-discomfort can be used to psychologically evaluate the stress level of a patient according to the degree of comfort or discomfort. The cortisol level of blood serum is used as a stress index to evaluate the physical degree of stress. This study reports the result of an attempt to reduce a patient's stress using a 30-minute hot-pack hot fomentation for a lower-back care.

In Japan, all healthcare professionals must obtain national approval to begin their careers in medicine. That is, final-year students of medical technology cannot start working without passing the national examination, and therefore all of the final-year students take the national examination. All of the medical technology students endure a long period of mental stress because of the clinical exercises and internship they undertake at a medical care facility and the education program they receive at school. Furthermore, universities provide various countermeasures in the education program to help ensure that all of the students pass the national examination. The number of students who pass the test influences the evaluation of university quality. Therefore, university staff helps their students prepare as best as possible.

Various reports and papers have considered the preparation for the national examination in Japan, such as support of the students and the educational program staff. However, the level of stress endured by a student during his/her preparation for the national examination or how a student can be relieved of the various kinds of stress is unknown.

In this paper, we propose lower-back hot fomentation as a stress relief method. This approach to stress relief is validated on senior students from 21 to 22 years old who will take a national examination for clinical Medical technologist in a month. After a 30-minute hot fomentation care using a hot-pack, the experiment to examine the psychological and physical conditions of all the subjects who had felt stiffness in their shoulders and in the lumbar region of their backs was undertaken. The psychological evaluation used a visual analog scale (VAS), which assesses the degree of comfort or discomfort. The physical evaluation used the cortisol level of blood serum as an index of stress. (1) The experimentally determined VAS values outcome indicated that hot fomentation significantly relieved stress. (2) The cortisol level of blood serum (a proven neuroimmunological index of stress) decreased significantly following the hot fomentation care. (3) However, the changes in the VAS values and the cortisol level changes of blood serum are not significantly correlated.

2 Research Method

2.1 Subjects

The subjects in this experiment were senior college students, from 21 to 22 years old, in the School of Clinical Laboratory Technology, University A, who complained of stiffness in the shoulder and lumbar region of the back. They planned to take a national examination for clinical Medical technologist a month after the experiment.

2.2 Ethical Considerations in Primary Research

\The primary research was conducted with the approval of the ethics review committee of University A after our submission of the required document (Approval Number of Experiment 20-51 of the Ethics Review Committee).

After initially explaining the objective of the research and experiment to the subject students in both oral and written forms, the subjects underwent the experiment when the approval and agreement between the subjects and the experiment staff were obtained. Furthermore, in the experiment, we gave sufficient considerations to the following requirement:

(1) The personal information is sufficiently protected so that is cannot identify an individual.
(2) The biographical data are not used for any purpose other than this research.

2.3 Experimental Method

The experiment was performed in a laboratory at University A from 18:00 to 20:00 in the middle of January 2011; the care was provided to students who would take the national examination of clinical Medical technologist in a month. While the students lay on a bed for 30 minutes, hot-packs were placed on the part of the back corresponding to the second vertebrae cervicales CII to the forth lumbar vertebra. Before and after this care, VAS assessments were given, and blood samples were taken.

(1) The hot-packs used were DeRoyal Sofsorb Absorbent Pads & Sheets (Duro-Med Industires, Inc., USA) that are applied when their centers reach a temperature of 45 degree Celsius by heating in a microwave oven. Each hot-pack, which is 40 degree Celsius on the surface, is covered with a dry towel. After informing each subject, their entire back from the posterior neck and vertebra prominens CVII to both shoulders and the third and fourth lumbar vertebrae is covered with the sheet.
(2) The cortisol level of the blood serum is used as a neuroimmunological index to measure physical stress. All of the students had blood samples drawn just before and after the care.
(3) The VAS was used as an index to measure the psychological response of the subjects. In the VAS, "the most uncomfortable feeling" was assigned to 10 in the scale, and "the most comfortable feeling" was assigned to 1.
(4) The ambient temperature of the room was set to 25 ±2 degrees Celsius with a natural sound state and without any conversation.

2.4 Statistical Analysis

Using a non-parametric Wilcoxon signed-rank test, the cortisol level of blood just before and after the care and the VAS results are tested for significant differences. We also calculated the Fisher's correlation between the cortisol level and the VAS to test the significance of their correlation.

3 Results

3.1 The Change in Cortisol Levels Caused by a 30-Minute Hot-Pack Care

For the 20 senior students (21 to 22 years old) who experience stiffness in the shoulder and in the lumbar region of the back and who will take a national examination for clinical Medical technologist in a month, the cortisol levels of blood serum were 8.4 ± 0.9 $\mu g/dl$ and 6.0 ± 0.9 $\mu g/dl$ before and after the 30-minute hot fomentation for a lower-back care using a hot-pack, respectively. These data were significantly different ($p < 0.001$), as indicated by Figure 1.

3.2 The Change in Psychological Responses Caused by a 30-Minute Hot-Pack Care

For the 20 subjects, the psychological response of the VAS values were 6.3 ± 0.3 and 4.6 ± 0.5 before and after the 30-minute lower-back hot fomentation care, respectively. These data were significantly different ($p < 0.01$), as indicated by Figure 2.

3.3 *Correlation Coefficient between the Cortisol Levels and Psychological Responses*

For the 20 subjects, the correlation coefficient of the care-induced changes between the VAS values of the psychological responses and the cortisol levels of blood serum was 0.3, p=0.5. The values are not significantly correlated.

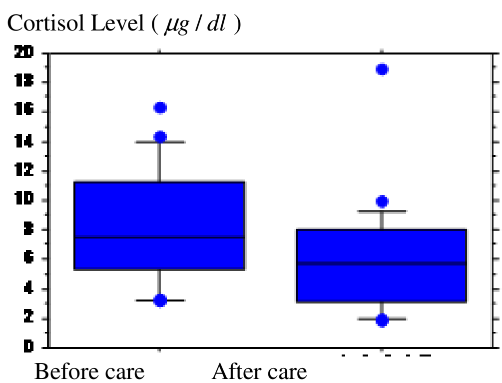


Fig. 1 Change in the mean cortisol level caused by a 30-minute hot-pack care

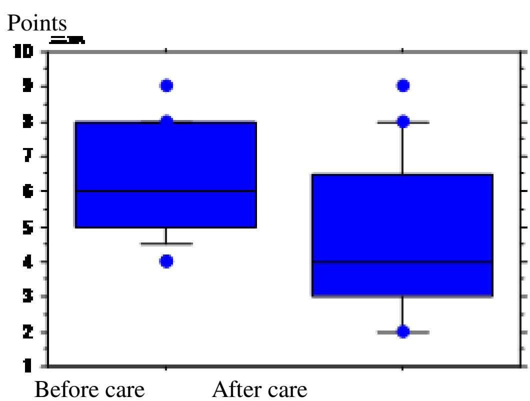


Fig. 2 Change in the mean VAS value caused by a 30-minute hot-pack care

4 Discussion

4.1 The VAS Evaluation of the Psychological Influence of Hot Fomentation Care

Using the VAS, the psychological state of the subject was measured before and after the hot fomentation care for the 20 subject students who were experiencing stiffness in the shoulder and the lumbar region of the back one month before the national examination of clinical Medical technologist. The VAS value dropped from 6.2 to 4.6. This change indicated that the heat-retaining effect of the lower-back care relieved the unpleasant emotions and stress and led to a more comfortable psychological response. The hot-pack care relieved stress and pain by thermal heating of the back, and the VAS values decreased. Hasebe (2005) [1], Matsushita (2008) [2], and Kubo *et al.* (2011) [3] reported that hot fomentation care using a hot-pack increases the blood flow volume of subjects and mitigates shoulder and loin stiffness. For psychological aspects, Buttagat *et al.* (2012) [4] reported that hot-pack fomentation care is capable of increasing the feeling of comfort and enhancing relaxation.

4.2 The Stress Index As Determined by the Blood Serum Cortisol Level

Stahl and Dorner (1982) [5] reported that physical stressors increase the cortisol concentration of blood serum. As senior students have been preparing for their national examination of clinical Medical technologist for a long period of time, they have experienced continuous psychological stressors. In particular, just before the national examination, they must spend many hours in preparation. Students sit in a chair for 6 to 8 hours a day, leading to stiffness in the shoulders and the lumbar region of the back those results from maintaining this position. Furthermore, they endure additional stressors caused by the anxiety around the results of the national examination, which causes much psychological and physical burden.

The lower-back fomentation care by a hot-pack decreased the blood serum cortisol level and significantly affected the national examination stressors. In general, the stress response can be understood to involve two response systems: (1) the endocrine system of the hypothalamic-pituitary-adrenal cortex axis and (2) the autonomic nervous system of the hypothalamus-pons-medulla oblongata-spinal marrow-adrenal medulla axis. The endocrine cortisol level of the blood serum serves as an effective index of stress. Pani *et al.* (2011) [6] reported that the cortisol level increases in response to psychological stress but decreases in response to comfortable stimuli. Al-Ayadhi (2005) [7] reported that the uncomfortable and tense state of sympathetic hyperactivity produced by stressors was changed into a comfortable state using a hot-pack fomentation care and that, furthermore, the biological response of the autonomic nervous system was recognized. Fredrickson and Levenson (1998) [8] clarified that the heat retention affected activation of the parasympathetic nervous system and inhibition of sympathetic hyperactivity. In this paper, we report that a hot-pack fomentation care can increase comfortable stimuli and decrease the cortisol level of blood serum, in agreement with the above reports. We believe that the hot-pack fomentation care provides stress relief.

4.3 The Relationship between the Blood Serum Cortisol Level and the VAS Value of Psychological Response

In this study, the stress relief by lower-back hot fomentation using a hot-pack was examined using the cortisol level of blood serum as a stress index and VAS values as a measure of the psychological response. After treating the subject with hot-packs on lower-back portion, the decrease in cortisol level and VAS values are recognized but the correlation between both the change of the cortisol level and the change of the VAS values was not significant. In general, hot fomentation produces relaxation and decreases the cortisol level of blood serum. Pawlow and Jones (2005) [9] also observed decreases in VAS values, as did Yamamoto and Nagata (2011) [10]. The hot fomentation care of the lower-back promotes the recovery from uncomfortable stress and creates a more comfortable state, inhibits the activities of skin sympathetic nerves and, as a result, improves the peripheral circulatory state. These effects relax the mind and body to rest after the burden of stress, as Watanabe *et al.* (2006) reported [11]. Nevertheless, the correlation between the cortisol level change and the VAS value change were not significant An explanation for this lack of correlation might be that the national examination for students causes a greater amount of anxiety and tension than the amount of relaxation provided by the lower back hot fomentation. Furthermore, statistically, the value of the cortisol level is more precise (one tenth of one percent), but the psychological VAS values are less precise because they employ discrete values from 1 to 10 on a scale of 10. Therefore, another psychological scale should be employed.

5 Conclusions

This paper reported the result of analysis of relief stress using a 30-minute hot fomentation care with a hot-pack on the lower-back of subjects. This method was applied to final-year students of medical technology who cannot begin working until they pass their exam. The experiment attempted to examine the psychological and physical states of the subjects who were experiencing stiffness in the shoulder and the lumbar region of the back. The psychological evaluation was performed using a visual analog scale (VAS) to assess the degree of being comfort or discomfort. The physical evaluation consisted of measurements of the blood serum cortisol level as an index of stress. The result showed that (1) VAS values indicated the effective relief of stress, (2) the cortisol level of blood serum as a neuroimmunological index is significantly decreased and (3) but the correlation between the VAS value change and the cortisol level change was not recognized significantly.

References

[1] Yoshiko, H.: Effects of hot compress care with a hot water bottle on physiological parameters and subjective sensations in healthy women. Japan Journal of Nursing Science 2(2), 107–114 (2005) (in Japanese)
[2] Kakushi, M., Akinori, M., Chuwa, T.: Efficacy of Waon Therapy for Fibromyalgia. Internal Medicine 47(16), 1473–1476 (2008) (in Japanese)
[3] Kubo, K., Yajima, H., Takayama, M., Ikebukuro, T., Mizoguchi, H., Takakura, N.: Changes in blood circulation of the contralateral achilles tendon during and after acupuncture and heating. International Journal of Sports Medicine 32(10), 807–813 (2011) (in Japanese)
[4] Buttagat, V., Eungpinichpong, W., Chatchawan, U., Arayawichanon, P.: Therapeutic effects of traditional Thai massage on pain, muscle tension and anxiety in patients with scapulocostal syndrome: A randomized single-blinded pilot study. J. Bodyw Mov. Ther. 16(1), 57–63 (Epub. June 22, 2012)
[5] Stahl, F., Dorner, G.: Responses of salivary contisol levels to stress situations. Endokrion 80(2), 158–162 (1982)
[6] Pani, S.C., Al Askar, A.M., Al Mohrij, S.I., Al Ohali, T.A.: Evaluation of stress in final-year Saudi dental students using salivary cortisol as a biomarker. J. Dent. Educ. 75(3), 377–384 (2011)
[7] Al-Ayadhi, L.Y.: Neurohormonal changes in medical students during academic stress. Ann. Saudi. Med. 25(1), 36–40 (2005)
[8] Fredrickson, B.L., Levenson, R.W.: Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. Cognition and Emotion 12, 191–220 (1998)
[9] Pawlow, L.A., Jones, G.E., Pawlow, L.A., Jones, G.E.: The impact of abbreviated progressive muscle relaxation on salivary cortisol and salivary immunoglobulin A (sIgA). Appl. Psychophysiol Biofeedback 30(4), 375–387 (2005)
[10] Yamamoto, K., Nagata, S.: Physiological and psychological evaluation of the wrapped warm footbath as a complementary nursing therapy to induce relaxation in hospitalized patients with incurable cancer: a pilot study. Cancer Nurs. 34(3), 185–192 (2011) (in Japanese)
[11] Watanabe, E., Fukuda, S., Hara, H., Maeda, Y., Ohira, H., Shirakawa, T.: Differences in relaxation by means of guided imagery in a healthy community sample. Altern. Ther. Health Med. 12(2), 60–66 (2006) (in Japanese)

Hybrid MCDM Model for the Value Evaluation of the Restoration of Historical Objects

Mei-Chen Lo, Tadeusz Trzaskalik, Maciej Nowak, Tian-Jong Hwu,
Gwo-Hshiung Tzeng, and Jerzy Michnik

Abstract. As time goes by, every historic object dilapidates and wears out. As the result, the values that it used to represent become obliterated and its effete on the public is weakened. Conservation and restoration of art works aim at preserving the extant matter and, if possible, at bringing the antiques to their former glory in historical value; the more so that the historical value of the objects also increases with time. Conservators' work, independently of their special fields of interest, should be preceded by research whose goal is the determination of the guidelines for the conservation efforts and the selection of the best methods of action in interdependence and feedback among dimensions and criteria. A thorough analysis determines several possibility methods of action, emphasizing various groups of values. In this study, a hybrid Multiple Criteria Decision Making (MCDM) model based on DNP (DEMATEL-based ANP) is used to assess the historical objects relative values viewed by different groups of experts in inter-relationship problem of real world. The procedures provide a reference guide of improvement directions and efficient art work reservation of historical objects for achieving the aspiration levels.

Mei-Chen Lo · Tian-Jong Hwu
Department of Business Management, National United University
No.1, Lienda Rd., Miaoli 36003, Taiwan
e-mail: {meichen_lo,tjhwu}@nuu.edu.tw

Gwo-Hshiung Tzeng
Institute of Project Management, Kainan University
No. 1, Kainan Road, Luchu, Taoyuan County 338, Taiwan
e-mail: ghtzeng@mail.knu.edu.tw

Gwo-Hshiung Tzeng
Institute of Management of Technology, National Chiao Tung University
No. 1001 Ta-Hsueh Road, Hsinchu Country 300, Taiwan
e-mail: ghtzeng@mail.nctu.edu.tw

Tadeusz Trzaskalik · Maciej Nowak · Jerzy Michnik
Department of Operations Research, University of Economics in Katowice
ul. 1 Maja 50, 40-287 Katowice, Poland
e-mail: {tadeusz.trzaskalik,jmichnik}@ue.katowice.pl
nomac@ae.katowice.pl

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 21–31.
springerlink.com

Keywords: Hybrid Multiple Criteria Decision Making (H-MCDM), historical portable organ, art restoration, value analysis, DNP (DEMATEL-based ANP), aspiration level.

1 Introduction

As time goes by, every historic object dilapidates and wears out. As the result, the values that it used to represent become obliterated and its effete on the public is weakened. Conservation and restoration of art works aim at preserving the extant matter and, if possible, at bringing the antiques to their former glory; the more so that the historical value of the objects increases with time.

Conservators' work, independently of their special fields of interest, should be preceded by research whose goal is the determination of the guidelines for the conservation efforts and the selection of the best methods of action. Inventory, documentation and research efforts are completed by a value analysis whose purpose is to precisely define several values of the object so as to emphasize and reveal the most important of them. A thorough analysis determines several possible methods of action, emphasizing various groups of values. The basic value groups of historic objects and monuments have been formulated by Walter Frodl [2]. These groups, expanded by musical issues, are used in this paper.

The possibility of a variant-based approach to the issue of the value analysis of historic items suggests that the methodology of the multi-criteria decision support can be used for the selection of the best variant of conservation method of the individual item or monument. The possibility of shaping the selected values after the reconstruction of the object allows regarding the values as decision criteria. Possible ways of the instrument reconstruction constitute here decision variants.

In the 17th- and 18th-century Poland the portable organ, called the positive organ, was a very popular instrument'; almost every parish was equipped with one. It was not on1y a church instrument, since the portab1e organ was used a1so to accompany dancers in ballrooms. Its popularity was due above all to the ease of hand1ing and the possibility of easy transportation. Unfortunately, on1y 18 copies of this once so common instrument are nowadays extant in Poland (according to current research). One of the extant instruments from this group, found only recently, comes from Sokoły near Łapy in the Podlasie region of Poland. For many years the instrument had been stored disassembled, undergoing atmospheric and biological damage. Its condition made it impossible to use it either as a visual historic item ("piece of furniture") or as a musical instrument. Such condition is called in Polish conservation science terminology a "destrukt".

The value analysis of historic items is not on1y a theoretical consideration, but aims at determining the guidelines of conservation efforts and, in connection with experience and conservation science, allows for the selection of the best conservation methods for individual works of art. The precise estimation of value of the extant elements of the instrument became thus a research problem; on this basis the determination of several (10 to 20) variants of conservation programs will be made. The purpose of this paper is the joint application of the analysis evaluating an historic organ and the MCDM method in the selection of the guidelines for conservation efforts in the case of the recently discovered organ.

This paper consists of five sections. In Section 1 selected problems related to the analysis for the evaluation of an historic organ are described. Specified groups of values have been used for the construction of decision criteria. The history and original condition of the instrument in question have been described in Section 2. Possible methods of restoration of this instrument, treated as decision variants in multivariate analysis, have been presented in Section 3. Section 4, the hybrid MCDM method is described and presents the application of the MCDM method proposed for the analysis of the problem in question, and Section 5, conclusions following from it.

2 About the Historical Organ

The first person to recognize and define the value of a separate group of historic objects - historic musical instrument s - was the German scholar and musician Albert Schweitzer'. Thanks to his authority the cause of preservation of historic organs gained many advocates among musicians as well as conservators and researchers.

The value analysis of historic objects, used nowadays in conservation science with respect to all kinds of historic objects and monuments, has been defined by Walter Frodl in the middle of the 20th century, and was subsequently expanded and completed [3]; in the Polish legislation it resulted in an act concerning the preservation and protection of historic monuments [1].

According to this document an "historic monument or object" is "a building or an object, its element or subsystem, man-made or related to human activities which are an evidence of an epoch or an event from the past, whose preservation is of social value due to its historic, artistic or scientific value". Taking into account the synthetic character of the group of objects dealt with in this paper, historic organs, a precise definition of such values will help improve value analysis.

2.1 Values of an Historical Organ as Decision Criteria

In the following discussion suggest a division of the values of historical organs into four groups: historic [6] and [7], artistic, musical and utilitarian values. We will now describe the values constituting each of the four groups.

Historic values determine the character of the object as a document and its influence on the development of historical knowledge. Among the values of this group are *scientific values,* due to the fact that an organ is an historic object, requiring a scholarly description. Also in this group are *technical values,* determining the ingenuity of the construction, the quality of the workmanship and the scientific value of its current condition. Also *historic emotional values,* perceived not only by scientists and scholars, but also by the public at large, belong here.

The *ownership values,* i.e., values stemming from the ownership of the original item (without hypothetical additions) are connected with honest approach of the conservators to the historic object, in which that what is preserved should be emphasized above all, as opposed to that what we think might have been there. The group of *artistic values* is related to the perception of historic organs as works of art, and this is connected with the instrument's case. To this group belong *historic-artistic values,* determining whether the solutions chosen by the builders are typical or atypical as well as the importance of the original, its copy or its hypothetical reconstruction.

Artistic qualities affect the public independently of the current fashion or style. The *artistic effect* of the ease of historic organ should match musical impressions received by the audience from the musical compositions heard by it. *Musical values* become apparent during a musical performance. This study deals here with the issue of style *(historical musical value)* and of sound *(musical quality).* All of them taken together may reinforce the *musical influence* on the amateur listener. It can happen that the regaining of musical value and the preservation of the original technical solutions are conf1icting goals. In such case we face the problem of *utilitarian values* of the historic instrument. The notions of live organ and dead organ are related to this group of values. A *musically dead organ* is an instrument that nowadays cannot fulfill its function of a musical instrument. A *live instrument* is an instrument capable of being used in musical performance, affecting the audience in various ways. Like any historic object, an organ as a piece of furniture can be also visually dead - not suitable for being exhibited, or else visually alive (independently of its musical "vitality") - beautiful, but unplayable.

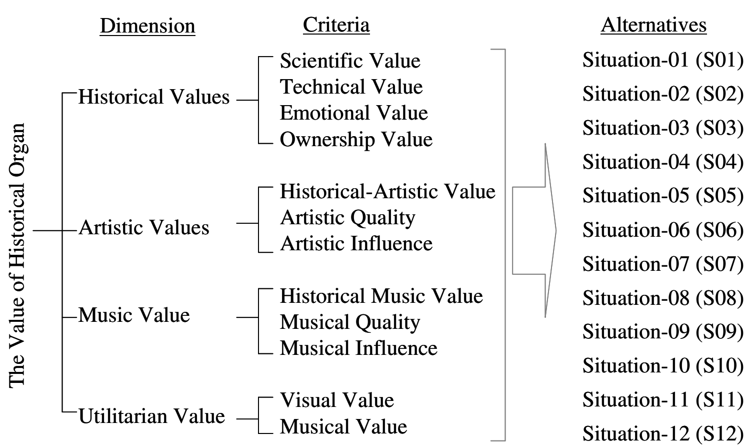


Fig. 1 The Value Structure of Historical Organ

Therefore, the value structure above (Fig. 1) is constructed and used as decision criteria in the problem of the selection of the best conservation alternatives/situations, discussed below. The decision criteria admit the values from 0 (lowest grade) to 4 (highest grade). The decision situations (alternatives) defined later in the paper will be evaluated by experts, utilizing during the evaluation according to his or her expertise of the topic.

3 The Instrument Undergoing the Restoration

The basic feature distinguishing a portable organ from a stationary one are its small dimensions and a design allowing for placing of all elements characteristic for the organ-like instrument s (pipes, wind chest, action, bellows) in a small, easy to handle case. The "compression" of the instrument's mechanism is achieved by making the dimensions of the wind chest as small as possible, restricting the action to the direct transfer of the movement from the key to the pallet and mitering (often repeated) of the pipes or the use of common side walls of wooden pipes. Bellows of small dimensions always require a certain space for proper functioning, and that is why they are located on the instruments, underneath, or next to the case wall opposite the keyboard. The placing of the keyboard is also related to the localization of the bellows, which follows from the construction of the wind chest.

The positive organ from Sokoły is an instrument of the two-chamber type, in which the lower one (larger) contains two wedge bellows, while the upper one (smaller), wind chest, pipes and keyboard. It is characteristic for this instrument that the lower chamber cover can be taken out and, after the bellows have been blocked, the upper chamber can be inserted into the lower one. Once this "package" is closed, the instrument is secured and can be transported conveniently. After the arrival at its destination, a two-part positive organ, when taken apart, is independent and does not require any auxiliary furniture.

The positive organ from Sokoły is preserved as a non-functional and visually unattractive object ("destrukt") - each element was stored separately and individual parts were damaged. About 70% of the case, 90% of the mechanism and 10% of the sound system have been preserved. In this condition the value of the positive organ is recognizable by a narrow group of researchers who are able to visualize how to combine the individual parts.

3.1 Possible Methods of Instrument Restoration

As this study further work on the possible methods of instrument restoration as decision alternatives [8]. On the basis of research and evaluation of the condition of the individual parts of the instrument (or their lack) 12 renovation treatments of the rediscovered instrument have been suggested.

- *Situation I* (S01): Preservation of the instrument as a non-functional, visually unattractive object ("destrukt") and its exhibition in the form of a group of museum exhibits.
- *Situation II* (S02): Integration of the elements of the instrument using racks necessary to place the individual elements in proper places.
- *Situation III* (S03): Integration of the parts of the instrument with full completion of the construction elements of the case; completion of the missing parts of the mechanism. The pipes remain secured, but do not play.

- *Situation IV* (S04): Integration of the parts of the instrument with full completion of the construction elements of the ease according to their former shape. The pipes remain secured, but do not play.
- *Situation V* (S05): Integration of the parts of the instrument with full completion of the construction elements of the case; completion of the missing parts of the mechanism. Reconstruction of the polychrome. The pipes remain secured, but do not play.
- *Situation VI* (S06): Integration of the parts of the instrument with full completion of the construction elements of the ease. Bringing the extant pipes to working condition and reconstruction of the missing pipes, so as to match the sound capabilities of the extant pipes.
- *Situation VII* (S07): Visually, the object is moderately attractive; utilitarian musical value appears, especially for people appreciating the original, historical sound. Exhibition of the extant historic pipes in a display ease without giving them their former technical functionality. Reconstruction of the entire sound system according to preserved models.
- *Situation VIII* (S08): Integration of the parts of the instrument with full completion of the construction elements of the ease according to their former shape. Bringing the pipes to a working condition and reconstruction of the missing pipes, so as to match the sound capabilities of the extant pipes.
- *Situation IX* (S09): Integration of the parts of the instrument with full completion of the construction elements of the ease according to their former shape, Exposition of the extant historical pipes in a display eases without bringing them to a working condition. Reconstruction of the whole sound system according to preserved models.
- *Situation X* (S10): Integration of the parts of the instrument with full completion of the construction elements of the ease according to their farmer shape. Bringing the pipes to a working condition and reconstruction of the missing pipes so as to match the sound of the sound capabilities of the preserved pipes.
- *Situation XI* (S11): Integration of the parts of the instrument with full completion of the construction elements of the case according to their former shape. Exhibition of the preserved historic pipes in a display case without bringing them to a working condition. Reconstruction of the whole sound system according to preserved models.
- *Situation XII* (S12): Preservation of the instrument in its non-functional, visually unattractive condition (as a "destrukt"). Making of an accurate copy. The evaluation focuses on the values of the copy, which is presented to the public.

The possible actions and results which caused by each different situation, it has made under the value evaluation by experts. Although, each of the result from different situation, it may guide different perception on the historic organ to be performed later; and the expectation of the restoration work display in the degree of its original states.

4 Methodologies of MCDM

This section is aimed to understand the evaluation indexes of the value of restoration historical organ. It is collected, selected, analyzed, simulated and tested by the literature and expert questionnaires to find the usefulness and the promotion spots as a basis. Lo & Tzeng [4] indicate that MCDM is a methodology that is able to consider multiple criteria at the same time and also helps the decision- maker to estimate the best case by sorting cases according to the characteristics or criteria ([4], [11], [9], [10]) of each from limited available cases.

4.1 Empirical Analysis

In this study, according to the literature review and expert experiences, an value evaluation system including four dimensions and 12 criteria that will exert an influence on alternative selection of restoration historical organ is established, as given in Table 4.1. A survey was conducted via questionnaires distributed to several groups comprised of the experts' knowledge and background. Their ratings for each criterion's relationship to sustainable development using a five-point scale ranging from 0 (no effect) to 4 (extremely influential) were collected.

This set of criteria provides this study with an overall evaluation system that facilitates further prioritization by the concept of ANP [5]. According to Table 1, the critical factors for value evaluation on restoration of historical organ which is including "Musical-utilitarian value (C_{42})", "Visual-utilitarian value (C_{41})", "Historical-musical value (C_{31})", "Historical-artistic value (C_{21})", "Artistic quality (C_{22})",and "Musical quality (C_{32})" according to the sequence (priority) of these factors from the calculation of ANP concept.

The MCDM model refers to making decisions in the presence of multiple, and often simultaneously faced/managed multiple criteria/objectives with critique criteria in real world. This study adopts MCDM approaches subject to calculate the weights and ranking that is used to solve the decision problem for the priority of value evaluation on restoration historical organ. We applied the influence matrix to compromise the trade-off between value aspects and situations' concerns. The main survey objects engage for the value in four phases (such as, values to Historical, Artistic, Music and Utilitarian), within related areas of experts or scholars who are senior in their domain knowledge and experiences on the evaluations work.

Table 1 The evaluation criteria

Aspects/Dimensions		Criteria	Weight
C_1 Historical Values	C_{11}	Historical-scientific value	0.057
	C_{12}	Historical-technical value	0.064
	C_{13}	Emotional value	0.076
	C_{14}	Ownership value	0.053
C_2 Artistic values	C_{21}	Historical-artistic value	0.086
	C_{22}	Artistic quality	0.084
	C_{23}	Artistic influence	0.080
C_3 Music values	C_{31}	Historical-musical value	0.093
	C_{32}	Musical quality	0.082
	C_{33}	Musical influence	0.075
C_4 Utilitarian values	C_{41}	Visual-utilitarian value	0.111
	C_{42}	Musical-utilitarian value	0.139

According to Table 2, the prominence $(d_i + r_i)$, "Historical Values (C_1)" is the highest impact of the strength of relation that means the most important influencing factors; in addition, "Utilitarian Values (C_4)" are all the factors that affect the least degree of other factors. According to the relation $(d_i - r_i)$, we also can find "Music Values (C_3)" is the highest degree of impact relationship that affects other factors directly. These dimensions also have the interact characteristics. Opposite, "Historical Values (C_1)" is the most vulnerable to impact that compare with other dimensions.

A grade of value from 0 (lowest grade) to 10 (highest grade) has been made by the set of experts' value evaluations, shown as Table 3. Each situation/alternative is depending of the values that the instrument would gain after the reconstruction according to the given decision situation.

Table 2 Influence Evaluation by dimension

	C_1	C_2	C_3	C_4	***D***	***d+r***	***d-r***	Rank
C_1	0.067	0.045	0.046	0.049	0.206	0.475	-0.062	4
C_2	0.068	0.048	0.068	0.057	0.242	0.458	0.025	2
C_3	0.067	0.066	0.050	0.065	0.248	0.456	0.039	1
C_4	0.065	0.057	0.045	0.030	0.197	0.397	-0.003	3
R	0.268	0.216	0.209	0.200	--	--	--	--

Table 3 Value Evaluation by Criteria

	Criteria	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12
C_{11}	Historical-scientific value	10	8	6	6	6	6	6	6	6	6	6	0
C_{12}	Historical-technical value	4	6	10	8	8	6	10	6	10	6	10	0
C_{13}	Emotional value	10	10	10	8	6	10	6	6	4	4	2	0
C_{14}	Ownership value	10	10	10	9	5	8	5	9	4	5	0	0
C_{21}	Historical-artistic value	0	2	4	6	6	4	4	6	6	8	8	0
C_{22}	Artistic quality	0	0	2	4	8	2	2	4	4	8	8	8
C_{23}	Artistic influence	2	2	6	8	10	6	6	8	8	10	10	10
C_{31}	Historical-musical value	0	0	0	0	0	10	4	10	4	10	4	4
C_{32}	Musical quality	0	0	0	0	0	8	10	8	10	8	10	10
C_{33}	Musical influence	0	0	0	0	0	8	10	8	10	8	10	10
C_{41}	Visual-utilitarian value	2	4	6	8	10	6	6	8	8	10	10	10
C_{42}	Musical-utilitarian value	0	0	0	0	0	8	10	8	10	8	10	10

The value evaluation by situations/alternatives is demonstrated as Table 4. Thus, two alternatives, S10 and S11 get the best evaluation. Thus, the sequence of performance of value evaluation is S10, S11, S08, S09, S06, S07, S12, S04 and then, S03, S05, S02, S01. As continuing lowering the value of the concordance threshold does not result in generating more detailed information, so the procedure and propose the decision making is recommended, as it has got the best evaluation in all rankings that have been constructed.

Base on these empirical results, we construct the network relationship map of each dimension as shown in Fig.2. It illustrates the critical problems in evaluating the restoration work on historical organ which including “Music Values (C_3)”, “Artistic Values (C_2)” and “Utilitarian Values (C_4) which are easily impacted by other factors. Therefore, this model provides the direction of problem solving from “Music Values (C_3)”, “Artistic Values (C_2)” and “Utilitarian Values (C_4). Some criteria take more effort on placing strategies are suggested according to experts’ verification, such as Historical-Artistic Value, Historical Music Values, Musical Quality, Artistic Quality, Artistic Influence, Emotional Value, etc.

Table 4 Value Evaluation by Situations/Alternatives

Criteria \ Alternatives	Local weight	Global weight	Aspired Level	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12
C_1	0.25			8.47	8.52	9.08	7.75	6.30	7.64	6.81	6.64	5.99	5.18	4.53	0.19
C_{11}	0.23	0.06	10	10.00	8.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	0.00
C_{12}	0.25	0.06	10	4.00	6.00	10.00	8.00	8.00	6.00	10.00	6.00	10.00	6.00	10.00	0.00
C_{13}	0.30	0.08	10	10.00	10.00	10.00	8.00	6.00	10.00	6.00	6.00	4.00	4.00	2.00	0.00
C_{14}	0.21	0.05	10	10.00	10.00	10.00	9.00	5.00	8.00	5.00	9.00	4.00	5.00	0.00	0.90
C_2	0.25			0.64	1.33	3.97	5.97	7.95	3.97	3.97	5.97	5.97	8.64	8.64	5.88
C_{21}	0.34	0.09	10	0.00	2.00	4.00	6.00	6.00	4.00	4.00	6.00	6.00	8.00	8.00	0.00
C_{22}	0.34	0.08	10	0.00	0.00	2.00	4.00	8.00	2.00	2.00	4.00	4.00	8.00	8.00	8.00
C_{23}	0.32	0.08	10	2.00	2.00	6.00	8.00	10.00	6.00	6.00	8.00	8.00	10.00	10.00	10.00
C_3	0.25			0.00	0.00	0.00	0.00	0.00	8.74	7.78	8.74	7.78	8.74	7.78	7.78
C_{31}	0.37	0.09	10	0.00	0.00	0.00	0.00	0.00	10.00	4.00	10.00	4.00	10.00	4.00	4.00
C_{32}	0.33	0.08	10	0.00	0.00	0.00	0.00	0.00	8.00	10.00	8.00	10.00	8.00	10.00	10.00
C_{33}	0.30	0.07	10	0.00	0.00	0.00	0.00	0.00	8.00	10.00	8.00	10.00	8.00	10.00	10.00
C_4	0.25			0.89	1.78	2.67	3.56	4.45	7.11	8.22	8.00	9.11	8.89	10.00	10.00
C_{41}	0.45	0.11	10	2.00	4.00	6.00	8.00	10.00	6.00	6.00	8.00	8.00	10.00	10.00	10.00
C_{42}	0.55	0.14	10	0.00	0.00	0.00	0.00	0.00	8.00	10.00	8.00	10.00	8.00	10.00	10.00
TOTAL	5.00	1.00		0.25	0.29	0.48	0.54	0.47	0.69	0.67	0.73	0.72	0.79	0.77	0.60
Performance Ranking				12	11	9	8	10	5	6	3	4	1	2	7

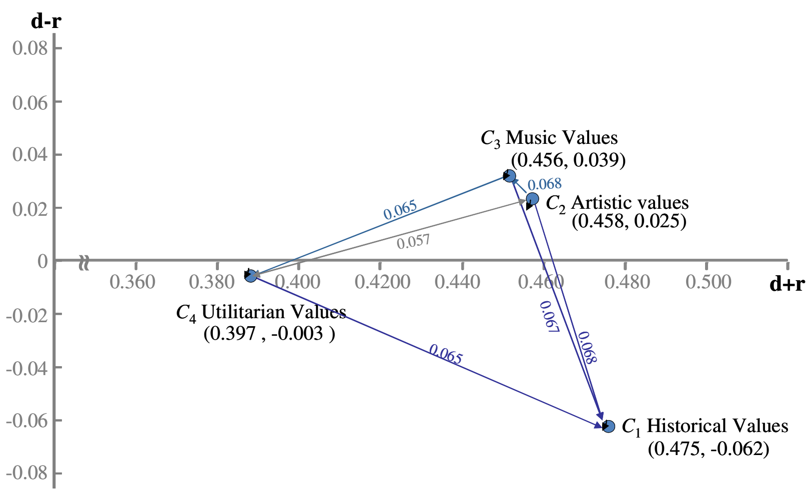


Fig. 2 The NRM of each dimension

5 Conclusions

The study described in this paper demonstrates the restoration of historical organ is within the realms of possibility. The possibilities of application of multi-criteria decision support in choosing an approach to conservation of historical organ have been presented. The work resulted in the renovation of a valuable Instrument made by Polish organ-builders. It seems that this methodology may be applied also for a wider range of objects of historical value, although this requires complex analysis of the set of criteria under consideration. Another issue requiring situation analysis would be the course of action in the case when decision alternatives/variants are evaluated by a group of experts.

This study is including different structure to be overcome. For this purpose, an idea generation method has been proposed in which the influence matrix is adopted based on many fields knowledge. Here, we employed with an innovative method with group expert's participation and multi-criteria decision making, the concept of ANP handling with many potential uncertainty factors which is able to evaluate and provide the most suitable guide for evaluating the restoration value of historical organ.

Acknowledgements. This study is granted (No. 099WMRFD1717003) by National Science Council–Taiwan and Polish Science Association as of the international academic collaboration program (Taiwan-Poland). And we appreciate Dr. Małgorzata Trzaskalik-Wyrwathe for her enthusiasm support and knowledge sharing on the study objects.

References

1. Brimblecombe, P., Price, C., Sabbioni, C., Van Balen, K.: Sustainable solutions in the conservation and protection of historic monuments and archaeological remains: a critical assessment of European research needs. In: Air Pollution and Cultural Heritage. Taylor & Francis (2004)
2. Frodl, W.: Pojęcia i kryteria wartościowania zabytków i ich oddziaływanie na praktykę konserwatorską. BmiOZ. Seria B, tom XIII, Warszawa (1966)
3. Kiesow, G.: Einfiihrung in die Denkmalpf1ege. Darmstadt (1982)
4. Lo, M.C., Tzeng, G.H.: Fuzzy MCDM Application for Strategy Evaluation. In: The 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), Taipei, Taiwan (2011)
5. Saaty, T.L.: Decision making with dependence and feedback: The Analytic Network Process. RWS Publications, Pittsburgh (1996)
6. Trzaskalik-Wyrwa, M.: Polskie pozytywy przenośne fenomen techniczny dawnego budownictwa organowego [Pólish-Positive Organ a Technological Phenomenon of the EarlyOrgan Making]. ZabYtki, nfI(10), s.62–s.67 (2006a)
7. Trzaskalik-Wyrwa, M.: Zrozumieć zabytkowe organy [ToUnderstand Historical Organs]. Ruch. Muzyczny (2), s.32–s.37 (2006b)
8. Trzaskalik-Wyrwa, M., Nowak, M., Trzaskalik, T.: Application of multicriteria analysis to restoration of historical portable organ. In: Trzaskalik, T. (ed.) Multiple Criteria Decision Making 2005. Scientific Publications: The Karol Adamiecki University of Economics in Katowice (2006)

9. Tseng, M.L.: A causal and effect decision-making model of service quality expectation using grey-fuzzy DEMATEL approach. Expert Systems with Applications 36(4), 7738–7748 (2009a)
10. Tseng, M.L.: Using extension of DEMATEL to integrate hotel service quality perceptions into a cause-effect model in uncertainty. Expert Systems with Applications 36(5), 9015–9023 (2009b)
11. Tzeng, G.H., Huang, J.J.: Multiple Attribute Decision Making: Methods and Applications. Taylor & Francis (2011)

Immunization of Networks via Modularity Based Node Representation

Tetsuya Yoshida and Yuu Yamada

Abstract. We propose an approach for immunization of networks via modularity based node representation. Immunization of networks has often been conducted by removing nodes with large centrality so that the whole network can be fragmented into smaller subgraphs. Since contamination is propagated among subgraphs (communities) along links in a network, besides centrality, utilization of community structure seems effective for immunization. However, despite various efforts, it is still difficult to identify true community labels in a network. Toward effective immunization of networks, we propose to remove nodes between communities *without* identifying community labels of nodes. By exploiting the vector representation of nodes based on the modularity matrix of a network, we propose to utilize not only the norm of vectors, but also the relation among vectors. Two heuristic scoring functions are proposed based on the inner products of vector representation and their filtering in terms of vector angle. Preliminary experiments are conducted over synthetic networks and real-world networks, and compared with other centrality based immunization strategies.

1 Introduction

Various resources are connected to each other and form networks, including the Internet, human networks such as social networks [5]. Links among resources makes it easier to exploit other resources by overcoming geographical or temporal distance. However, fast spreading of information over networks can have negative aspects, such as computer viruses or epidemics of diseases.

Tetsuya Yoshida · Yuu Yamada
Graduate School of Information Science and Technology,
Hokkaido University
N-14 W-9, Sapporo 060-0814, Japan
e-mail: {yoshida,yamayuu}@meme.hokudai.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 33–44.
springerlink.com

The spread of epidemics (e.g., virus) can occur through the interaction between nodes in a network. Removing contaminated nodes (or, vaccinating nodes) can prevent the spread of epidemics over the whole network. However, usually the amount of available doze is much smaller than the total number of nodes. Thus, it is important to selectively utilize the available doze for removing nodes from a network for the immunization of networks.

Since contamination is propagated among subgraphs (communities) along links in a networks, the standard approach is to immunize nodes which play the major role in propagation. Based on the assumption that such nodes are in some sense "central" nodes in networks, various heuristic immunization strategies have been proposed based on the notion of node centrality [7]. In addition, by assuming that community labels of nodes are available, an immunization strategy which utilizes the community labels was also proposed [4] based on a perturbation approach [10]. However, despite various efforts, it is still difficult to identify true community labels from networks [2, 7].

Toward effective immunization of networks, we propose to remove nodes between communities *without* identifying community labels of nodes. By exploiting the vector representation of nodes based on the modularity matrix of a network [6, 11], we propose to utilize not only the norm of vectors, but also the relation among vectors. Two heuristic scoring functions are proposed based on the inner products of vector representation and their filtering in terms of vector angle. Preliminary experiments are conducted over synthetic networks and real-world networks. Comparison with other centrality based immunization strategies shows that the proposed approach is promising.

2 Immunization of Networks

2.1 Preliminaries

We use a bold italic lowercase letter to denote a vector, and a bold normal uppercase letter to denote a matrix. $\mathbf{X}_{ij}$ stands for the element in a matrix $\mathbf{X}$, and $\mathbf{X}^T$ stands for the transposition of $\mathbf{X}$. $\mathbf{1}_n \in \mathbb{R}^n$ stands for a column vector where each element is 1.

Let n stands for the number of nodes in a network G, and m stands for the number of links in G [1]. Since most social networks are represented as undirected graph without self-loops [5], we focus on this type of networks in this paper.

The connectivity of a network is usually represented as a square matrix $\mathbf{A} \in \{0,1\}^{n \times n}$ which is called an adjacency matrix. $\mathbf{A}_{ij} = 1$ when the pair of vertices (v_i, v_j) is connected; otherwise, 0. For an undirected graph without self-loops, the corresponding adjacency matrix $\mathbf{A}$ is symmetric and its diagonal elements are set to 0. The vector $\boldsymbol{k} = \mathbf{A}\mathbf{1}_n$ denotes the degree vector, where k_i represents the degree (number of links) of i-th node in a network.

[1] We also call a network as a graph, a node as a vertex, and a link as an edge.

2.2 Immunization Strategies Based on Centrality

Various kinds of "node centrality" have been studied and utilized in order to identify important nodes in social network analysis [5, 7]. A node with large centrality plays an important role in a network in some sense. Thus, removing such node has been widely used as a heuristic immunization strategy.

Since nodes with many links can be considered as a hub in a network, the degree (number of links) of a node is called degree centrality. On the other hand, betweenness centrality focuses on the shortest path along which information is propagated over a network. By enumerating the shortest paths between each pair of nodes, betweenness centrality of a node is defined as the number of shortest paths which go through the node.

Similar to the famous Page Rank, eigenvector centrality utilizes the leading eigenvector of the adjacency matrix $\mathbf{A}$ of a network, and each element (value) of the eigenvector is considered as the score of the corresponding node. Based on the approximate calculation of eigenvector centrality via perturbation analysis, another centrality (called dynamical importance) was also proposed in [10]. By assuming that community labels of nodes in a network can be specified, perturbation analysis is utilized for approximately calculating the leading eigenvector among communities in [4].

2.3 Modularity of Networks

Besides immunization, community discovery from networks has also been studied [5, 7]. Modularity has been widely utilized as a quality measure of communities based on the identified community labels in a network [6]. Under the so-called *null model*, modularity Q of a network G is defined as:

$$Q = \frac{1}{2m} \sum_{C \in \mathcal{P}} \sum_{i,j \in C} (\mathbf{A} - \mathbf{P})_{ij} \tag{1}$$

where $\mathbf{P} = \boldsymbol{k}\boldsymbol{k}^T/2m$, $\boldsymbol{k}$ is the degree vector of G. $\mathcal{P}$ stands for the partition of G, and C runs over the communities in $\mathcal{P}$. Communities with larger modularity are considered as better partitioning of nodes in a network.

3 Immunization of Networks via Modularity Based Node Representation

3.1 Community Centrality Based on Modularity

For community discovery, it is shown that maximization of modularity in eq.(1) can be sought by finding the leading eigenvector of the following matrix, which is called a modularity matrix [6]:

$$\mathbf{B} = \mathbf{A} - \mathbf{P} \tag{2}$$

By utilizing several eigenvectors of $\mathbf{B}$ in eq.(2) with the largest positive eigenvalues, the modularity matrix $\mathbf{B}$ can be approximately decomposed as:

$$\mathbf{B} \simeq \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \tag{3}$$

where $\mathbf{U}$=[$\boldsymbol{u}_1, \cdots, \boldsymbol{u}_q$] are the eigenvectors of $\mathbf{B}$ with the descending order of eigenvalues, and Λ is the diagonal matrix with the corresponding eigenvalues.

A new node score called community centrality was proposed in [6] based on the above approximated decomposition. First, a new data representation (data matrix) $\mathbf{X}$ was proposed based on eq.(3) as:

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}^{1/2} \tag{4}$$

Each row of $\mathbf{X}$ corresponds to the representation of a node. Hereafter, the i-th node in a network is represented as a column vector $\boldsymbol{x}_i$ based on eq.(4). With this vector representation, community centrality (CC) was defined as [6] [2]:

$$CC(\boldsymbol{x}_i) = \boldsymbol{x}_i^T \boldsymbol{x}_i \tag{5}$$

3.2 Community Boundary Nodes Based on Modularity

As in the heuristic strategies in Section 2.2, removing nodes with large community centrality seems a reasonable immunization strategy. However, a node with large community centrality would be around the "center" of a community by definition. Thus, although a node with larger community centrality would be connected to other nodes in the same community, since it would not be connected to nodes in other communities, it might not be so effective to fragment communities in a network.

Instead of identifying and removing central nodes in communities, we propose to remove nodes which reside between communities. If the community label of each node in a network can be identified, it would be possible to estimate community centers as well as nodes between communities. However, although various methods have been proposed for community discovery from networks [9, 8, 6, 7, 11], it is still difficult to identify the true community labels. Besides, it is proved that finding the communities with maximum modularity for a network is NP-complete [2].

Toward effective immunization of networks, we try to identify and remove nodes between communities *without* estimating or utilizing community labels of nodes based on the vector representation in eq.(4) from the modularity matrix.

---

[2] Both the square sum ($\boldsymbol{x}_i^T \boldsymbol{x}_i$) and the absolute sum (norm) ($||\boldsymbol{x}_i||$) were called community centrality in [6].

3.3 A Hyper-Plane Based Filtering Approach

Although we cannot utilize community labels of nodes, whether a node reside between communities still depends on the relation among nodes in a network. The notion of community centrality in eq.(5) only utilizes the norm (or, square norm) of the vector representation $\boldsymbol{x}_i$ for each node in eq.(4). In order to identify nodes between communities, we propose to utilize the relation among nodes in a network as well.

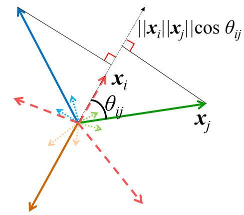


Fig. 1 Modularity based representation

For each node representation $\boldsymbol{x}_i$, we consider the contribution from another nodes $\boldsymbol{x}_j, j \neq i$, and quantify the contribution as the inner product $\boldsymbol{x}_i^T \boldsymbol{x}_j$. The inner product represents to what extent other node $\boldsymbol{x}_j$ contributes to the direction $\boldsymbol{x}_i$ (see Fig. 1). Note that $\theta_{ij} = \cos^{-1} \frac{\boldsymbol{x}_i^T \boldsymbol{x}_j}{||\boldsymbol{x}_i||||\boldsymbol{x}_j||}$ represents the angle between vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

We categorize nodes in a network into the following three types:

i) nodes around the center of a community
ii) nodes in the fringe of a community
iii) nodes between communities

If a node $\boldsymbol{x}_i$ is in **i)**, it would be "far away" from other node $\boldsymbol{x}_j$ in different communities in terms their angle θ_{ij}. Thus, $\boldsymbol{x}_i^T \boldsymbol{x}_j$ would be small for most nodes unless $\boldsymbol{x}_j$ is in the same community. On the other hand , if a node $\boldsymbol{x}_i$ is in **ii)**, by definition of community centrality, its norm $||\boldsymbol{x}_i||$ would be small. Thus, $\boldsymbol{x}_i^T \boldsymbol{x}_j$ would be small. Finally, if a node $\boldsymbol{x}_i$ is in **iii)**, its norm $||\boldsymbol{x}_i||$ would not be too small, and $\boldsymbol{x}_i^T \boldsymbol{x}_j$ can be counted from nodes in different communities.

Based on the above argument, for each node $\boldsymbol{x}_i$, we consider the contribution from another nodes $\boldsymbol{x}_j$ as $\boldsymbol{x}_i^T \boldsymbol{x}_j$, and sum up the contributions. However, simple sum over all nodes cannot be utilized from the following observation [6].

Observation 1. *For an undirected network with n nodes, for each node i in the network, $\sum_{j=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_j = 0$.*

Proof. By the definition of $\mathbf{P}$ in eq.(1), for an undirected network G with n nodes, $\mathbf{1}_n$ is the eigenvector of $\mathbf{B}$ with eigenvector 0. Since we select several eigenvectors with the largest positive eigenvalues to construct $\mathbf{X}$ in eq.(4), all the column vectors of $\mathbf{U}$ are orthogonal to $\mathbf{1}_n$. Thus, since $\mathbf{1}_n^T \mathbf{X} = \mathbf{0}_q^T$ holds, $\sum_{j=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_j = \boldsymbol{x}_i^T (\sum_{j=1}^{n} \boldsymbol{x}_j) = \boldsymbol{x}_i^T (\mathbf{1}_n^T \mathbf{X})^T = \boldsymbol{x}_i^T \mathbf{0}_q = 0$. $\square$

From the above observation, instead of *all* the nodes, it is necessary to selectively consider the contribution from other nodes. Since the inner product represents the contribution of other node $\boldsymbol{x}_j$ to the direction of $\boldsymbol{x}_i$, if $\boldsymbol{x}_i^T \boldsymbol{x}_j < 0$, $\boldsymbol{x}_j$ does not contribute to $\boldsymbol{x}_i$. Thus, we propose the following node score:

$$hpf(\boldsymbol{x}_i) = \sum_{j:\ \boldsymbol{x}_i^T \boldsymbol{x}_j \geq 0} \boldsymbol{x}_i^T \boldsymbol{x}_j \tag{6}$$

In eq.(6), only nodes with non-negative contributions are summed up. As shown in Fig. 2, when we consider a bisecting hyper-plane perpendicular to $\boldsymbol{x}_i$, contributions from nodes in the same side with $\boldsymbol{x}_i$ are considered in eq.(6). Thus, we name the node score in eq.(6) as Hyper-Plane based Filtering (hpf).

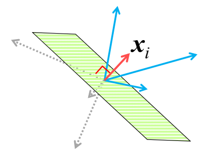


Fig. 2 Hyper-plane based filtering

3.4 A Conic Filtering Approach

For each node i, the score in Section 3.3 filters out contributions from the nodes in the opposite side of the hyper-plane. However, since the score is based on the inner product, a node with large norm ($||\boldsymbol{x}_i||$) tends to have larger score in eq.(6). Thus, in addition to the nodes between communities, nodes around the center of a community also tend to have larger score in eq.(6).

Community centrality in eq.(5) only considers the norm of each vector *separately*. In addition to the norm of vectors, we propose to exploit the relation among vectors in order to identify nodes between communities. Note that θ_{ij} represents the angle between vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Especially, $\cos\theta_{ij}$ has been widely utilized as cosine similarity in text analysis.

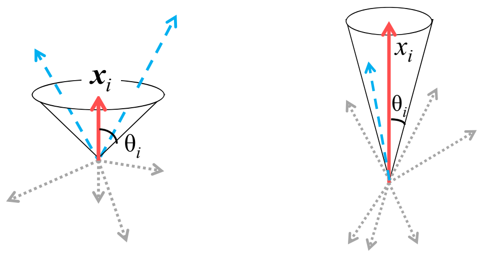


Fig. 3 Conic-filtering

Based on the above argument, we propose another filtering based scoring funciton based on a cone around each node. For the modularity matrix of a network, suppose we use q eigenvectors with positive eigenvalues in eq.(3) [3]. As illustrated in Fig. 3, for each node $\boldsymbol{x}_i$, we consider a (hyper-) cone around $\boldsymbol{x}_i$ in $\mathbb{R}^q$. The tip of the cone is the origin of $\mathbb{R}^q$, the height is $h_i = ||\boldsymbol{x}_i||$, and the conical angle θ_i (see Fig. 3). By considering a (hyper-) cone for each node $\boldsymbol{x}_i$, we filter out the nodes with which the angle exceeds θ_i for each node i.

[3] We assume $q > 1$, since we consider vector representation of nodes.

The conical angle θ_i for each node controls to what extent other nodes are filtered out. When vectors are randomly distributed around the origin in $\mathbb{R}^q$, the number of vectors (nodes) inside a cone would be proportional to the volume of the cone. By exploiting the property of vector representation in eq.(4), we set the conical angle θ_i so that the volume of the cone is invariant to all the nodes. To realize this, for each node, we set the conical angle θ_i as:

$$\theta_i = \arctan(\frac{c}{h_i^q})^{1/(q-1)} \tag{7}$$

where $h_i = ||\boldsymbol{x}_i||$, and c is some positive constant.

With the above conical angle θ_i, the following property holds:

Property 1. When the conical angle θ_i is set as eq.(7) for each node, the volume of the cone is invariant to all the nodes.

Proof. Since each cone is in $\mathbb{R}^q$, its volume V_i is proportional to the product of its height h_i ($=||\boldsymbol{x}_i||$) and the volume of its base. Since the base of a cone in $\mathbb{R}^q$ forms a $q-1$ dimensional hyper-sphere, by denoting its radius as r_i, the volume of the base is proportional to r_i^q, i.e., $r_i \propto r_i^q$. Furthermore, with the conical angle θ_i and h_i, we can represent r_i as $r_i = h_i \tan\theta_i$. Thus, the volume of the cone $\propto h_i^q(\tan\theta_i)^{q-1}$. By setting $h_i^q(\tan\theta_i)^{q-1} = c$ for some constant c, we obtain eq.(7). □

For instance, for a node with large norm (i.e., around a community center), its conical angle gets smaller (middle figure in Fig. 3). On the other hand, for a node with small norm (i.e., in the fringe of a community), its conical angle gets larger (right figure in Fig. 3).

The above idea is formalized as the following node scoring function, which is called a Conic Filtering (cf):

$$cf(\boldsymbol{x}_i) = \sum_{j:\ \theta_{ij} \le \theta_i} \boldsymbol{x}_i^T \boldsymbol{x}_j \quad \textit{with } \theta_i \textit{ in eq.}(7) \tag{8}$$

The constant c in eq.(7) is a parameter, and we set as c=1 in the experiments.

4 Preliminary Evaluations

4.1 Experimental Settings

Datasets. Preliminary evaluations were conducted over both synthetic networks and real-world networks. Utilized networks are shown in Table 1 and Table 2.

As synthetic networks, an *ad hoc* random network ER in Table 1 was generated using Erdös-Rény (ER) model [3]. The link probability was set to

Table 1 Synthetic Networks

dataset	# nodes	#links (ave)
ER	500	1929.7
BA	500	1946.3

Table 2 Real-World Social Networks

dataset	#nodes	#links
dolphins	62	159
polbooks	115	613
netscience *	379	914
IV'04 *	112	255

0.0155. The scale-free network BA in Table 1 was generated using Barabási-Albert (BA) model [1] by setting the degree distribution $p(k) \propto k^{-3}$, where k denotes the degree of a node. The initial degree was set to 4. Since these are random networks, we constructed 10 networks for each type and report the average result.

The first three real-world networks in Table 2 are available as GML (graph markup language) format[4]. The last network (IV'04) is a co-authorship networks among researchers[5]. Since the third and fourth networks (netscience* and IV'04*) are disconnected, we conducted experiments on the maximum connected component in these networks.

Quality Measures. By following the quality measure in [4], the relative size S of the largest connected component (LCC) in a network was measured against the node occupation probability p. After removing some nodes from a network with n nodes, these are calculated as:

$$S = \frac{\#nodes\ in\ LCC}{n}, \quad p = \frac{\#remaining\ nodes}{n} \tag{9}$$

The smaller S is, the better a immunization strategy of networks is, since it can prevent the spreading of contamination over the whole network.

Compared Methods. For comparison, immunization strategies based on node centrality in Section 2.2 and Section 3.1 were evaluated. The node with the maximum centrality was repeatedly selected and removed in each strategy:

D : degree centrality (gray line with "x" in Fig. 4 and Fig. 5)
B : betweenness centrality (black line with "+")
RB : repeated calculation of betweenness centrality (black dotted line)
EVC : eigenvector centrality (yellow line with square)
CC : community centrality (blue line with upper triangle)

The proposed methods are shown with lower triangle (hpf (green line) and cf (red line) in Fig. 4 and Fig. 5).

Except for RB, node centrality was calculated only once with respect to the whole network (including our proposals). On the other hand, since betweenness centrality (B) is known to be effective (but with huge computational

[4] http://www-personal.umich.edu/˜mejn/netdata/
[5] http://iv.slis.indiana.edu/ref/iv04contest/

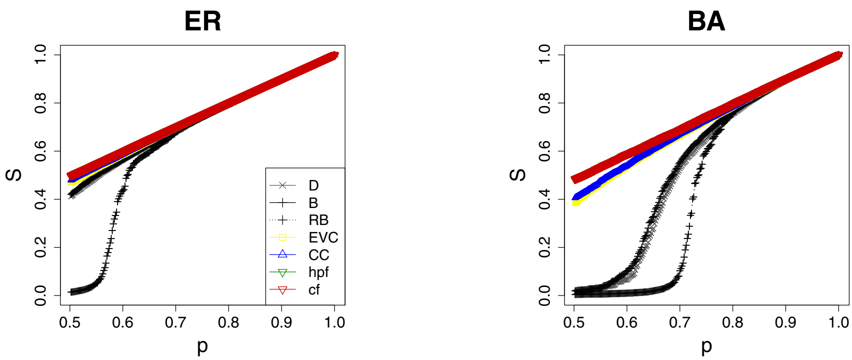


Fig. 4 Results (synthetic networks)

cost) for immunization of networks, betweenness centrality was re-calculated in RB whenever a node was removed from a network.

4.2 Results on Synthetic Networks

Although explicit community labels of nodes are not utilized nor identified, since our methods (hpf and cf) and CC are based on the modularity of a network, these methods would not be effective for random networks (with low modularity). Results on synthetic networks (average of 10 runs) are shown in Fig. 4. The horizontal axis is the node occupation probability p, and the vertical one is the relative size S of LCC in eq.(9).

As shown in Fig. 4, the results confirmed the above hypothesis, since S did not decrease (except for the number of removed nodes) with the methods based on the modularity. On the other hand, although other methods were also not effective for random networks (i.e. ER model), still RB and EVC showed a drop of S around p=0.7 for scale-free networks (i.e., BA model). Since "hub" nodes act as bridges among distant nodes in scale-free networks, identifying and removing such nodes with RB or EVC contributes to fragmenting networks.

4.3 Results on Real-World Networks

Results on real world networks in Table 2 are shown in Fig. 5. As expected, with large computational cost, RB showed the best performance for all networks. Except for iv04, the performance of EVC (based on the leading eigenvector of adjacency matrix) was better than that of CC (based on the leading eigenvector of modularity matrix).

Although the proposed methods (hpf and cf) and CC are based on the the vector representation in eq.(4), our proposals outperformed CC for polbooks

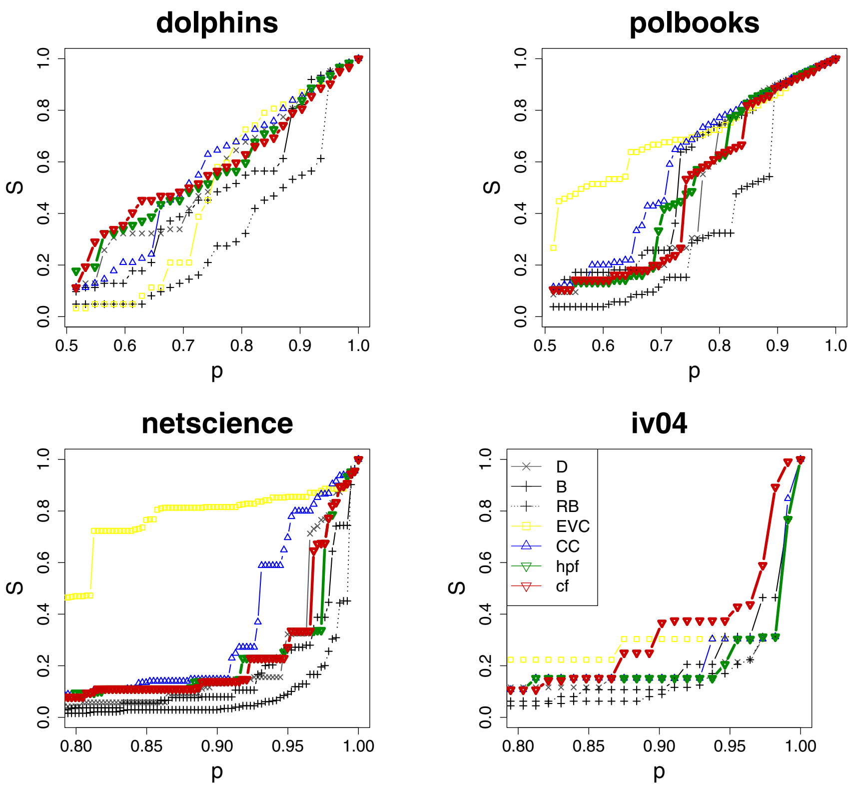


Fig. 5 Results (real-world networks)

and netscience, and were almost equivalent for dolphins and iv04. In addition, these methods slightly outperformed EVC for polbooks and netscience. However, not so much difference was observed between the proposed methods: hpf showed almost equivalent performance with RB for iv04, but cf outperformed hpf slightly for polbooks. In addition, they were not so effective for dolphins.

4.4 Discussions

Besides the immunization of networks, modularity in eq.(1) has been widely utilized for community discovery from networks. Although finding the communities with maximum modularity for a network is NP-complete [2], toward effective immunization of networks, we proposed to utilize vector representation of nodes based on the modularity of a network, *without* identifying community labels.

The proposed methods (hpf and cf) outperformed CC in most networks in terms of S in eq.(9). However, despite the difference in the filtering process, not so much difference was observed between them in the experiments. In addition, the performance of proposals was rather similar to that of D for the real-world networks in the experiment. The utilized networks might be rather small to reveal the difference of immunization strategies. Larger networks with heterogenous substructures (e.g., networks with both dense subgraphs and sparse subgraphs) will bring the difference of strategies.

As shown in [2], generally there exist some trade-offs between the quality measures (e.g., modularity, relative size S in eq.(9)) and the computational cost. Proposed methods could not outperform RB in terms of S, albeit the latter requires much computational cost. We plan to conduct more evaluations over larger networks (with heterogeneous substructures) to investigate the effectiveness of our proposals in terms of the trade-offs.

5 Concluding Remarks

We proposed an approach for immunization of networks via modularity based node representation. Since contamination is propagated among subgraphs (communities) in a networks, we proposed to remove nodes between communities *without* identifying their community labels so that the whole network can be fragmented into smaller subgraphs. By exploiting the vector representation of nodes based on the modularity matrix of a network, we proposed to utilize not only the norm of vectors, but also the relation among vectors. Two heuristic scoring functions were proposed based on the inner products of vector representation and their filtering in terms of vector angle.

Preliminary experiments are conducted over synthetic networks and real-world networks, and compared with other centrality based immunization strategies. The results are encouraging, and indicate that the proposed approach is promising. We plan to conduct more in-depth analysis of vector representation and their relations, and extend the proposed method in near future.

Acknowledgments. This work is partially supported by Casio Science Promotion Foundation.

References

1. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science 286, 509–512 (1999)
2. Brandes, U., Delling, D., Gaertler, M., Görke, R., Hoefer, M., Nikoloski, Z., Wagner, D.: On modularity clustering. IEEE Transactions on Knowledge and Data Engineering 20(2), 172–188 (2008)
3. Erdös, P., Rény, A.: On random graphs. Publicationes Mathematicae 6, 290–297 (1959)
4. Masuda, N.: Immunization of networks with community structure. New Journal of Physics 11, 123018 (2011), doi:10.1088/1367-2630/11/12/123018
5. Mika, P.: Social Networks and the Semantic Web. Springer (2007)

6. Newman, M.: Finding community structure using the eigenvectors of matrices. Physical Review E 76(3), 036104 (2006)
7. Newman, M.: Networks: An Introduction. Oxford University Press (2010)
8. Pons, P., Latapy, M.: Computing communities in large networks using random walks. Journal of Graph Algorithms 10(2), 191–218 (2006)
9. Raghavan, U., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Physical Review E 76, 036106 (2007)
10. Restrepo, J.G., Ott, E., Hunt, B.R.: Characterizing the dynamical importance of network nodes and links. Physical Review Letters 97, 094102 (2006)
11. Yoshida, T.: Toward finding hidden communities based on user profile. Journal of Intelligent Information Systems (2011) (accepted)

Information Fusion Approach in Computer-Aided Detection of Pathological Changes

Mirosław Trzupek and Marek R. Ogiela

Abstract. The fusion of information is a domain of research in recent years. One of the more important fields in which such techniques could be applied is the analysis of medical images showing the complicated morphology of selected structures or organs. These techniques allow the image to be analysed comprehensively by identifying the most important parameters of given structures based on aggregating information obtained at various stages of processing and analysing this image. This paper presents a proposal for the semantic analysis of coronary vascularisation carried out using 3D reconstructions obtained in spiral CT examinations. This will be illustrated by the recognition of pathological changes in coronary arteries of the heart.

Keywords: image understanding systems, intelligent medical image processing and analysis.

1 Introduction

In recent years, it is observed an intense progress in the field of medical diagnostic imaging devices, which significantly increased the number of possible diagnostic imaging associated with a given patient, which then must be correctly interpreted by the diagnostician. On the other hand, more and more emphasis is put on improving the quality and effectiveness of medical care, and in particular to minimize possible mistakes that may occur during the evaluation and interpretation of medical imaging data. To meet the expectations of reducing medical errors, improving the efficiency of interpretation of large imaging data sets and to improve access and exchange of information, it is necessary to support the use of advanced methods of computer-aided medical image diagnosis. In this field the authors of this paper have conducted research and systematically put

Mirosław Trzupek · Marek R. Ogiela
AGH University of Science and Technology,
Al. Mickiewicza 30, PL-30-059 Krakow, Poland
e-mail: {mtrzupek,mogiela}@agh.edu.pl

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 45–50.
springerlink.com

forward various proposals of new solutions improving the capability to support medical image diagnostics [1, 2, 5, 7, 11].

Previous publications by the authors focused on images acquired by 64-slice spiral computed tomography (CT) [1, 7, 9, 11], whereas this article describes attempts to adapt information aggregation techniques and image recognition methods to the semantic analysis of coronary vascularisation reconstructions acquired by a 128-slice spiral CT apparatus.

A look at the capability of carrying out computer analyses of such medical images shows that the set of tools supporting the visual assessment of coronary vascularisation reconstructions can be considered satisfactory. This is not, however, true of the availability of tools supporting the assessment and the right semantic interpretation of a given image. Such algorithms, which would imitate the thought processes taking place in the mind of the physician and thus support a comprehensive description of the analysed image and formulating the appropriate diagnosis, are not available yet. Nevertheless, attempts made in this regard (also by the authors of this article) are beginning to produce systems of varied effectiveness and support range [1, 2, 11, 12]. What is more, the rapid development of medical computer technologies observed in recent years means that when a new generation of diagnostic apparatuses appears, those quite recently used in everyday clinical practice become obsolete and are rarely used. This fact should be borne in mind when developing systems for medical image interpretation, as adapting existing solutions to apparatuses of a newer generation may require elaborating completely new solutions, which frequently becomes yet another scientific challenge. For this reason it is worth emphasising that the scientific proposals of the authors of this article are highly universal in this regard

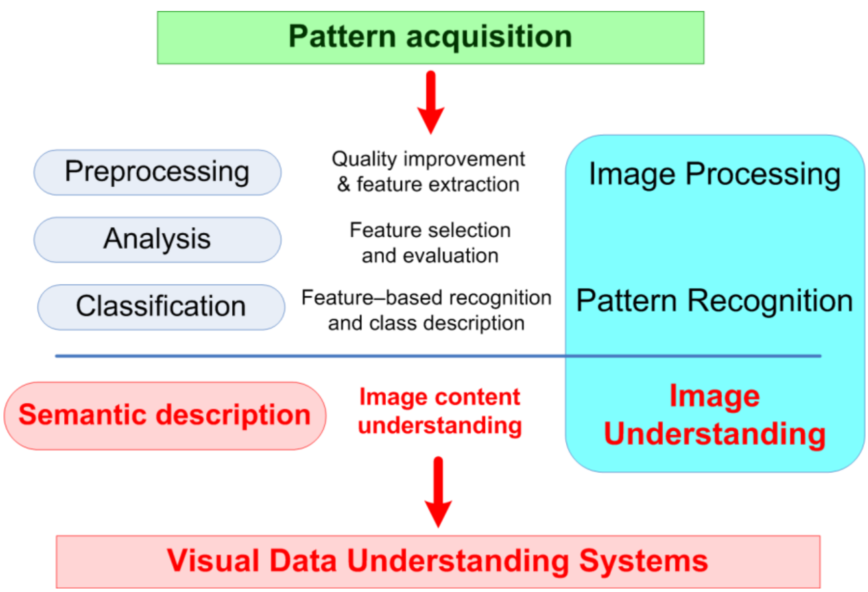


Fig. 1 Main stages in computer image understanding

and represent a set of advanced techniques allowing various medical images to be semantically analysed, regardless of their quality or other characteristics caused by the technical, hardware properties of the diagnostic apparatuses used [1, 4, 5, 9, 11]. These techniques allow the image to be analysed comprehensively by identifying the most important parameters of given structures, and then analysing their meaning by aggregating various information obtained at the stage of pre-processing or classifying individual structures. This operation produces a semantic description in the form of a semantic record which the specialist can use when formulating his/her diagnosis. This research is motivated by statistics showing that recent years have seen an increase in the incidence of diseases connected with the patency of arteries, including the coronary arteries, which are particularly important for the health and life of the patient [12].

To structure further considerations, Fig. 1 shows the basic stages and purposes of advanced semantic image analysis.

2 Computer-Aided Detection of Pathological Changes Based on the Example of Coronary Arteries

Canvas for the solutions presented below are the research efforts, which have already been successfully undertaken by the authors for planar images acquired by traditional angiography to diagnose patients [4, 5]. Such images showed coronary arteries with a contrast medium in them, allowing the lesions present in them to be visualised. Then, mathematical linguistic methods were used to create a generalised description of the image, and the application of sequential attributed grammars allowed lesions to be detected in the analysed coronarography images.

The transition to analysing images acquired using spiral computed tomography, particularly the 64-slice or 128-slice types, requires the use of new tools, as images of this type, unlike those from traditional angiography, have the form of a three-dimensional structure of coronary vessels visualised in 3D space. The additional degree of freedom caused by the transition from a 2D to a 3D visualisation requires the use of more advanced tools, as these images feature a greater number of visible details, and thus contain more differences due to individual patient characteristics [8]. For this reason, instead of sequential grammars, the 3D structure of coronary vascularisation has to be described using more powerful (in the sense of the description power) formalisms, i.e. graph grammars. Such grammars represent tools that are definitely more powerful in describing images than tree or sequential grammars [4]. To make it possible to represent the analysed structure of coronary vascularisation from tomograms with a graph, it is necessary to execute a number of pre-processing operations of the image. One of them is skeletonising the structure in order to find the points in which main vessels branch into collateral ones. Points found in this way become the vertices of the graph modelling the 3D structure of coronary arteries, while the edges within the graph thus produced, which connect individual vertices, represent the spatial relationships between individual arteries. This yields a graph model of the analysed coronary arteries which precisely maps the topological characteristics of the analysed 3D structure. For the above model, the authors proposed independent graph grammars for the left and the right artery, and with a suitably defined set of productions, these grammars generate languages in the form of graphs modelling the analysed 3D structures of coronary vascularisation. This method has been described in detail in the article [11] using visualisations acquired with a 64-slice spiral CT apparatus as examples. It is worth noting that, additionally, the set of image data was divided into three subsets depending on the topology of the coronary vascularisation (balanced distribution of arteries, right artery dominant, left artery dominant). This differentiation aimed at obtaining additional information about the significance of the danger to the patient's health depending on the location of the lesion in a given type of coronary vascularisation (e.g. a stenosis found in the left coronary artery constitutes a greater danger to the life of a patient with a dominant left artery than if he/she had a dominant right artery). At the following stages, the graph thus produced, modelling the examined structure, undergoes detailed analyses to detect pathological locations in particular arteries represented by different edges of the graph. Sequential attributed grammars are used for this purpose, and the methodology applied is similar to how the locations of stenoses were detected in planar images acquired during diagnostic examinations using traditional angiography [4, 5]. Once a stenosis has been detected in a given artery, its magnitude, size and type (concentric or eccentric) have been determined, and this information has been confronted with the graph model representing the entire coronary vascularisation structure, it is known in which artery the stenosis is found, so therapeutic premises can be established about the threat to the examined patient's life.

All the above information also makes it possible to create a semantic record containing the semantic description of the examined structure. What is more, such descriptions can form a kind of indexing key, allowing image data to be archived and searched for. The described methodology, making use of advanced mathematical linguistic mechanisms, has been described in detail in the authors'

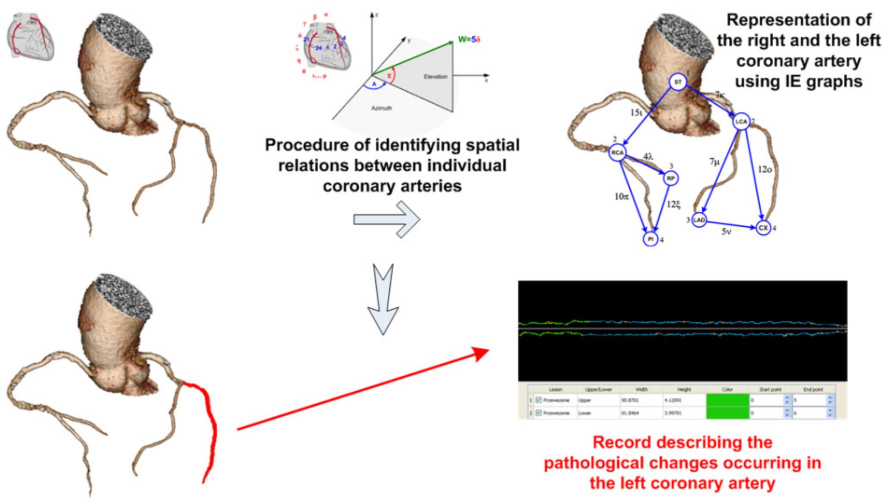


Fig. 2 Main stages in the analysis and recognition of morphological changes for the right and the left coronary artery

articles dealing with this subject [1, 7, 9, 11]. Below (Fig. 2), a visualisation acquired by 128-slice spiral CT is used as an example for a step-by-step presentation this methodology.

3 Summary and Further Work

The obtained results prove the universality of the proposed approach and great opportunities for using the previously proposed solutions to analyse images acquired with a 128-slice CT apparatus as well. The set of test data was small but experimental results reveal that the proposed approach can successfully detect pathology changes in coronary arteries including the determination of locations, their number, extend and the type (i.e. concentric or eccentric). Moreover, thanks to using comprehensive analysis consist in aggregating diagnostically significant information about the morphology of individual vessels, their topological relationships and functional condition, it is possible to obtain additional information about the significance of the danger to the patient's health depending on the location of the lesion. However, further advanced standardisation work in this regard as well as the verification and improvement of the presented solutions using a much larger set of image data will be necessary in the future.

What is more, the research work also covers the application of the proposed solutions for indexing and quickly finding specialised image data in distributed medical databases. Such searches could use semantic keys and would thus allow finding cases meeting specified substantive conditions related to image contents. This would make it possible to solve at least some of the problems of intelligently archiving this type of data in medical multimedia databases and finding semantic image data fulfilling certain semantic criteria defined using example image patterns. Such patterns would significantly extend the presented concept of semantic interpretation of a given image, which is based on aggregating information obtained at various stages of processing and analysing this image. The final semantic interpretation would also be dependent on previously correctly interpreted patterns stored and found in specialised image repositories.

Acknowledgments. This work has been supported by the National Science Centre, Republic of Poland, under project number N N516 478940.

References

1. Hachaj, T., Ogiela, M.R.: Automatic detection and lesion description in cerebral blood flow and cerebral blood volume perfusion maps. Journal of Signal Processing Systems for Signal, Image, and Video Technology 61(3), 317–328 (2010)
2. Hachaj, T., Ogiela, M.R.: A system for detecting and describing pathological changes using dynamic perfusion computer tomography brain maps. Computers in Biology and Medicine 41, 402–410 (2011)
3. Ogiela, L., Ogiela, M.R.: Cognitive Techniques in Visual Data Interpretation. Springer, Heidelberg (2009)

4. Ogiela, L., Ogiela, M.R.: Advances in Cognitive Information Systems, Cognitive Systems Monographs, vol. 17. Springer, Heidelberg (2012)
5. Ogiela, L., Tadeusiewicz, R., Ogiela, M.R.: Cognitive Computing in Intelligent Medical Pattern Recognition Systems. In: Huang, D.-S., Li, K., Irwin, G.W. (eds.) Intelligent Control and Automation. LNCIS, vol. 344, pp. 851–856. Springer, Heidelberg (2006)
6. Ogiela, M.R., Tadeusiewicz, R., Ogiela, L.: Intelligent Semantic Information Retrieval in Medical Pattern Cognitive Analysis. In: Gervasi, O., Gavrilova, M.L., Kumar, V., Laganá, A., Lee, H.P., Mun, Y., Taniar, D., Tan, C.J.K. (eds.) ICCSA 2005. LNCS, vol. 3483, pp. 852–857. Springer, Heidelberg (2005)
7. Ogiela, M.R., Tadeusiewicz, R., Trzupek, M.: Picture grammars in classification and semantic interpretation of 3D coronary vessels visualisations. Opto-Electronics Review 17(3), 200–210 (2009)
8. Sirol, M., Sanz, J., Henry, P., et al.: Evaluation of 64-slice MDCT in the real world of cardiology: A comparison with conventional coronary angiography. Archives of Cardiovascular Diseases 102(5), 433–439 (2009)
9. Trzupek, M.: Semantic Interpretation of Heart Vessel Structures Based on Graph Grammars. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010. LNCS, vol. 6374, pp. 81–88. Springer, Heidelberg (2010)
10. Trzupek, M.: Semantic Modelling of Coronary Vessel Structures in Computer Aided Detection of Pathological Changes. In: Tjoa, A.M., Quirchmayr, G., You, I., Xu, L. (eds.) ARES 2011. LNCS, vol. 6908, pp. 220–227. Springer, Heidelberg (2011)
11. Trzupek, M., Ogiela, M.R., Tadeusiewicz, R.: Intelligent image content semantic description for cardiac 3D visualisations. Engineering Applications of Artificial Intelligence 24, 1410–1418 (2011)
12. Wang, Y., Liatsis, P.: A Fully Automated Framework for Segmentation and Stenosis Quantification of Coronary Arteries in 3D CTA Imaging. In: 2009 Second International Conference on Developments in eSystems Engineering DESE, pp. 136–140 (2009)
13. Yusuf, S., et al.: Global burden of cardiovascular diseases, Part I. General considerations, the epidemiologic transition, risk factors, and impact of urbanization. Circulation 104, 2746–2753 (2001)

Manufacturing Intelligence to Forecast the Customer Order Behavior for Vendor Managed Inventory

Chen-Fu Chien, Chia-Yu Hsu, and Sheng-Chiao Lin

Abstract. As semiconductor foundry is built-to-order, the variation of demands of various customers affect the total quantity of customer order directly, and thus create variability on inputted wafer start on wafer fabrication facility (fab), which will affect the WIP bubble and cycle time as well as the capacity utilization and profitability. This research aims to develop a manufacturing intelligence methodology to predict customer demands based on their behaviors and historical data to extract useful information to support the decision maker for manufacturing strategy and production plan in light of demand uncertainty and market fluctuation. In particular, we proposed a manufacturing intelligence framework in which distributed lags structure and neural tree were employed to analyze the relationships among customer finished goods, order, and other decision factors. An empirical study was conducted for validation. The derived empirical rules can effectively help the decision maker to make timely production decisions given different order situations while maintaining fab utilization and cycle time well.

Keywords: manufacturing intelligence, vendor managed inventory (VMI), neural tree, distributed lags structure, forecast, semiconductor manufacturing.

1 Introduction

The semiconductor companies rely on competitive advantages different from other industries. Similar advanced equipment and processes are applied to produce a variety of semiconductor products including microprocessors, memories, digital signal processor, and application-specific logic. Many semiconductor companies compete on selling interchangeable products or providing foundry services with

---

Chen Fu Chien · Sheng Chiao Lin
Department of Industrial Engineering and Engineering Management,
National Tsing Hua University, 30013, Hsinchu, Taiwan
e-mail: `cfchien@mx.nthu.edu.tw`

Chia-Yu Hsu
Department of Information Management, Yuan Ze University, 32003, Chungli, Taiwan

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 51–60.
springerlink.com

similar process technologies to customers worldwide. When the lots finish the manufacturing process, the downstream customers decide to take the lots right away or to be held in the fab's warehouse according to the trend of market and capacity collocation. VMI (Vendor Managed Inventory) is defined as the quantity of finished goods that would be stored in the warehouse which could help the engineer to distribute the capacity to customers. This kind of inventory could regard as the buffer for the customers. In supply chain management, the upstream and downstream can make a contract for inventory level of the finished good by exchanging the information of sales data and inventory level. If the upstream and downstream do not have this kind of contract, the upstream vendor has to forecast the downstream customers' behavior of queuing finished goods. With the advanced information technology, data mining approaches have been applied to explore the huge database automatically or semi-automatically and extract useful rules or patterns for improve the decision quality [2]. Furthermore, manufacturing intelligence approaches have been developed to derive decision rules for enhancement of operation efficiency and effectiveness [1, 3, 4, 9]. Although there are a number of data mining approaches to extract rules or patterns to manage customer relationship, little research has been done on exploration of customer behavior in semiconductor industry.

This study aims to develop a manufacturing intelligence framework, in which the data mining methods are developed to explore the behavior of customers for VMI from historical transactions in semiconductor foundry. First, this study conducts the distributed lags structure in the pattern extracted model. The previous attributes might have extended effect on the performance of customer in latter period. The distributed lags structure could detect the delay effect of previous period on recent period. Next, The VMI with similar performance will be clustered to build different performance clusters. Then, the neural tree is constructed to improve the understandability by back-propagation (BP) algorithm. In particular, Taguchi method is used to determine the parameters of BP algorithm. Furthermore, the proposed framework is validated by the empirical data which were conducted from a semiconductor fabrication and examined by the domain experts.

2 Fundamental

The semiconductor fabrication produces the wafers according to the production plan to meet the delivery targets, while enhancing the quality, maintaining capacity utilization, and reducing the cost [5]. The customers' shipping quantity are different to the wafer out because of the order modification of the customers, fab yield gain or loss, customers close the order, fab abnormal event impact, manufacturing capability of fab, customers design issue, extra service of IC assembly and IC Final test, other special requirement of the customers. VMI is one kind of inventory management method to control the sales data and quantity of inventory. After the vendor receiving the sales data from downstream customers and the inventory level, the vendor replenish the stocks according to the pre-set inventory level. Base on the sharing of important information between the vendor and the retailer, both of them can improve demand forecasting, replenish stock planning and promotion management.

The key to VMI's success is good connectivity between the customer and the supplier through which the suppliers have direct access to consumption information in order to make better forecast and better respond to the customers' inventory needs in terms of quantities to ship and locations to replenish [8]. Thus, it could be known that to construct a good VMI forecasting model needs a connection between the vendor and the retailer, or needs an operation research model to optimize the VMI quantity. This study develops a manufacturing intelligence framework by adapting the vendor's historical data which obtained inside the vendor to forecast the VMI. This framework could help the decision maker to allocate the capacity and reduce the inventory cost.

3 Proposed Framework for VMI Forecast

3.1 Problem Definition

To forecast VMI of semiconductor fabrication is a complex task because VMI is influenced by several complex economical and manufacturing factors such as semiconductor industry market, customer demand, fab manufacturing capability and company sales policy. The variation of prosperity and the production strategy of customers often affect the quantity of fabrication output. In addition, the demand is higher than supply in high season, the fabrication often shifts the surplus order to latter period to balance production rate and increase capacity for extra order. In low season, the fabrication allocates the order equally to every period for capacity balance and reduction of the capacity waste. Therefore, VMI needs to be managed and forecasted accurately through the key performance indices (KPIs).

3.2 Data Preparation

The distributed lags method is used to explore the influence of considered attributes in former period to current VMI. Some surplus columns of useless lags of predictors are removed. For example, there are seven data firstly in Fig. 1 and assume x is one predictor for y. The lags period is determined to be three, then one predictor x become three predictors, x_1, x_2 and x_3. After the data shift, the number of data decreases to four. Therefore, the number of data for analyzing is determined by the total number of data and the lags periods.

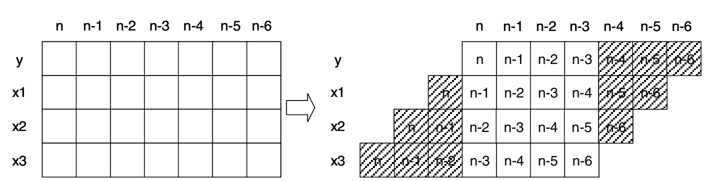


Fig. 1 Distributed lag structure

Next, a two stage clustering method is applied to divide training data into several groups based on the similarity of VMI quantity. The clustering results are set for the classification target for neural tree construction. First, Ward's method is used to determine the appropriate number of clusters [11]. In each step of the clustering process, the nearest two clusters will be combined to minimize the variance within clusters. Then, Self Organizing Map (SOM) [7] is applied to reallocate the data based on the determined cluster number. The SOM network provides advantages over classical pattern-recognition techniques because it utilizes the parallel architecture of a neural network and provides a graphical organization of pattern relationships [6]. Furthermore, Kruskal-Wallis (KW) test is applied to screen the non-significant attributes among the VMI groups according to the user define distributed lag periods.

3.3 Neural Tree Construction

This study uses neural tree structure with back-propagation (BP) learning to construct the VMI forecast model. Neural tree could solve the disadvantage of understandability for neural network. In particular, the connections between input layer and hidden layer in neural tree can be reduced, in which could identify the responsibility and significance of each input factor and its relative descended factors easily. Each node in the hidden layer represents the attributes that influence the VMI and the descended nodes of each attribute in the input layer is the significant former period of the attribute.

BP is a supervised algorithm for learning appropriate weights of a multilayer feed-forward network [10]. The activation function for a non-input unit is

$$a_i = f\left(\sum_{\substack{j \in \text{previous} \\ \text{layer}}} w_{ij} a_j + b_i \right) \tag{1}$$

$$f(x) = \frac{1}{1+e^{-x}} + c \tag{2}$$

where c is a constant, e.g., c = 0 or c = -0.5. The BP algorithm involves a forward pass and a backward pass. In the forward pass, the output of unit i for pattern p is

$$a_{pi} = \frac{1}{1+e^{-net_{pi}}} \tag{3}$$

In particular, the net input to node i for pattern p is

$$net_{pi} = \sum_{\substack{j \in \text{previous} \\ \text{layer}}} w_{ij} a_{pj} + b_i \tag{4}$$

where a_{pj} is the activation value of unit j for pattern p, and w_{ij} is the weight from unit j to i, and b_i is a basis associated with unit i. In the backward pass, BP algorithm revises the weight based on gradient descent or the least square error function as follows.

$$E = \sum_{p} E_p \tag{5}$$

$$E_p = \frac{1}{2}\left\| t_p - o_p \right\|_2^2 \tag{6}$$

where t_p is the target output for the pth pattern, and o_p is the actual output for the pth pattern. The generalized delta rule is based on clever use of the chain rule for differentiation, $\Delta_p w_{ij} = \eta \delta_{pi} a_{pj}$ ($\Delta_p w_{ij}$ is the change of w_{ij} when pattern p is presented) where η is the learning rate and

$$\delta_{pi} = (t_{pi} - o_{pi}) f'(net_{pi}) \quad \text{if unit } i \text{ is an output unit} \tag{7}$$

$$\delta_{pi} = f'(net_{pi}) \sum_{\substack{k \in next \\ layer}} \delta_{pk} w_{ki} \quad \text{if unit } i \text{ is a hidden unit} \tag{8}$$

However, the larger of the learning rate, the larger of the changes in the weights for BP network training. A larger learning rate may lead to oscillation of weights. One way to increase the learning rate without leading to oscillation is to modify the back propagation learning rule according to the following equation:

$$\Delta w_{ij}(n+1) = \eta \delta_{pi} a_{pj} + \alpha \Delta w_{ij}(n) \tag{9}$$

where n is the number of iterations and α is a constant (momentum coefficient) that determines the effect of past weight changes on the current direction of movement in weight space.

Before modeling neural tree, the Taguchi parameter design is used to decide the best combination parameter setting including learning rate (η), momentum (α), and maximum training iterations. An $L_9(3^4)$ orthogonal array is used to deploy the experiment of parameter design. The inner orthogonal arrays represent the control factors and outer orthogonal arrays represent the noise factors. The root mean of sum of square error (RMSE) can be used for the performance of the neural tree model by the different parameter settings. To consider the result for training RMSE and testing RMSE, the signal to noise (S/N) ratio under smaller-the-better (STB) is used and estimated under different combinations of parameters. The design with biggest S/N value is selected for neural tree construction. Furthermore, the neural tree can be pruned according to the weights on each branch which are represented the magnitude and significance of each period from the attributes. The importance of each factors are ranked by the Pareto chart and thus the selected factors are used to re-build model for VMI forecast.

$$RMSE(y)=\sqrt{\frac{\sum_{p=1}^{M}\sum_{i=1}^{N}(o_{pi}^{2}-t_{pi}^{2})}{M\cdot N}} \tag{10}$$

$$SN_{STB}=-10\cdot\log_{10}(MSD)=-10\cdot\log_{10}(\frac{1}{n}\sum_{i=1}^{n}y_i^2) \tag{11}$$

4 Empirical Study

4.1 Background and Problem Definition

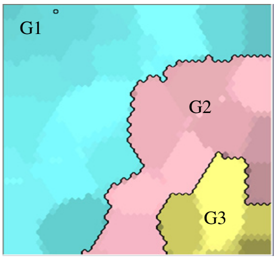
In order to validate the proposed MI framework, an empirical study was conduct from a leading semiconductor company in Taiwan. In practice, the customers sometimes queue the finished goods in fab without asking for the extra service in IC assembly and IC final test, these finished lots become VMI. In addition, if the extra services are done but the lots receive immediately, the lots also become VMI. Indeed, the VMI is the majority of Backend WIP. In order to control VMI and provide a basis of capacity allocation for reduction of inventory cost, the VMI needs to be forecast and to identify the important attributes for decision making.

After discussion with the domain experts, five attributes including fab utilization, customer hold rate, SEMI B/B ratio, and Company A's B/B ratio, pull-in rate are selected. In particular, SEMI B/B rate reflects next season's customers' behavior and Company A's B/B rate reflects next month's customers' behavior. The delay of these two attributes will be considered in model construction step. In addition, the lags period is determined to be six because they are interesting that how the attributes in former half year affect recent VMI. The definitions and observed period of each attribute are listed in Table 1.

Table 1 Definition and explanation of attributes

Attribute	Definition	Period
Utilization	(the real production time of wafer) ÷ (the effective production time of wafer) × 100%	t-1 ~ t-6
Customer Hold Rate	(quantity of customers hold the material) ÷ (customers allocated support demand (CASD))	t-1 ~ t-6
SEMI B/B Rate	(fab booking equipments) ÷ (facility vendor provide to semiconductor fab)	t-3 ~ t-8
Company A's B/B Rate	(quantity of customers booking wafers) ÷ (quantity of fab provide wafers to customers)	t-2 ~ t-7
Pull-in Rate	(difference between output and customers allocated support demand (CASD)) ÷ (CASD)	t-1 ~ t-6
•VMI%	(quantity of VMI) ÷ (CASD)•	-

P.S. For Company A, CASD represents the initial quantity of customer order which announced by the sales department.



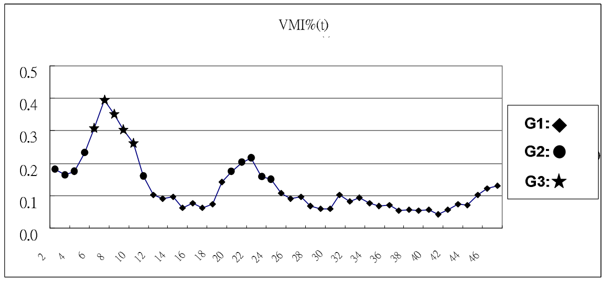


Fig. 2 SOM result

Fig. 3 VMI trend chart with three groups

4.2 Data Preparation

There are total 58 transactions collected by monthly. All transactions are transformed into distributed lags format with six months, and then the available number of transactions for analysis becomes 52. According to the time period, the first 46 transactions (90%) are used for model construction and the remaining 6 transactions (6%) are used to evaluate the model validity. Next, VMI with similar performance is grouped by two stage clustering method. In first stage, the Ward's method is applied to determine that the number of cluster is three. In second stage, the cluster number is set as three and the transactions are re-grouped by SOM neural network as shown in Fig. 2. These three groups perform with different level of VMI%. As shown in Fig. 3, VMI% tends to be higher (G3) in the initial period because of everything in this company is unstable and Company A doesn't have deep partnership with the customers. In the latter period, Company A becomes steady, the VMI level tend to be lower (G1). In G2, it could be explained as the transition stage or the change of operation strategy.

Furthermore, K-W test is used to test the significance of each factor among different VMI% levels. The non-significant one is eliminated to decrease the noise and simplify the model. Based on the significance level is 0.05, the utilization in period (*t-5*), (*t-6*) and SEMI B/B rate in period (*t-7*), (*t-8*) are non-significant.

4.3 Neural Tree Construction

BP neural tree is adopted to construct the VMI% forecast model based on the remaining 26 factors. Before training the neural tree, the input data value is normalized into interval 0.15~0.85. A $L_9(3^4)$ Taguchi orthogonal array is shown as in Table 2. The fifth parameter design with the highest S/N value is selected first. The 12 factors which have 80% importance are selected to re-build the neural tree model as shown in Fig. 4. The Taguchi parameter design of η, α, and maximum iteration is applied again. Fig. 5 shows that the seventh parameters design has the highest S/N value. It means this model performs best with lowest error and variability, and the seventh model is selected to forecast the VMI%.

Table 2 $L_9(3^4)$ Orthogonal array

No.	Inner Orthogonal Array			Outer Orthogonal Array mode of test data / number of test data		S/N
				Random/ 10	End/ 15	
	factors			data		
No.			iteration	Y1	Y2	S/N
1	0.2	0.9	5000	0.074716	0.010757	25.45287
2	0.2	0.925	10000	0.074704	0.010756	25.45430
3	0.2	0.95	15000	0.074711	0.010754	25.45356
4	0.225	0.9	10000	0.074709	0.010756	25.45377
5	0.225	0.925	15000	0.074699	0.010755	25•45496
6	0.225	0.95	5000	0.074719	0.010756	25.45261
7	0.25	0.9	15000	0.074709	0.010755	25.45379
8	0.25	0.925	5000	0.074709	0.010756	25.45378
9	0.25	0.95	10000	0.074720	0.010754	25.45251

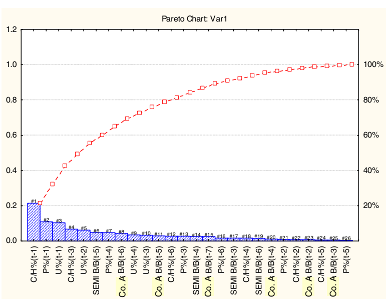


Fig. 4 Pareto chart of input contribution

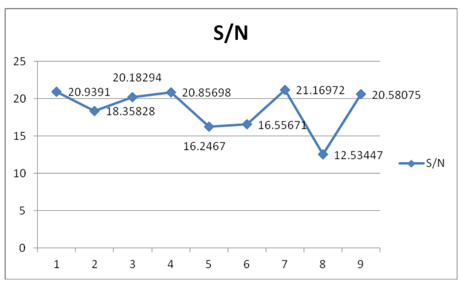


Fig. 5 S/N result for different parameter setting

4.4 Result Evaluation and Discussion

The model can provide information for the engineer that VMI% level is going up or down. The remaining 6 testing data also can be classified correctly. If the VMI% rises, the model can remind the engineers to re-allocate the capacity to avoid extra waste. According to the result of neural tree model as listed in Table 3, three attributes including customer hold, utilization and pull-in are mainly influenced VMI%. This means when customers hold the material and stop process in fab, the customers change the lots delivery quantity, or Company A's utilization shifts, the VMI% raises.

Table 3 The importance of attributes

Attribute	Items	Percent	Importance of attribute
C/H%	C/H%(t-1)	21.3292	30.7166
	C/H%(t-3)	6.6309	
	C/H%(t-6)	2.7565	
U%	U%(t-1)	10.3287	23.0026
	U%(t-2)	6.0756	
	U%(t-3)	3.2338	
	U%(t-4)	3.3645	
P%	P%(t-1)	10.9721	15.7113
	P%(t-4)	4.7392	
Co. A B/B	Co. A B/B(t-5)	2.8034	7.1002
	Co. A B/B(t-6)	4.2968	
SEMI B/B	SEMI B/B(t-5)	4.8018	4.8018

5 Conclusion

The demands variation of various customer orders affect the total loading of the fab directly, and thus create variability on inputted wafer start, which will affect the WIP bubble and cycle time as well as the capacity utilization and profitability. With knowing the information of customer behavior, the decision maker can effectively control the quantity of productions and optimize production allocation decisions in light of involved uncertainty of customer behaviors and thus be able to reduce the waste of material and cost of inventory. The extracted attributes are useful to track the root cause of variation of VMI and forecast the VMI% next month. The empirical study results showed the practical viability of this approach and thus manage VMI effectively to adjust the capacity allocation. Further research can be done to develop a real time update model by every unit period and discuss the interaction between each factor.

Acknowledgments. This research is supported by National Science Council, Taiwan (NSC99-2221-E-007-047-MY3; NSC 100-2410-H-155-048).

References

[1] Chen, L., Chien, C.-F.: Manufacturing intelligence for class prediction and rule generation to support human capital decisions for high-tech industries. Flex Serv. Manuf. J. 23(3), 263–289 (2011)

[2] Chien, C.-F., Wang, W., Cheng, J.: Data mining for yield enhancement in semiconductor manufacturing and an empirical study. Expert Syst. Appl. 33(1), 192–198 (2007)

[3] Chien, C., Chen, Y., Peng, J.: Manufacturing intelligence for semiconductor demand forecast based on technology diffusion and production life cycle. Int. J. Prod. Econ. 128, 496–509 (2010)

[4] Chien, C.-F., Hsu, C., Hsiao, C.: Manufacturing intelligence to forecast and reduce semiconductor cycle time. J. Intell. Manuf. (2011), doi:10.1007/s10845-011-0572-y

[5] Chien, C.-F., Wu, C., Chiang, Y.: Coordinated capacity migration and expansion planning for semiconductor manufacturing under demand uncertainties. Int. J. Prod. Econ. 135, 860–869 (2012)

[6] Dayhoff, J.E.: Neural networks architectures: an introduction. Van Nostrand Reinhold, New York (1990)

[7] Kohonen, T.: The self-organizing map. Proc. IEEE 78(9), 1464–1480 (1990)

[8] Kuk, G.: Effectiveness of vendor-managed inventory in the electronics industry: determinants and outcomes. Inf. Manage. 41, 645–654 (2004)

[9] Kuo, C., Chien, C.-F., Chen, C.: Manufacturing intelligence to exploit the value of production and tool data to reduce cycle time. IEEE Trans. Autom. Sci. Eng. 8(1), 103–111 (2010)

[10] Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning internal representation by error propagation. Parallel Distrib. Process. 1, 318–362 (1985)

[11] Ward, J.H.: Hierachical Grouping to Optimize an Objective Function. J. Am Stat. Assoc. 58, 236–244 (1963)

MCDM Method for the Behavior of Innovative New Product Development

Mei-Chen Lo

Abstract. In the highly competitive environment, new product development (NPD), technology, equipment and materials have applications progress rapidly by promoting the concept of innovation. The manager not only intends to achieve the highest customer satisfaction, product value and product continuity, but also requires in low cost, high quality, mass production and time to market in customization. In this paper, we propose an effective approach for promoting innovative new product development by multiple-criteria decision making (MCDM) method. The proposed hierarchy structure is applied to provide the priority for improving the high-tech product development. The results of this study provide a guidance to continuously improve, track and meet the quality assurance of production, research and development consequently; the customer needs can be responded.

Keywords: New Product Development (NPD), Multiple Criteria Decision Making (MCDM), Analytic Network Process (ANP).

1 Introduction

In the highly competitive environment, new product development (NPD), technology, equipment and materials have progressed rapidly. It is a gradual trend that competitors continuously innovate on their product and product life-cycle becomes shorter [8]. Innovative technology is critical to nation's economic development and prosperity. Econometric studies identify technological advancement as perhaps the major contributor to modern economic growth, attributing more growth to technology. Technological innovation also has facilitated a rise in the standard of living for millions of consumers.

Innovation comprises combinations of knowledge that result in new products, processes, input and output markets, or organizations that include not only technical innovations, but also organizational and managerial innovations, new

Mei-Chen Lo
Department of Business Management, National United University, No.1, Lienda Rd., Miaoli 36003, Taiwan
e-mail: meichen_lo@nuu.edu.tw

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 61–68.
springerlink.com

markets, new sources of supply, financial innovations, and new combinations. Clark and Guy [3] mentioned that innovation is a critical factor in enhancing a firm's competitiveness, which generally is understood to refer to the ability of a firm to increase in size, market share, and profitability at the firm level [3]. The literature distinguishes different types of innovation: incremental, radical, technological, process, product, organizational, operational, managerial, social, or institutional (e.g., [5,12,18]). Innovation processes therefore can be viewed as a sequence of exploration (in which existing products and processes are adapted incrementally or radically through the search for and application of new assets) and exploitation (in which the variety of products and processes decreases while their efficiency increases) [18].

2 The Framework of NPD

Traditional, companies strengthen their competence to innovate by developing the capabilities of employees within the organization [6]. However, this approach began to change in the 1980s [9]. A company's competitiveness increasingly depends on its capabilities beyond its internal boundaries. Promoting the firm's innovative activities integrate both resources from internal and external inputs in a managed way enables companies to realize radically new product innovation.

Although, a high-tech company's competitiveness has increasingly depends on its capabilities beyond its internal boundaries. Promoting innovation, the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation [2], has emerged as a model where firms commercialize both external and internal ideas/technologies and use both external and internal resources. In terms of R&D, the leverage of external R&D ideas, technologies as well as resources to achieve a firm's goals of innovation, has already become an important concept for achieving the activities of promoting innovation. Opportunities in fields like micro-electronics, communications, information systems, bio-technologies, and energy technologies will provide products that now exist only in the imagination.

2.1 Technological Innovation

The potential for technology to enhance further the consumers' quality of life is thus enormous. Realization of this potential of technological change lies to a significant extent in the capacity of our economic and legal systems to encourage such change and utilize its benefits fully. Society's concern with consumer and environmental protection has led to a variety of regulations affecting products sold in consumer markets.

As Fig. 1, a strategy that companies may adopt is to delight customers and to exceed their satisfaction. Through the use of products with attractive quality, service, cost, user friendly design, customers may be retained and thereby market share can be captured and sustained. Essentially, customer requirements and expectations can be satisfied and exceeded with such products.

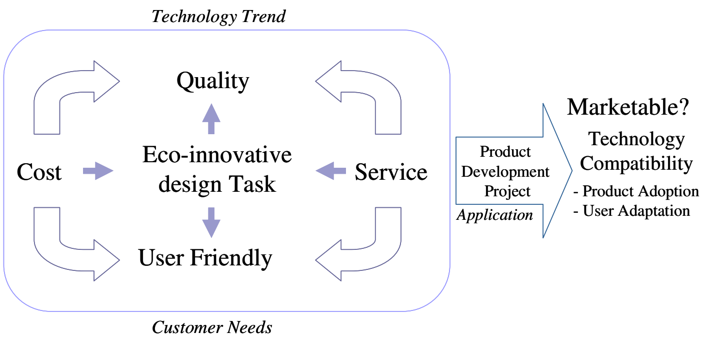


Fig. 1 Interrelations on Promoting NPD (Internal concerns)

2.2 New Product Development Project

NPD process may be accepted as a dynamic process in which each decision-making on each stage must be evaluated, selected, and prioritized. All the stages of the process are affected by uncertain, changing information and dynamic opportunities [1]. Hence, much attention should be paid to attractive quality creation when managing a product development project. Another important implication we concerns the timely delivery of innovative products. According to the model, attractive attributes over time become one-dimensional attributes and then further on become basic attributes. In other words, those products that were perceived as innovative and attractive are not considered to be innovative at the present time, and consequently customer satisfaction may not be achieved and exceeded. Timely development and introduction of products with innovative features are, thus, important. Swink [14] believes the quality and up to 80% of NPD costs that has decided in the product design stage.

2.3 Environmental Concerns

The development of technology plays a crucial role in modern economic growth, but it is also the key factor of environmental crisis. Manufacturers usually emphasize the novelty and economic usefulness of their innovative products but neglect its environmental impact [4]. As it becomes the competitive issues to companies, those issues and trends are already affecting many businesses. Environmental performance is increasingly a determinant of a company's reputation, among employees, customers, and stakeholders alike. A poor environmental reputation can harm recruitment, retention, and morale, damage sales, and threaten a company's "license to operate." Good performance can have the opposite effect. And the visibility of both is increasing as legislation, stakeholder, and peer pressures require more and more external reporting on the topic and media coverage becomes ever more global and instantaneous.

2.4 Process Model

The initial product idea generation and identification stages have received considerable attention. Product ideas may arise from various sources, e.g. technology, market needs, competitors and user solutions [7]. For an initial product idea, potential customers should be identified before collecting their requirements and conducting further analysis. As a consequence, segments probably exist among customers. It is useful to identify possible customer segments for maximizing market share. After knowing who the potential customers are, the project team then proceeds to collect their specific needs and requirements based on the initial product idea. As Fig. 2, the process model of innovative products, five development stages were found to impact their performance significantly and positively, namely (i) idea generation, (ii) concept and screening, (iii) internal activities and analysis, (iv) external environmental analysis (v) technical development and strategy deployment.

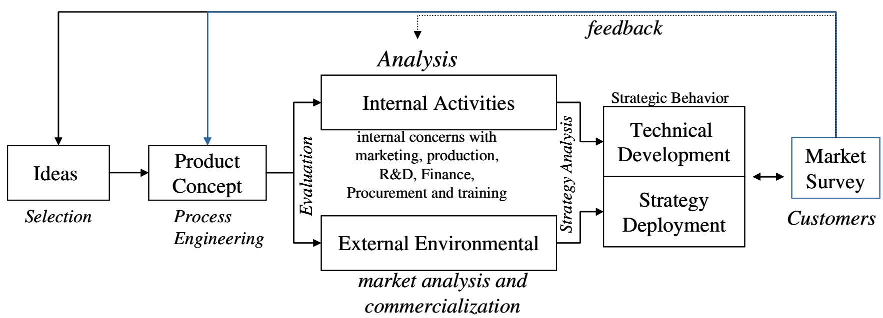


Fig. 2 Process Model of Innovative Product Development

According to above issues of existing innovative NPD, we explored the development characteristic of the high-tech product and the structure of NPD (as Fig. 3) in order to find the optimization of problem-solving.

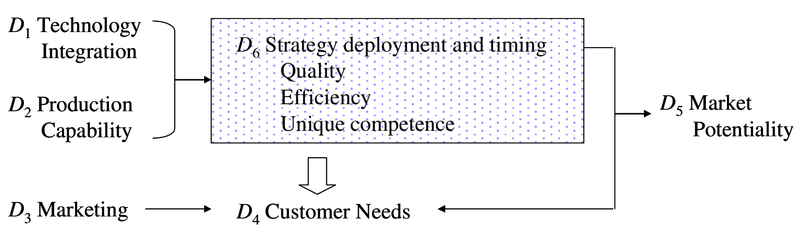


Fig. 3 Assumptions of Promoting Innovative product

3 Methodologies of MCDM

This section is aimed to understand the evaluation indexes of promoting innovative product development related to high-tech sectors. It is collected, selected, analyzed, simulated and tested by the literature with expert questionnaires to find the usefulness and the promotion spots as a basis. We used the existing high-tech product nature [10, 11] and the characteristics of the high-tech industry to explore the related issues of promoting activities. Lo & Tzeng [10] indicate that Multiple Criteria Decision-Making (MCDM) is a methodology that is able to consider multiple criteria at the same time and also helps the decision- maker to estimate the best case by sorting cases according to the characteristics or criteria [15, 16, 17] of each from limited available cases.

In this study, according to the literature review and expert experiences, a evaluation system including six dimensions and 24 criteria that will exert an influence on promoting NPD is established, as given in Table 1. A survey was conducted via questionnaires distributed to several groups comprised of the experts' background from marketing, research and development (R&D), management, engineering, finance, procurement, production, planning and facilities. Their ratings for each criterion's relationship to sustainable development using a five-point scale ranging from 0 (no effect) to 4 (extremely influential) were collected.

Table 1 The Evaluation Criteria of Promoting New Product Development

Dimensions	Criteria	Weight	Rank
D_1 Technology Integration	C_1 Integration capability	0.041	18
	C_2 Compatibility	0.043	3
	C_3 Technology diffusion speed	0.042	13
	C_4 Core technology	0.041	16
D_2 Production Capability	C_5 Capacity allocation (configuration)	0.040	23
	C_6 Production integration and structure	0.042	10
	C_7 Efficiency and yield	0.043	4
	C_8 Cost control	0.042	14
D_3 Marketing	C_9 Complementary/alternative	0.041	20
	C_{10} Time-to-market (market timing)	0.042	12
	C_{11} Marketing practices	0.043	2
	C_{12} Opponents simulation - market response	0.041	17
D_4 Customer Needs	C_{13} Installation base	0.042	8
	C_{14} Customer response	0.040	22
	C_{15} Customization	0.043	5
	C_{16} Training-knowledge platform	0.041	15
D_5 Market Potentiality	C_{17} Competence position	0.040	24
	C_{18} Related supporting industries	0.042	11
	C_{19} Infrastructure environment (built)	0.043	6
	C_{20} Product/industry life cycle	0.043	7
D_6 Strategy deployment and timing	C_{21} Product development project	0.042	9
	C_{22} Product mix and match	0.043	1
	C_{23} Technology acquisition (Obtaining)	0.040	21
	C_{24} Alliances and partnerships	0.041	19

This set of criteria provides this study with an overall evaluation system that facilitates further prioritization by the concept of ANP [13]. According to Table 1, the critical factors for promoting NPD which is including “Consistency for the product mix and match (C_{22})”, “Marketing practices (C_{11})”, “Compatibility (C_2)”, “Efficiency and yield (C_7)”, “Customization (C_{15})”,and “Infrastructure environment (C_{19})” according to the priority sequence of these factors from ANP.

4 Empirical Analysis

The MCDM model refers to making decisions in the presence of multiple, and often simultaneously faced/managed multiple criteria/objectives with critique criteria in real world. The problems of MCDM can be broadly classified into two categories including multiple objective decision making (MODM) for plan and multiple attribute decision making (MADM) for evaluation and selection. We use the related literatures and experiment to summarized and discussed. MCDM approaches subject to calculate the weights and ranking that is used to solve the decision problem for the priority of promoting new product development. We applied the influence matrix to compromise the trade-off between innovative product promoting and development concerns. The main survey objects engage for the NPD in six phases (such as, technology integration, production capability, marketing, customer needs, market potentiality, strategy deployment and timing), within related areas of experts or scholars who are senior in their domain knowledge and experiences on project management in high-tech industry.

Table 2 Evaluation Index of Total Influence by Dimension

	D_1	D_2	D_3	D_4	D_5	D_6	Influence (D)	***d+r***	***d-r***	Weight	Rank
D_1	0.035	0.043	0.040	0.040	0.040	0.044	0.157	0.314	0.000	0.170	2
D_2	0.042	0.032	0.038	0.042	0.041	0.040	0.154	0.313	-0.005	0.166	3
D_3	0.040	0.040	0.029	0.038	0.041	0.042	0.147	0.293	0.001	0.159	6
D_4	0.041	0.044	0.040	0.029	0.039	0.040	0.153	0.303	0.004	0.166	4
D_5	0.038	0.038	0.040	0.037	0.029	0.040	0.152	0.313	-0.009	0.164	5
D_6	0.038	0.042	0.041	0.042	0.042	0.033	0.162	0.329	-0.004	0.175	1
Influenced (R)	0.157	0.159	0.146	0.149	0.161	0.167	--	--	--	--	--

According to Table 2, the prominence $(d_i + r_i)$, “Strategy deployment and timing (D_6)” is the highest impact of the strength of relation that means the most important influencing factors; in addition, “ Marketing (D_3)” are all the factors that affect the least degree of other factors. According to the relation $(d_i - r_i)$, we also can find “Market Potentiality (D_5)” is the highest degree of impact relationship that affects other factors directly. These dimensions also have the interact characteristics. Opposite, “Customer Needs (D_4)” is the most vulnerable to impact that compare with other dimensions.

Base on these empirical results, we construct the network relationship map of each dimension as shown in Fig. 4. It illustrate the critical problems in promoting NPD including “Production Capability (D_2)”, “Market Potentiality (D_5)” and “Strategy deployment and timing (D_6) which are easily impacted by other factors. Therefore, this model provides the direction of problem solving from “Productivity (D_1)”, “Marketing (D_3)” and “Customer Needs (D_4)”. Some strategies are suggested according to experts‘verification, such as technology leading devise, product diversification, deploy market share, strategy alliance and efficiency improvement, etc.

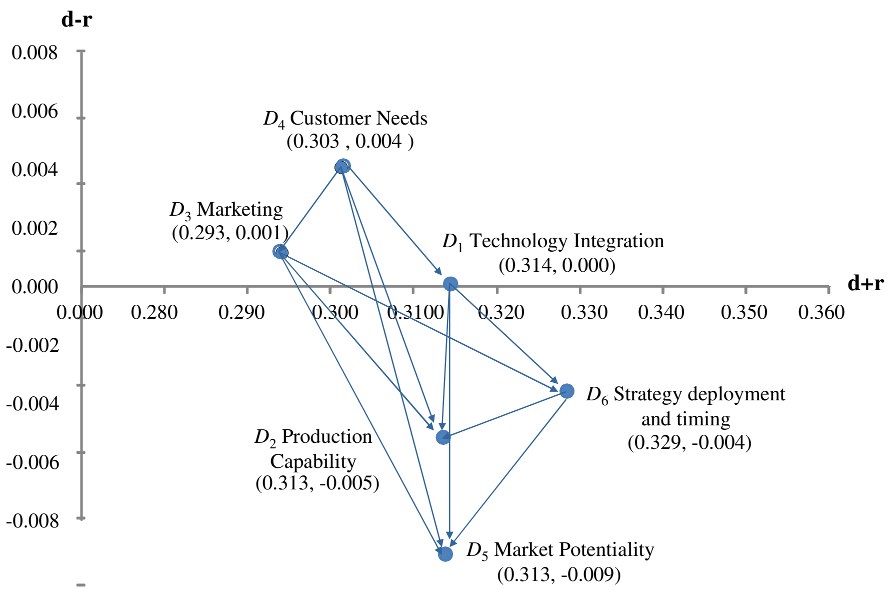


Fig. 4 The NRM of each dimension

5 Conclusion

The research described in this paper demonstrates the development of innovative high-tech product is within the realms of possibility. This study is including different structure to be overcome. For this purpose, an idea generation method has been proposed in which the influence matrix is adopted based on many fields knowledge. Here, we employed with an innovative method with group expert’s participation and multi-criteria decision making, the concept of ANP handling with many potential uncertainty factors which is able to evaluate and provide the most suitable guide for promoting the innovative product. Finally, this study provide a guidance to project team a continuous improvement, track for achieving the quality assurance of production, consequently, the customer needs can be satisfied.

References

1. Büyüközkan, G.G., Feyzolu, O.: A fuzzy-logic-based decision-making approach for new product development. International Journal of Production Economics 90(1), 27–45 (2004)
2. Chesbrough, H.: Open innovation. Harvard Business School Press, Harvard (2003)
3. Clark, J., Guy, K.: Innovation and Competitiveness: a Review. Technology Analysis & Strategic Management 10(3), 363–395 (1998)
4. DeSimone, L.D., Popoff, F.: Eco-efficiency: The Business Link to Sustainable Development. Massachusetts Institute of Technology (1997)
5. Hammer, M.: Deep Change. How Operational Innovation Can Trans-form Your Company. Harvard Business Review 82(4), 84–93 (2004)
6. Hargadon, A., Sutton, R.I.: Building an Innovation Factory. Harvard Business Review 78(3), 157–166 (2000)
7. Hauser, J.R., Urban, G.L., Weinberg, B.D.: How Consumers Allocate Their Time When Searching for Information. Journal of Marketing Research 30(4), 452–466 (1993)
8. Levitt, T.: Exploit the Product Life Cycle. Harvard Business Review 43(6), 81–94 (1965)
9. Lichtenthaler, U., Lichtenthaler, E.: A Capability-Based Framework for Open Innovation: Complementing Absorptive Capacity. Journal of Management Studies 46(8) (December 2009), doi:10.1111/j.1467-6486.2009.00854.x
10. Lo, M.C., Tzeng, G.H.: Fuzzy MCDM Application for Strategy Evaluation. In: The 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), Taipei, Taiwan (June 27-30) (2011)
11. Lo, M.C., Michnik, J., Cheng, A.L.P.: Technology Assessment as A Guidance to Business Management of New Technologies. In: The 2007 IEEE In-ternational Conference on Industrial Engineering and Engineering Management (IEEM 2007), Singapore (December 2-5) (2007)
12. Nadler, D.A., Tushman, M.L.: The Organization of the Future: Strate-gic Imperatives and Core Competencies for the 21st Century. Organizational Dynamics 28(1), 45–60 (1999)
13. Saaty, T.L.: Decision making with dependence and feedback: The Analytic Network Process. RWS Publications, Pittsburgh (1996)
14. Swink, M.: Threats to New Product Manufacturability and the Effects of Development Team Integration Processes. Journal of Operations Management 17(6), 691–709 (1999)
15. Tzeng, G.H., Huang, J.J.: Multiple Attribute Decision Making: Methods and Applications. Taylor & Francis (2011)
16. Tseng, M.L.: A causal and effect decision-making model of service quality expectation using grey-fuzzy DEMATEL approach. Expert Systems with Applications 36(4), 7738–7748 (2009a)
17. Tseng, M.L.: Using extension of DEMATEL to integrate hotel service quality perceptions into a cause-effect model in uncertainty. Expert Systems with Applications 36(5), 9015–9023 (2009b)
18. van Kleef, J.A.G., Roome, N.J.: Developing Capabilities and Compe-tence for Sustainable Business Management As Innovation: a Research Agenda. Journal of Cleaner Production 15(1), 38–51 (2007)

Medical Decision Support System for Assessment of Dermatoglyphic Indices and Diagnosis of Down's Syndrome

Hubert Wojtowicz and Wieslaw Wajs

Abstract. The paper describes design of an intelligent information system for assessment of dermatoglyphic indices of Down's syndrome in infants. The system supports medical diagnosis by automatic processing of dermatoglyphic prints and detecting features indicating presence of genetic disorders. Application of image processing and pattern recognition algorithms in pattern classification of fingerprints and prints of the hallucal area of the sole is described. Application of an algorithm based on multi-scale pyramid decomposition of an image is proposed for ridge orientation calculation. A method of singular points detection and calculation of the ATD angle of the palm print is presented. Currently achieved results in dermatoglyphic prints enhancement, classification and analysis are discussed.

1 Introduction

Certain genetic disorders in infants manifest themselves through the unusual combinations of dermatoglyphic patterns. Early detection of these patterns can help to establish the diagnosis and start a treatment as first demonstrated in the case of Down's syndrome [6]. Dermatoglyphics are the dermal ridge configurations on the digits, palms, and soles. Dermal ridges begin to develop at about the 13th week of prenatal life as the fetal mounds on the digit tips and the interdigital, thenar, hypothenar, and corresponding areas of the foot begin to regress. The pattern formation is complete by the 19th week. Factors disturbing normal development of fetus may also influence formation of dermal ridges structures. There are a number of

Hubert Wojtowicz
University of Rzeszow, Faculty of Mathematics and Nature, Institute of Computer Science
e-mail: hubert.wojtowicz@gmail.com

Wieslaw Wajs
AGH University of Science and Technology, Faculty of Electrical Engineering,
Institute of Automatics
e-mail: wwa@agh.ia.edu.pl

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 69–78.
springerlink.com

dermatoglyphic patterns that occur more frequently in certain syndromes than in the general population. Detection of the characteristic differences in frequency of dermal configurations between affected and normal children requires a simultaneous analysis of dermatoglyphic prints of fingers, palms and soles. For the detection of Down's syndrome a diagnostic index was developed called dermatoglyphic nomogram [7]. Successful detection of Down's syndrome in infants using this index relies on a correct recognition of dermatoglyphic patterns by an anthropologist.

2 The Aim of the Work

The scientific aim of the project is development of new method used for problem solving of pattern recognition and understanding of the classification process of the Down's syndrome symptoms. The new method is based on the combination of text knowledge found in the scientific literature describing Down's syndrome with knowledge obtained by the analysis of dermatoglyphic indices characteristic for Down's syndrome using digital pattern recognition techniques. A scientific goal is design of a classifier realizing statistical comparison of knowledge included in the text description with knowledge provided by the mathematical methods of the analysis of indices present in dermatoglyphic patterns. This approach is similar to the approach used by an anthropologist, which the system realizes by juxtaposing knowledge described in the form of natural language sentences and information provided by the appropriate digital equipment and on the basis of this juxtaposition performs arbitral classification of the investigated pattern. The conception of the project assumes development of a prototype of the telemedical system, in which data in the form of dermatoglyphic images is transferred through the telecommunication networks from the distant hospital database centers and processed. The aim of the telemedical system is screening analysis of incoming data. Another goal is a support of the diagnosis process conducted by medical personnel in cases of ambiguous classification of the telemedical system.

The aim of the research conducted is creation, based on gathered data and domain knowledge described in medical literature, of an automatic system supporting the diagnosis process and detecting infants' genetic disorders as follows:

1. The system recognizes characteristic combinations of particular patterns of soles, palms and fingers and on that basis infers the occurrence of genetic disorders. It is expected, that the application of this system improves treatment's effectiveness, i.e. the number of complications caused by the treatment in the later years of infants' life is going to be lower.
2. The system supports doctor's work by the analysis of large amounts of patients' data and decreases the probability of a mistake in strenuous biometric analysis such as counting number of ridges, determining ridges width or calculating the ATD angle.

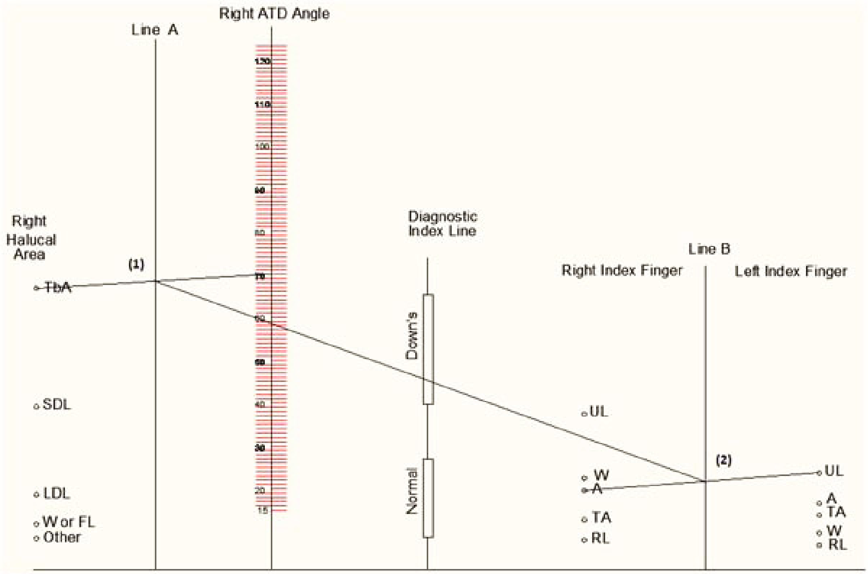


Fig. 1 Example of the dermatoglyphic nomogram. Abbreviations: TbA - tibial arch; SDL - small distal loop; LDL - large distal loop; FL - fibular loop; W - whorl; UL - ulnar loop; A - arch; TA - tented arch; RL - radial loop.

3 Description of Dermatoglyphic Nomogram

Dermatoglyphic nomogram is based upon four pattern areas and uses four dermatoglyphic traits chosen for their high discriminant value. These traits include: pattern types of the right and left index fingers, pattern type of the hallucal area of the right sole and the ATD angle of the right hand.

Figure 1 presents an example of the nomogram. In order to determine the diagnostic value of the index three lines must be constructed. The first of the lines connects particular determined type of pattern of the hallucal area of the sole with the selected value on the scale of the diagnostic line representing the determined value of the ATD angle of the right palm. In the case shown in the diagram it is the connection of the tibial arch pattern with the angular value equal to 70. The plotted line intersects with line A at the point marked (1). The second line connects the determined type of the pattern of the index fingerprint of the right hand with the pattern of the corresponding index fingerprint of the left hand. In this case it is the connection of the pattern type called plain arch (PA) with the pattern type called ulnar loop (UL). The plotted line intersects with line B at the point marked (2). The last third line connects the point of intersection labeled as (1) with the point of intersection labeled as (2) and crosses main diagnostic line. The point of intersection of the third line with the diagnostic line determines the diagnosis. There are three diagnostic cases, which correspond to the particular intervals of the diagnostic line.

The first case corresponds to the intersection of the lines in the interval denoted as ”Down’s” - an infant according to the determined value of the nomogram has Down’s syndrome. The second case corresponds to the intersection of the lines in the interval denoted as ”Normal” - an infant according to the determined value of the nomogram is healthy. The third case corresponds to the intersection of the lines in the interval between ”Down’s” and ”Normal” - it cannot be determined whether the infant has a genetic disorder or not.

Designations referring to the anatomy of the hand are used in describing palmar dermatoglyphics and in presenting methods of interpreting them. Terms of anatomical direction (proximal, distal, radial, ulnar) are employed in describing the locations of features and in indicating directions toward the respective palmar margins. Fingerprint pattern left loop is called ulnar loop when found on any of the fingers of the left hand and it is called radial loop when found on any of the fingers of the right hand. Pattern right loop in the left hand is called radial loop and in the right hand is called ulnar loop [2].

4 Classification Method of Fingerprint Patterns

Fingerprint classification is one of the tasks of dermatoglyphic analysis. Many classification methods were developed and described in the literature. Classification method used in dermatoglyphic analysis is called the Henry method. It classifies fingerprints into five distinct classes called: left loop (LL), right loop (RL), whorl (W), arch (A) or plain arch (PA) and tented arch (TA) (Fig. 2).

Classification scheme based on the Henry method is a difficult pattern recognition problem due to the existence of small interclass variability of patterns belonging to the different classes and large intraclass variability of patterns belonging to the same class. In the Henry classification scheme the whorl pattern has large intraclass variability. In this scheme term ”whorl” has a very broad sense and any pattern having two or more triradii would be assigned to the class of whorls. Small interclass variability results from the natural transitions between patterns and applies to the group of patterns, which include: reduced ulnar loops, plain arches, reduced tented arches, reduced whorls and reduced radial loops. In the paper we present a classification scheme based on the extraction of fingerprint ridge orientation maps from the enhanced images. Vectors constructed from the directional images are used for training of SVM multi-class algorithms. For the induction process of SVMs we use RBF and triangular type kernels.

5 Classification of the Hallucal Area of Sole Patterns

Classification of the patterns in the hallucal area of the soles is another task of dermatoglyphic analysis. The patterns in the hallucal area are classified into five distinct classes called: large distal loop (LDL), small distal loop (SDL), whorl (W), tibial arch (TA) and tibial loop (TL) (Fig. 3).

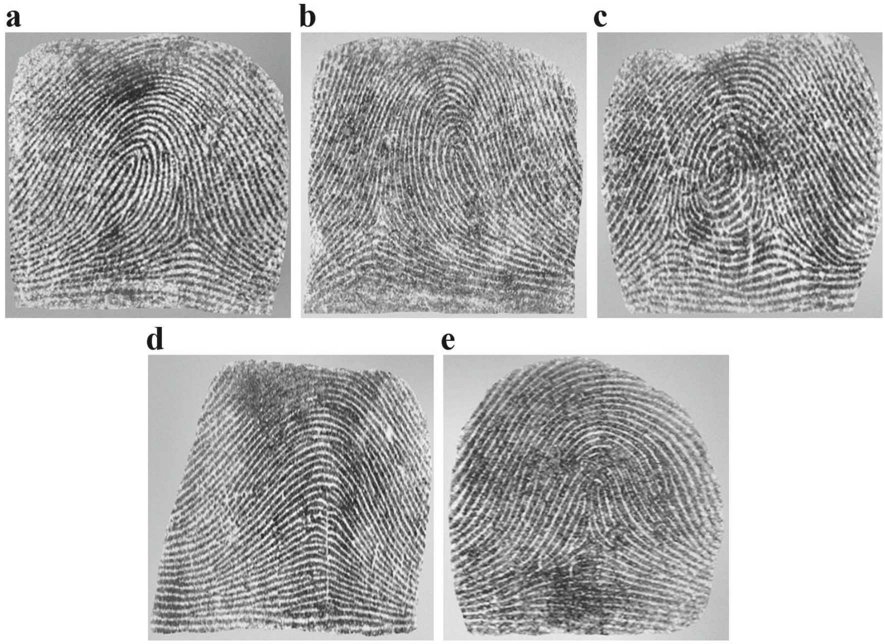


Fig. 2 Example fingerprints: (a) left loop; (b) right loop; (c) whorl; (d) plain arch; (e) tented arch

Classification of the patterns in the hallucal area of the soles is also a difficult pattern recognition problem due to the existence of small interclass variability of patterns belonging to the different classes and large intraclass variability of patterns belonging to the same class. The whorls of the hallucal area of the sole, similarly as the whorls of the fingerprints, are characterized by a very large variability. The main types of this class in the hallucal area are: a typical concentric or spiral whorl; a whorl presenting an S-shaped central design; seam, a whorl showing abrupt interruption of the concentric circuit; lateral pocket or two interlocked loops; a central pocket loop. Small interclass variability occurs in this dermatoglyphic area between the classes of small distal loop and tibial arch.

6 Method of the ATD Angle Calculation

The ATD angle of the right palm is calculated by finding locations of digital triradii labeled A and D and location of axial triradius labeled t. Four triradii labeled A, B, C and D are located at the base of the palm. An axial triradius, labeled t is most often located near the point where the palm is connected to the wrist. Two percent of normal infants and all of infants with Down syndrome have the axial triradius positioned near the center of the palm. Triradius in this elevated location is

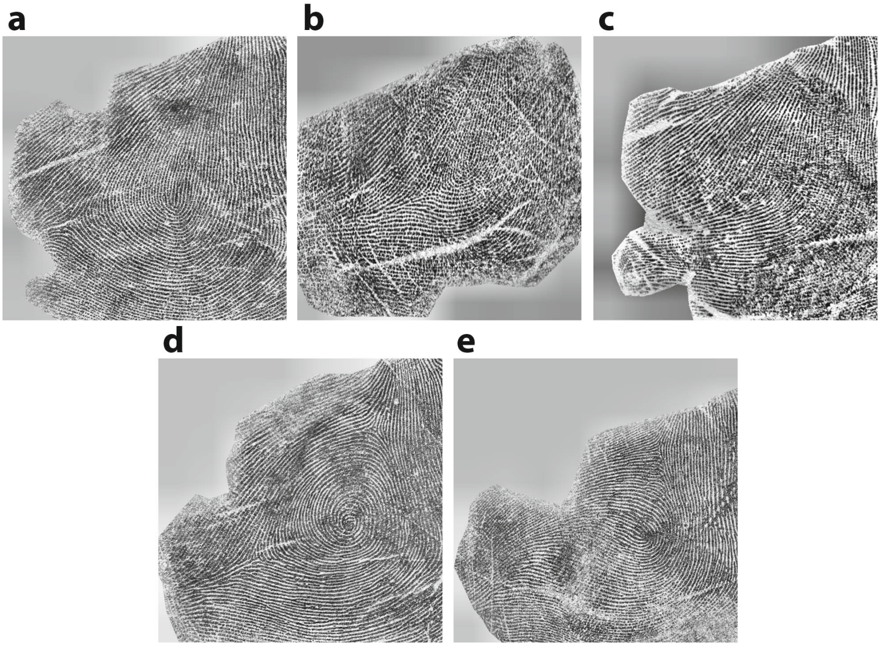


Fig. 3 Example patterns in the hallucal area of the soles: (a) large distal loop , (b) small distal loop, (c) tibial arch, (d) whorl, (e) tibial loop

termed t''. A triradius t' located approximately halfway in between these two extreme positions is found on twenty-one percent of the normal population. The mean value of the ATD angle for normal female infants is $\sim 39°$ and for normal male infants is $\sim 43°$. The mean value of the ATD angle for infants with Down's syndrome is $\sim 81°$. Locations of palmprint's singular points are calculated from the orientation map. The orientation map is computed using algorithm based on the principal component analysis and multi-scale pyramid decomposition of the image. Principal Component Analysis is applied to the image to calculate maximal likelihood estimations of the local ridge orientations and the multi-scale pyramid decomposition of the image helps to improve the accuracy of the calculated orientations. Algorithm calculating ridge orientations is robust to noise and allows for reliable estimation of local ridge orientations in the low quality areas of images. After the process of estimation of ridge orientations a Poincare index is calculated, which in turn allows extraction of singular points from the palmprint image. Basing on the locations of singular points the ATD angle of the palmprint is calculated. A two stage algorithm is applied in order to reliably estimate positions of singular points. In the first stage improved formula of Poincare index is used to calculate candidate locations of singular points [9]. In the second stage a coherence map of the palmprint image is calculated. The coherence map is calculated for each pixel in the image on the basis of its eigenvalues. Pixel eigenvalues are calculated from the confusion matrix

Fig. 4 Singular points of the palmprint located using two stage algorithm taking advantage of improved Poincare index and Gauss-Hermite moments

containing values calculated for each pixel (i,j) by applying the combination of two dimensional orthogonal Gaussian-Hermite moments to the segment of the image centered in pixel (i,j). The proposed approach combines information about singular points obtained from the improved Poincare index calculated from the orientation map and the information obtained from the coherence map calculated from the pixel map of the image. The two stage approach allows for reliable detection of only true singular points even in low quality areas of the palmprint image [9]. Example of the calculation of palmprint's singular points using the proposed algorithm is presented in Fig. 4.

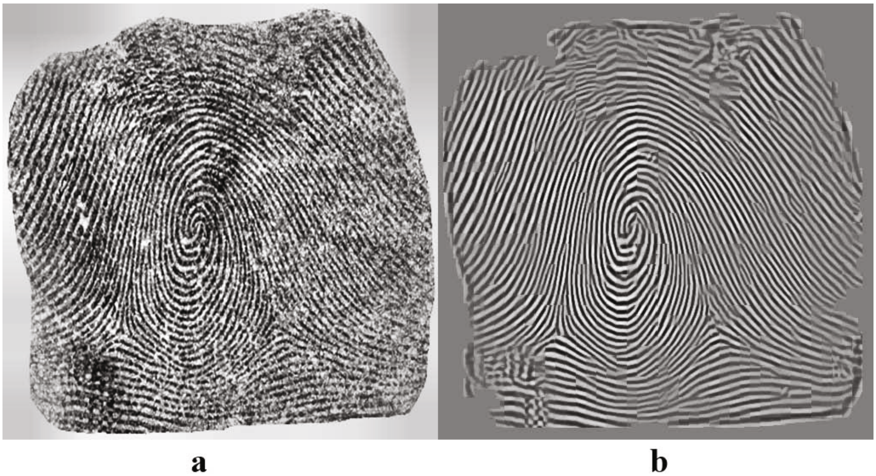


Fig. 5 Image of the fingerprint impression belonging to the whorl class preprocessed using contrast enhancement algorithm CLAHE (a), and then by filtration algorithm STFT (b)

7 Feature Extraction

Accurate extraction of features and classification of fingerprints depends on the quality of the images containing the impressions. Determination of the quality of the impressions partly depends on subjective criteria. It corresponds to the clarity of the dermal ridges structure. Impressions of good quality are characterized by high contrast and clearly discernible structure of ridges and valleys, poor quality impression are of low contrast and ill-discernible structure of ridges and valleys. The overall approach used for extraction of the dermatoglyphic features is divided into several stages. In the first stage removal of the background noise from the images is accomplished by application of the segmentation algorithm, which separates background from the parts of the image where ridges are present. The region mask representing the foreground area is used to remove background area from the image. The image is truncated according to the coordinates of the boundary points of the foreground area and then resized to the square frame. Contrast enhancement is performed in the next stage using locally adaptive contrast enhancement algorithm (Fig. 5a). In the last stage of pre-processing image quality enhancement is carried out using a contextual image filtration STFT (Short Time Fourier Transform) algorithm [1], which generates information about ridges flow directions, frequency of ridges and local image quality estimation (Fig. 5b). Ridge orientation maps are computed using algorithm based on Principal Component Analysis and multi-scale pyramid decomposition of the images [3]. The multi-scale method provides robustness to the noise present in the image. Examples of the estimation of the ridge local orientations in the distorted areas of the fingerprint impression from Fig. 5 are shown in Fig. 6a and Fig. 6b.

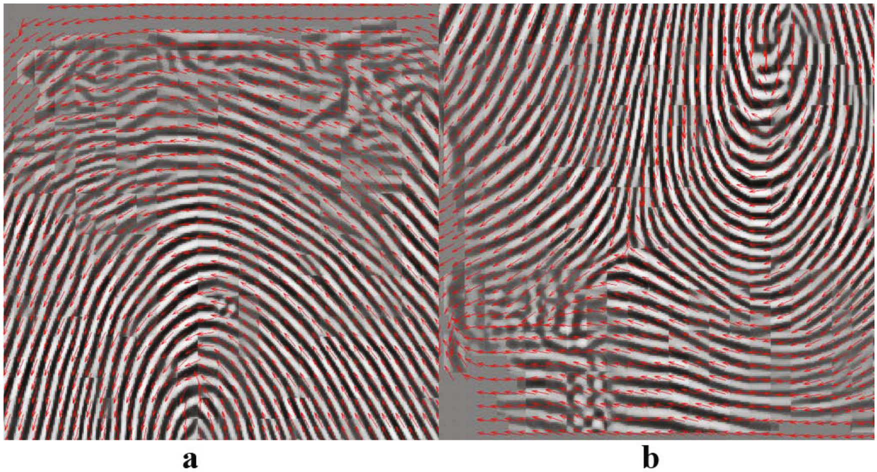


Fig. 6 Local values of ridge orientations in the noised areas of fingerprint impression: (a) top area of the impression above the upper core of the whorl, (b) left bottom area of impression containing left triradius and the lower core of the whorl

8 Fingerprint Classification with SVM Algorithm

The data set used in the experiment consists of 600 fingerprint 8-bit grey scale images. The size of the images is 512 x 512 pixels. The classes of left loop, right loop, whorl, arch and tented arch are uniformly distributed in the data set and consist of 120 images each. Images were in the first stage pre-processed using CLAHE algorithm and then filtered using STFT. Ridge orientation maps were computed for the filter enhanced images using PCA and multi-scale pyramid decomposition algorithm [3]. Feature vectors extracted from the images are used as input to an ensemble of SVM classifiers trained with one vs one voting method. RBF and triangular [4] types of kernel functions are used by the classifiers. Training and testing datasets each consist of 300 ridge orientation maps calculated from the fingerprint images. Both datasets contain 60 maps for each of the five classes. Cross-validation and grid search were used to obtain kernel parameters for the training of the SVM algorithms. Both the SVM trained with the RBF kernel and the SVM trained with the triangular kernel achieved classification accuracy of 91,3 % on the test data set.

9 Summary

In the paper a medical decision support system is proposed for the diagnosis of Down's syndrome. The system infers the occurrence of genetic disorder using an expert system. Sets of rules required for the creation of the expert system's clauses originate from the text description of the problem found in the medical literature. Premises of the conditions for the created expert system are the features obtained using image processing and pattern recognition methods. Automatic analysis of dermatoglyphic prints of fingers, palms and soles using pattern recognition methods is required for the determination of patients' diagnostic score. Currently achieved results of fingerprint classification allow a reliable determination of the left and right index finger patterns, which are two of the traits used in the nomogram diagnostic index. The most common pattern types in Down's syndrome which are left and right loops are classified with a 95.0% accuracy for the left loop type and a 96.7% accuracy for the right loop type when using the triangular kernel function in the SVM algorithm. Available data estimate that around 60000 people living in Poland were diagnosed with Down's syndrome. From this fact stems the importance and advantage of screening tests conducted using low cost non invasive methods over the limitations of costly genetic tests. Development and implementation of the decision support system proposed in the paper may lead to extension of research to other genetic syndromes, i.e. Turner syndrome, and in the later perspective to other syndromes' patterns.

Acknowledgments. This work was partially supported by the project "Innovative Resovia University" implemented under the program of Ministry of Science and Higher Education "Innovation Creator - support for innovative academic entrepreneurship" through a Contract No. 13/PMKI/U/11-11.09/2010 of 27.10.2010.

References

1. Chikkerur, S., Cartwright, A.N., Govindaraju, V.: Fingerprint image enhancement using STFT analysis. Pattern Recognition 40, 198–211 (2007)
2. Cummins, H., Midlo, C.: Finger Prints Palms and Soles - Introduction to Dermatoglyphics. Dover Publications Inc., New York
3. Feng, X.G., Milanfar, P.: Multiscale principal components analysis for image local orientation estimation. In: Proc. of the 36th Asilomar Conf. on Signals, Systems and Computers, vol. 1, pp. 478–482 (2002)
4. Fleuret, F., Sahbi, H.: Scale-invariance of support vector machines based on the triangular kernel. In: Proc. of the IEEE Int. Conf. on Computer Vision, ICCV (2003)
5. Preuss, M.: A screening test for patients suspected of having Turner syndrome. Clinical Genetics (10), 145–155 (1976)
6. Reed, T.: Dermatoglyphics in Down's syndrome. Clinical Genetics (6), 236 (1974)
7. Reed, T.E., et al.: Dermatoglyphic nomogram for the diagnosis of Down's syndrome. J. Pediat. (77), 1024–1032 (1970)
8. Tornjova-Randelova, S.G.: Dermatoglyphic characteristics of patients with Turners syndrome. Medicine Anthropologie 43(4), 96–100 (1990)
9. Yin, Y., Weng, D.: A New Robust Method of Singular Point Detection from Fingerprint. In: Proc. of Int. Symposium on Information Science and Engieering. IEEE (2008)

Parallel Servers Scheduling with Dynamic Sequence-Dependent Setup Time

Thanh-Cong Dinh and Hyerim Bae

Abstract. This paper deals with task scheduling problem in industrial consulting companies with dynamic sequence-dependent setup time by considering parallel servers (or multiple servers). The problem is transformed to multi-depot multiple asymmetric traveling salesman problem (MmATSP). A mixed integer programming model is formulated to minimize sum of tardiness time and sum of setup time.

Keywords: Dynamic sequence-dependent setup time, task scheduling, parallel servers, multiple servers, mixed integer programming, multi-depot multiple asymmetric traveling salesmen (MmATSP).

1 Introduction

Main business related to software in industrial consulting companies is software sales and daily maintenance. Maintenance tasks in these companies involve software customization and troubleshooting to avoid unnecessary error because of different operating system (OS) and software environment with customers'. For example, customers who use different software or different versions of given software may request numerous maintenance tasks. Thus, the company should keep all software in its testing server. However, due to the incompatibility of software and OS with customer request, this company may have to uninstall the incompatible software (or OS) and then reinstall the suitable one. This kind of tasks require significant cost such as analyzing time, working time (setup time), and an effort to keep the due date. Therefore, reducing the setup time and assuring on-time delivery becomes an important issue for this kind of company.

Dinh et al. (2011) introduced an approach to solve task scheduling problem in industrial consultant companies considering a single testing server. They addressed the case "dynamic sequence-dependent setup time" which means that

Thanh-Cong Dinh · Hyerim Bae
Business & Service Computing Lab., Industrial Engineering, Pusan National University, 30-san Jangjeon-dong Geumjong-gu, Busan 609-735, South Korea
e-mail: `cong.dinh@pusan.ac.kr, hrbae@pusan.ac.kr`

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 79–87.
springerlink.com

setup time between two tasks not only depends on their sequence, but also on the previous sequences. The model was described as single-machine scheduling problem and transformed to Asymmetric Traveling Salesman problem (ATSP). A mathematical model using integer nonlinear programming (INLP) was formulated with objective function is to minimize sum of squared tardiness. Due to the difficulty of NP-hard problem, the authors developed two heuristics based on Genetic Algorithm (GA) and Tabu Search (TS) to find an appropriate solution for this kind of problem.

This study, as an extended version of Dinh et al. 2011, has two additional salient features compare to previous version. First, the current research considers *m* servers run parallel. The problem is transformed to MmATSP which is a generalization of well-known multiple TSP. Second, the INLP model from Dinh et al. 2011 is extended and linearized to a mixed integer linear programming (MIP) model which objective is to minimize sum of tardiness and sum of setup time.

The paper is organized as follows. Section 2 reviews related works. Section 3 presents formulation of the problem. Numeric experiments are shown in section 4. Then, conclusion and future works are given in section 5.

2 Related Works

Multiple machines (or parallel machines) scheduling has been studied in the field of manufacturing. Chung et al. (2006) introduced a branch and bound algorithm to solve *m*-machine permutation flowshop problem with the total tardiness objective. Plateau and Rios-Solis (2010) considered unrelated parallel machines scheduling problems where jobs have earliness and tardiness penalties. The author used quadratic programming to minimize the sum of earliness and tardiness costs. However, to the best of our knowledge, there is no research about parallel servers scheduling in industrial software consulting company with dynamically sequence-dependent setup time. Therefore, this study is considered as the first research work in this field. In addition, research by Cirasella et al. (2001) has shown that ATSP solution is the most useful method for machine scheduling problem. Thus, ATSP is a suitable model to apply in this research.

3 Formulation of Problem

3.1 Problem Transformation

The *m* servers scheduling problem firstly is transformed to MmTSP in which salesmen and cities represent servers and tasks, respectively. Then, one of the approaches used for solving the MmATSP is to transform the problem to ATSP (Bektas 2006). Hence, in this paper, we convert the MmATSP with *m* salesmen and *n* cities into previous work ATSP (Dinh et al. 2011) by adding (*m*-1) home cities into the graph (Fig. 1). As mentioned above, a task can be considered as a city in ATSP and the initial status is the home city where salesman is located. Each server has its own initial status different from others; hence salesmen start at different home cities (depots). Completion of all tasks means the salesman visits all cities in his tour and then returns to his home city. The setup time is distance from one city to another and the processing time of task is the travel time in the city. Note that if c_{ij} is the distance between city i and city j, then $c_{ij} \neq c_{ji}$ (asymmetric) and the triangle inequality $c_{ij} \leq c_{ik} + c_{kj} \quad i, j, k, k \neq i, j$ is satisfied.

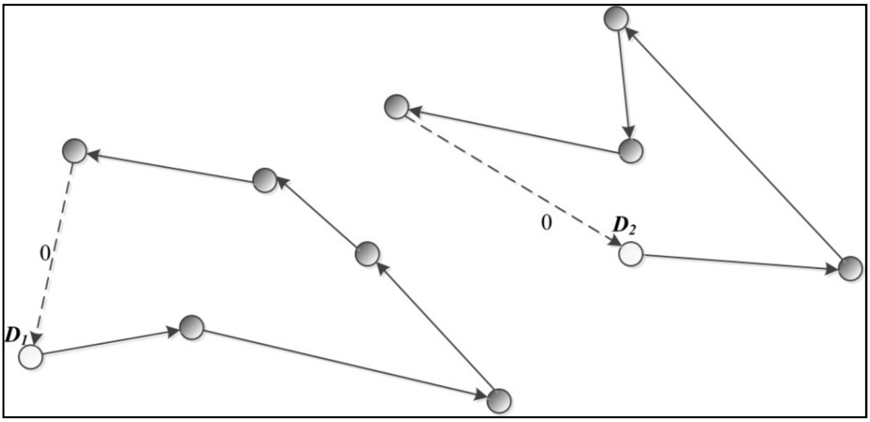


Fig. 1 An example of MmATSP with two salesmen start at two depots

3.2 MIP Model

The model considers eight assumptions: (1) the start time t is at zero. (2) For each task, the expected processing time and due date are known. (3) Software will be uninstalled immediately if it is incompatible with other required software. (4) Task is processed only one time without preemption. (5) Customer can require only one kind of OS for the task, and any OS is accepted if there is no requirement of a specific OS. (6) If there is a reinstall OS event, all current software in the server will be uninstalled without additional time consumption. (7) The setup time for returning back to the start node is zero. (8) The number of server is at least two, and all servers are idle at start time.

Let consider a graph in which nodes present cities (or tasks) and arcs are travel distances (or completion time) between cities. The model is describes in detail as follows.

Sets

V	set of nodes
D	set of home nodes, $D \subset V$
T	set of target nodes, $T \subset V$
K	set of software
N	set of OS
S	set of servers

Parameters

i_k	install time of software k
u_k	uninstall time of software k
r_n	reinstall time of OS n
p_i	expected processing time of task i
m_{kl}	1 if software k and l are incompatible, otherwise 0
s_{ik}	1 if task i requires software k, otherwise 0
o_{in}	1 if task i requires OS n, otherwise 0
M	a relative big number
w	working hours per day

Decision variables

E_{ik}	1 if the software k exists in the server when task i is handled
Q_{in}	1 if the OS n exists in the server when task i is handled
x_{ij}	1 if there is a directed arc from node i to node j, otherwise 0. It means that task j is the successor of task i
v_i	1 if server i is chosen, otherwise 0
e_{ij}	1 if there is no OS reinstall event, otherwise 0

Let α and β are weighted numbers that are associated with sum of tardiness time and sum of setup time, respectively. If W_i expresses tardiness time of task i, and S_{ij} is setup time during the changeover between task i and task j. Then the objective function of this model is presented as follows.

$$Min\ Z = \alpha \sum_{i \in T} W_i + \beta \sum_{i \in V, j \in T} S_{ij} \tag{2.1}$$

W_i can be calculated by:

$$W_i \geq C_i - wD_i \quad \forall i \in T \tag{2.2}$$

Where C_i and D_i express completion time of task i and its expected due date (in hours), respectively. Suppose C_i can be determined by following constraints:

$$C_j - p_j - C_i - S_{ij} \geq -M(1 - x_{ij}) \quad \forall j \in T, i \in V, i \neq j \tag{2.3}$$

$$C_i = 0 \quad \forall i \in D \tag{2.4}$$

Let U_{ij}, I_{ij}, R_{ij} are the times required to uninstall software, to install software and to reinstall OS during the changeover from task i to task j, respectively. Following constraints are to calculate S_{ij}:

$$S_{ij} = U_{ij} + I_{ij} + R_{ij} \quad \forall i \in V, j \in T, i \neq j \tag{2.5}$$

$$I_{ij} = U_{ij} = R_{ij} = 0 \quad \forall j \in D, \forall i \in T \tag{2.6}$$

$$r_n(Q_{jn} - Q_{in}) - M(1 - x_{ij}) \leq R_{ij} \leq Mx_{ij} \quad \forall j \in T, i \in V, i \neq j \tag{2.7}$$

$$R_{ij} \le M(1-e_{ij}) \quad \forall j \in T, i \in V, i \ne j \tag{2.8}$$

$$e_{ij} \ge 1 - \sum_{n \in N} \max(0, Q_{jn} - Q_{in}) \quad \forall j \in T, i \in V, i \ne j \tag{2.9}$$

$$\sum_{k \in K} u_k \max(0, E_{ik} - E_{jk}) - M(2 - e_{ij} - x_{ij}) \le U_{ij} \le M e_{ij} \quad \forall j \in T, i \in V, i \ne j \tag{2.10}$$

$$I_{ij} \ge \sum_{k \in K} i_k \max(0, E_{jk} - E_{ik}) - M(2 - e_{ij} - x_{ij}) \quad \forall j \in T, i \in V, i \ne j \tag{2.11}$$

$$\sum_{k \in K} i_k E_{jk} - M(e_{ij} + 1 - x_{ij}) \le I_{ij} \le M x_{ij} \quad \forall j \in T, i \in V, i \ne j \tag{2.12}$$

$$W_i \ge 0, C_i \ge 0, R_{ij} \ge 0, U_{ij} \ge 0, I_{ij} \ge 0 \quad \forall i, j \in V \tag{2.13}$$

Suppose that for an expression *F*, the function max(0, *F*) can be linearized by introducing a new binary variable *y*, then following constraints should be included in the model for each corresponding expression in constraints (2.9), (2.10) and (2.11):

$$\max(0, F) \ge F \tag{2.14}$$

$$F \le My \tag{2.15}$$

$$\max(0, F) \le My \tag{2.16}$$

$$F > -M(1-y) \tag{2.17}$$

Note that the constraint (2.17) should be changed to suitable form of MIP by adding a new parameter λ which is a relative small number enough to satisfy the condition in constraint (2.17). Hence, constraint (2.17) is transformed to equation as follow:

$$F \ge \lambda - M(1-y) \tag{2.18}$$

Constraints (2.7) and (2.10) are to calculate reinstall time and uninstall time. Constraints (2.8) and (2.9) are to ensure the occurrence of reinstall OS event. If there is no reinstall OS event occurs then reinstall time has to be zero, otherwise reinstall time will greater than zero and all current software will be uninstalled without any additional time. In the same context, if there is no reinstall OS event, the install time during changeover between task *i* and task *j* is calculated by constraint (2.12), otherwise constraint (2.11) has to be satisfied.

In order to ensure there is always one OS installed in the servers, following constraints are used:

$$Q_{in} \ge o_{in} \quad \forall i \in V, n \in N \tag{2.19}$$

$$\sum_{n \in N} Q_{in} = 1 \quad \forall i \in T \tag{2.20}$$

The assumption software will be uninstalled immediately if it is incompatible with other required software is expressed by constraints (3.21) and (3.22)

$$\text{If } m_{kl} = 1 \text{ then } E_{ik} + s_{ik} + E_{il} + s_{il} \le 2 \quad \forall i \in T, k \in K, l \in K, k \ne l \tag{2.21}$$

$$\text{If } m_{kl} = 0 \text{ then } E_{ik} + s_{ik} + E_{il} + s_{il} \le 4 \quad \forall i \in T, k \in K, l \in K, k \ne l \tag{2.22}$$

The requirement of software for performing any task i is described by constraint (3.23)

$$E_{ik} \ge s_{ik} \tag{2.23}$$

Finally, constraints of MmATSP are added:

$$\sum_{i \in V} x_{ij} = 1 \quad \forall j \in T, i \ne j \tag{2.24}$$

$$\sum_{j \in V} x_{ij} = 1 \quad \forall i \in T, i \ne j \tag{2.25}$$

$$\sum_{i \in T} x_{ij} = v_j \quad \forall j \in D \tag{2.26}$$

$$x_{ij} = 0 \quad \forall i \in D, j \in D \tag{2.27}$$

Constraints (2.24) and (2.25) are to enforce each node has degree two (one arc in, and one arc out). And, constraint (2.26) is to ensure that each home node (server) which is chosen has to have one arc out. Constraint (2.27) enforces that there is no connection between home nodes.

4 Numerical Experiments

We generated 6 problem instances that have different number of tasks, number of software, number of OS and number of servers. For weighted values of the objective we considered $\alpha = \beta = 1$. The list of problems is shown in

Table 1 List of problem instances

Problem instance	Problem size				No. of variables		No. of constraints
	No. of tasks	No. of software	No. of OS	No. of servers	Integer	Total	
1	4	2	20	3	2332	2466	4958
2	4	3	15	2	1434	1549	3019
3	6	2	7	3	1662	1906	3581
4	6	2	20	2	3068	3285	6909
5	15	2	7	2	5898	6970	14164
6	15	3	15	3	12561	13696	28297

All experiments are conducted on a computer with an Intel Core i5 2.27GHz and 4.00GB of RAM. Fig. 2 shows the computation times of the problems listed in Table 1.

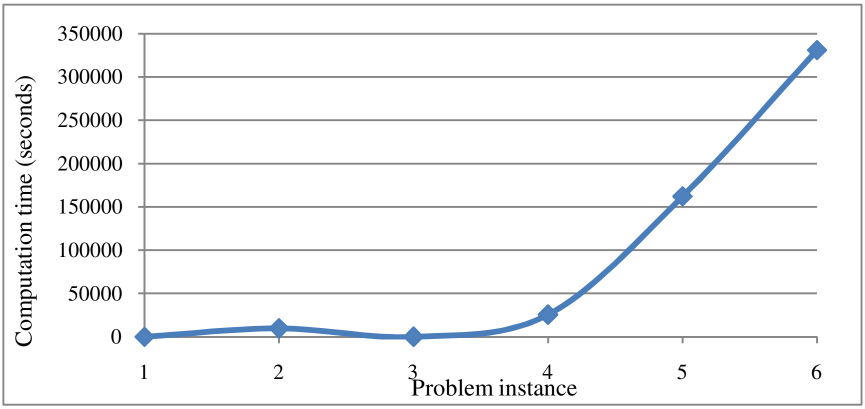
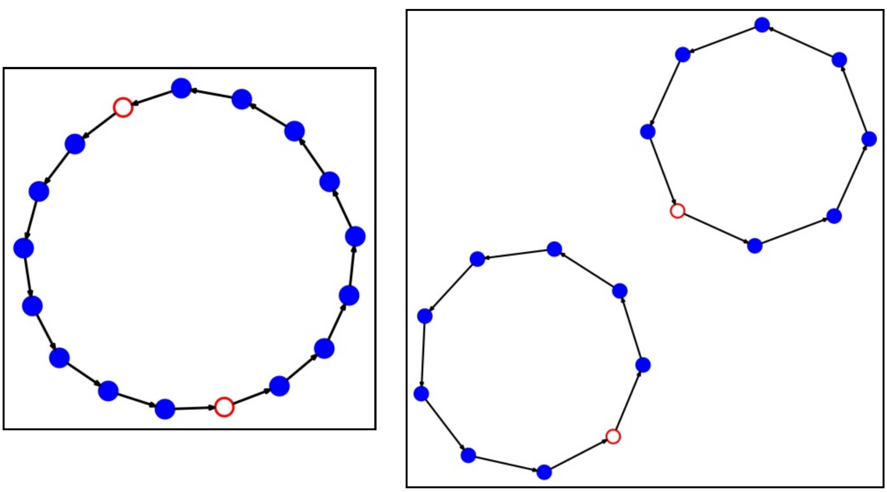


Fig. 2 Computation time of problem instances

Fig. 3 illustrates two best solutions of problem instance number 5. The unfilled nodes correspond to the home nodes, the other filled nodes are target cities (or tasks need to be performed). Although the two solutions give different tours, the objective results are the same at 77.76 hours. Fig. 4 illustrates a possible solution of problem instance number 6.



(a) (b)

Fig. 3 Two best solutions of problem instance number 5

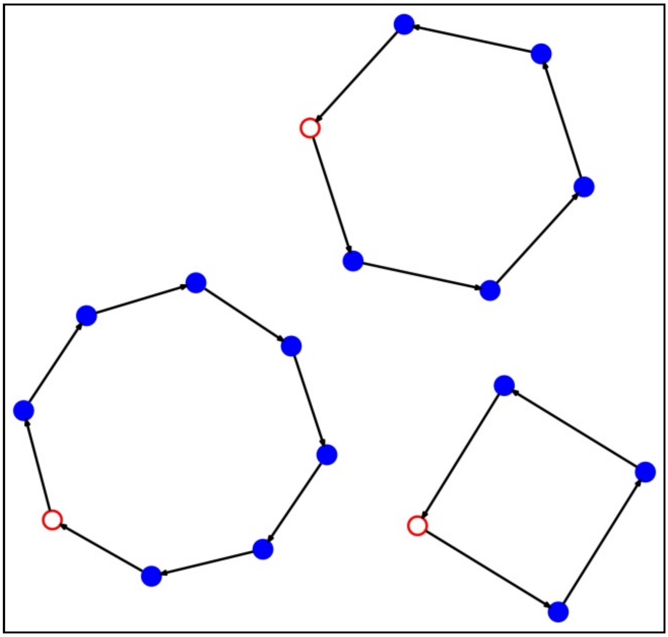


Fig. 4 A best solution of problem instance number 6

5 Conclusion

This research proposed a method to deal with parallel servers scheduling including dynamic sequence-dependent setup time in industrial consulting companies. A MIP model which objective is to minimize sum of tardiness time and sum of setup time is developed. However, the MIP model performed inefficiently when the problem becomes more complex. Experiments showed that as problem is larger, the computation time rapidly increases. Hence, heuristics should be developed to overcome this issue. Our future work is to apply Lagrangian relaxation based heuristic to work out the lower bound and approximate solution for this kind of problem.

Acknowledgments. "This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No.2010-0027309)".

References

Bektas, T.: The multiple traveling salesman problem: an overview of formulations and solution procedures. Omega-International Journal of Management Science 34(3), 209–219 (2006), doi:10.1016/j.omega.2004.10.004

Chung, C.-S., Flynn, J., Kirca, Ö.: A branch and bound algorithm to minimize the total tardiness for m-machine permutation flowshop problems. European Journal of Operational Research 174(1), 1–10 (2006), doi:10.1016/j.ejor.2004.12.023

Cirasella, J., Johnson, D.S., McGeoch, L.A., Zhang, W.: The Asymmetric Traveling Salesman Problem: Algorithms, Instance Generators, and Tests. Paper presented at the Revised Papers from the Third International Workshop on Algorithm Engineering and Experimentation

Dinh, T.C., Wang, Z., Bae, H., Ryu, K.: Task Scheduling with Dynamic Sequence-dependent Setup Time using Asymmetric TSP. In: International Conference on Flexible Automation and Intelligent Manufacturing 21st, Taiwan (2011)

Plateau, M.C., Rios-Solis, Y.A.: Optimal solutions for unrelated parallel machines scheduling problems using convex quadratic reformulations. European Journal of Operational Research 201(3), 729–736 (2010), doi:10.1016/j.ejor.2009.03.049

Research on Complexity and Evolution of Business Model Innovation

Wang Qian, Li Chaoyang, Jiang Qian, and Qi Jiangna

Abstract. Business models have become the focus of enterprises as the result of an increasingly fierce environment of market competition, with business model innovation becoming the most important way for enterprises to gain a competitive advantage. In this paper, we study the complexity of business model innovation from a systemic viewpoint, exploring the evolution rules of different complexities. This knowledge can help enterprises make better decisions to ensure the success of their business model innovation.

1 Introduction

Business model innovation is becoming the most important way for enterprises to gain a competitive advantage. In fact, because of increasingly fierce market competition, business model innovation has taken an increasingly prominent role within enterprise business activities. Businesses such as DELL, Amazon.com, and Amway have achieved great success due to business model innovation.

Most business model research focuses more on interpretation of the connotation and elements, believing that innovation of the business model are simply a matter of changing its constituent units. However, there are numerous factors covering a series of business activities, which can affect business model innovation. These factors include target customers, customer processes, business processes, products and services, distribution channels, logistics management processes, and so on. Obviously, the process of business model innovation is difficult to interpret because of this conclusion, or even to determine its characteristics and laws. As research in the topic has broadened, a number of scholars now believe that a business model is the comprehensive integration of a series of elements by the enterprise; elements such as economic logic, operational structure, strategic direction, and so on. The process of business model innovation, with its complex and

Wang Qian · Jiang Qian · Qi Jiangna
School of Business, Sun Yat-Sen University, Guangzhou China

Li Chaoyang
China Overseas Land & Investment Ltd., Shenzhen, China

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 89–98.
springerlink.com

systematic features, is affected not only by the specific morphology of its constituent units, but also by the contact between them [1~5]. Therefore, determining business model complexity, the interaction between component units, and the combined effects, will reveal the mechanisms of business model innovation and should provide a theoretical basis [6,7]. In this paper, we build a business NK model based on the "3-4-8" system from a value perspective. We then simulate the evolution process for business models of different complexity using statistical mechanics and random numbers. Finally, we summarize three rules of business model innovation by analyzing the simulation results.

2 Reviews

Initial business model research has mainly focused on model definition and composition elements from economic, operational, and strategic points of view. From the economic perspective, some scholars think it is the model that obtains the economic profit, while the model's definition and composition are related to economic indicators. There are some advantages to describing the definition and composition of a business model from an economic perspective: the method easily obtains actual data related to compositions; it further expands more quantitative research on business models; and it highlights the importance of the business model by emphasizing profits and the profit model.

Research that focuses on business model innovation mechanisms and driving forces can be a useful way to help enterprises find the right business model when under fierce competition. In innovation mechanism research, scholars treat the business model as a process of value creation through internal process and basic structure design, based on value chain theory. Meanwhile, in researching the driving forces of business model innovation, scholars are trying to determine the core force that encourages enterprises to carry out business model innovation [8]. Kagerma and Weiying Zhang believed that a new business model was the driving force behind enterprises maintaining a competitive advantage. Future business model innovation has the following trends: 1) innovation changes from product-oriented to consumer-oriented; 2) the operating model changes from enterprise-centric to customer-centric; and 3) the business model changes from product-driven to services and solution-driven. Meanwhile, Qian Wang [9] analyzed the mechanisms of an IT-driven business model and the selection of its innovation path from an IT-driven perspective, believing it would become an endogenous force for business model innovation.

The business model is a functional organic system serving as a carrier of the enterprise. The various component units possess a complex and close relationship, while the core is the effective integration of resources. Therefore, some scholars research business model systems and the impact of the constitute units from an operational perspective. Di Xu [7] studied business model system complexity and built the NK model, providing a theoretical basis for the application of business model innovation using complex system theory. Yuan lei [10] believe that business model is the value creation logic of corporations at all. Basing on thus cognition, he brings forward the'3-4-8' structure system of business model. The

’3-4-8’structure system is a three dimensional framework seeing about the business model from “far-midst-near” in the round. The business model innovation may change the business model core logic on different degree, which was applicable to the different environmental dynamic characteristic and industrial lifecycle respectively.

These studies on business model innovation have mainly focused on the theory of connotation and the business model elements. Research to date has rarely analyzed the interactions between those elements, or the mechanisms of innovation. In actual operations, any change of status and interaction of the component units can be considered a new business model innovation for the whole business operation. However, the traditional NK model cannot objectively reflect business model innovation evolution, nor does it provide a theoretical basis for business model innovation, as the model is constructed based on the ideal assumption that component units are homogeneous and provide an equal contribution.

3 The Business NK Model

3.1 Modified NK Model

In order to better study the complexity and innovative evolution of business models, the traditional NK model was improved based on the value creation logic of a given business model. This new model relaxes the ideal assumptions and considers the heterogeneity and different contributions of the component units. Within the framework of the value chain, the value creation logic of the business model has three contact interfaces in the system: customer value, partner value, and enterprise value. Each contact interface consists of four units—the value proposition, the value network, value maintenance, and value realization—with each unit consisting of a further factors: target customer and value content for the value proposition, network configuration and business location for the value network, partnership and isolating mechanisms for value maintenance, and revenue model and cost management for value realization [10]. This system is referred to as the "3-4-8" system, and its specific structure is shown in Fig.1. In the modified NK model, N represents the number of component units in the complex system of the business model, while K represents the complexity of the relationship between the component units. A large number of simulation results have shown that the fitness distribution is mainly affected by the parameters N and K, but not the quantity of factors.

Therefore, in this model, each unit can be simplified by using a binary set for each factor, respectively given by 0 and 1; meanwhile, the inter-relationship of each unit is defined by the host relationships. If changes in a unit do not refer to other units, then the relationship is represented by a 0, otherwise the relationship is represented by 1. If the change refers to multiple units, then the number of host relationships is cumulative.

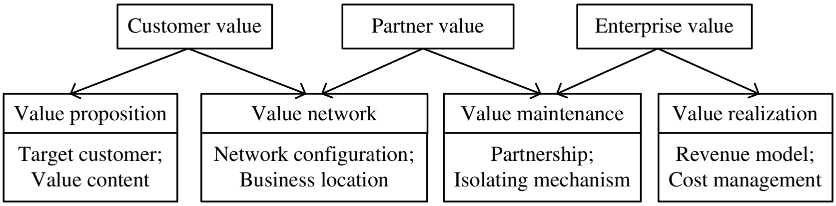


Fig. 1 Business model system based on the value creation logic (as the model of "3-4-8")

Definition 1: If component unit x is affected by unity, then there is a host relationship between x and y, given as x ;

Definition 2: In the business model, there should at least be a direct host relationship between two component units, such that $K_{min} = 12$ is the minimum complexity.

Definition 3: The sum of the number of host relationships among component units in the business model is called the complexity scale.

In order to simplify the computational complexity of the simulation, we assume that the host relationships are different but that the sum of its number is the same; i.e. seems to have the same complexity.

Set class A to represents the component units of the business model. Then A={a, b, c, d}, where a represents the value proposition units, b the value network units, c the value maintenance units, and d the value realization. The function of the host relationship within the component units is then $f_R(x) = \{xRy, x, y \in \{a, b, c,$ where K=1221, …, 3333, and each host relationship corresponds to a different complexity of the business model.

We can write the host relationship $\mathcal{R} \subseteq \mathcal{A} \times$, and its matrix as

$$\mathcal{M}_{RK} = (r_{ij})_{m\times n}, \quad \text{where } r_{ij} = \begin{cases} 1, & if\ xRy \\ 0, & other \end{cases} \tag{1}$$

Definition 4: The overall system fitness is the advantage of a specific environment. We set the overall system fitness for the business model as W, where

$$W = \sum_{i=1}^{4} \alpha_i \omega_i \tag{2}$$

Here, represents the system fitness of each unit and indicates the contribution of each unit for overall system fitness, where

$$\alpha_i < 1, \sum_{i=1}^{4} \alpha_i = 1, i = 1,2,3,4 \tag{3}$$

In order to reduce computational complexity, this paper divides each component unit into two types: an intermediate link and a terminal link. When we consider the effects of the different contributions of the component units, this refers to the same contribution of the value proposition and value realization, and it is the same as the value network and value maintenance.

3.2 Model of Business

For this study, we analyzed the business model complexity for K=1221, …, 3333, or a total of 21 cases. However, we only discuss some typical cases within the article, as a treatment of each individual case would be far too lengthy and would not really add any significant value in terms of demonstrating the results of this paper.

Case 1: complexity K=1231

When the model complexity is K=1231 according to the “3-4-8” business model system, the value proposition is directly affected by the value network. This means the value proposition must change when the value network innovates. At the same time, the value network is affected by the value proposition and value maintenance; if one of those values changes, it may affect the value network. The NK model is $f_{1231}(a) = \{a\mathcal{R}a, a\mathcal{R}b\}$, $f_{1231}(b) = \{b\mathcal{R}a, b\mathcal{R}b, b\mathcal{R}c\}$, $f_{1231}(c) = \{c\mathcal{R}a, c\mathcal{R}b, c\mathcal{R}c, c\mathcal{R}d\}$, $f_{1231}(d) = \{d\mathcal{R}c, d\mathcal{R}d\}$,and its relationship matrix is

$$\mathcal{M}_{R1231} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

In this model, the value maintenance unit is affected by other units. Any innovation of others units will change the value maintenance unit, which is the most flexible part of the business model. This business model usually occurs in larger enterprises with strong bargaining power, which can easily adjust partnership and isolating mechanisms according to their value proposition.

Case 2: Complexity K=2332

If complexity is K=2332 within the ‘3-4-8’ business model system, the value proposition and value realization are affected by the value network and value maintenance, respectively. On the other hand, the value network and value maintenance are affected by any component units of the business model, and are more supportive and flexible. This case’s NK model is $f_{2332}(a) = \{a\mathcal{R}a, a\mathcal{R}b, a\mathcal{R}c\}$, $f_{2332}(b) = \{b\mathcal{R}a, b\mathcal{R}b, b\mathcal{R}c, b\mathcal{R}d\}$, $f_{2332}(c) = \{c\mathcal{R}a, c\mathcal{R}b, c\mathcal{R}c, c\mathcal{R}d\}$,

$f_{2332}(d) = \{d\mathcal{R}b, d\mathcal{R}c, d\mathcal{R}d\}$and its relationship matrix is

$$\mathcal{M}_{R2332} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

With this business model, enterprises typically exhibit operating characteristics according to the value proposition and value realization. The company’s value network and maintenance value are relatively simple, changing according to the value proposition and value realization. An example is emerging internet companies, such as Facebook, who provide customers with a wide variety of Internet products. These companies do not charge users directly, but instead charge the enterprises which provide embedded advertising and provision of user data. This model provides a new value content and explores a new way for companies to profit.

Case 3: Complexity K=3223

When complexity is K=3223, the two component units of the business model system, the value network and value maintenance, also affect each other, as in the previous case. However, in contrast with the K=2332 case, the two values can only affect either value proposition or value realization at any given time, while these same values can still both be affected by any units at the same time. This case's NK model is $f_{3223}(a) = \{aRa, aRb, aRc\}$ $f_{3223}(b) = \{bRa, bRb, bRc, b$ $f_{3223}(c) = \{cRa, cRb, cRc, c\}$, $f_{3223}(d) = \{dRb, dRc, d\}$ and its relationship matrix is

$$\mathcal{M}_{K3223} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

In this business model, the value network and value maintenance are closely interrelated and influence each other; any innovation of either unit greatly changes the whole business model. At the same time, the value proposition and value realization are very flexible, and need to be adjusted to constantly adapt to changing scenarios. This business model exists in monopoly enterprises, which are good at dealing with partnerships. These enterprises are sensitive to any change in the industry chain, adjusting their value proposition and value realization when there is any change of the value network or value maintenance because of service location or new partnerships.

Case 4: K=3333

The most complex business model occurs when complexity K=3333 with the "3-4-8" business model system. In this model, there is a direct inter-relationship between each unit, and every unit is affected by all the other units in the system. In this business model, all units change when any one unit changes. Therefore, a variety of factors should be considered in the process of reform in this case. The case's model is $f_{3333}(a) = \{aRa, aRb, aRc, a\}$, $f_{3333}(b) = \{bRa, bRb, bRc, b$ $f_{3333}(c) = \{cRa, cRb, cRc, c\}$, $f_{3333}(d) = \{dRa, dRb, dRc, d\}$, and its relationship matrix is

$$\mathcal{M}_{K3333} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}。$$

Enterprises that use this business model may be doing so due to operating in a perplexing market environment, or if its service relationship is complicated. In short, any simple innovation affects the whole business model, and there is considerable difficulty and high risk involved in any business model innovation.

4 Experiment

4.1 Principle and Index

To analyze the importance of complexity, we can simulate business model evolution using random numbers and describe the business model using a fitness function. When a business unit change, we assign a random number as its new fitness, and the same is done for the units which are affected by the unit with the initial change. This results in a change in the total fitness for the business model, corresponding to the innovation of the business. At the same time, the nonlinearity and uncertainty of the simulation correspond to the uncertain profit which business innovation brings. According to statistical mechanics of complex systems, a single simulation is just one possible potential business model evolution. To get a more stable result, a significant number of simulations are required. Experiments show that the minimum number of iterations is 500,000 to achieve statistically meaningful results.

We observe business model evolution by looking at the total fitness of the model, including the mean fitness of local optima (MFLO) and the mean fitness of global optima (MFGO). The mean fitness of local optima is the mean fitness of all local optima in the evolution; MFGO is the mean fitness of the global optima which is the highest fitness in the evolution, such that

$$MFLO = \frac{1}{n \times m} \sum_{j=1}^{n} \sum_{m=0}^{15} \max_m(\omega_m(x)) \tag{5}$$

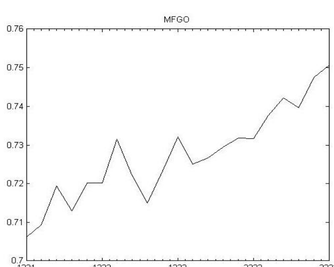
$$MFGO = \frac{1}{n \times m} \sum_{j=1}^{n} \sum_{m=0}^{15} \max_m(W_m) \tag{6}$$

Here, N is a set of four binary neighbor numbers from 0 to 15, $x \in NS_m$, $m = 0,1,\cdots$, is the average fitness of each set, and n=500,000.

4.2 Process

Experiment 1: Units are heterogeneous but share the same weight

In this experiment, we determine the evolution of business model innovation for $\alpha_i = 0.25, (i = 1,2,3$ as follows, without considering the different weights of the units. Figure 2(a) and Figure 2(b) show that MFLO and MFGO become larger as the model becomes more complex.



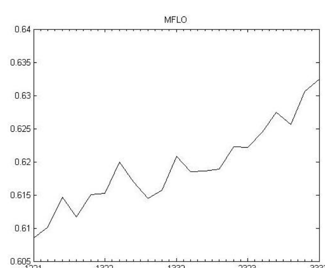


Fig. 2(a) $MFLO(\alpha_i = 0.25, i = 1,2,3,4)$ **Fig. 2(b)** $MFGO(\alpha_i 0.25, i = 1,2,3,4)$

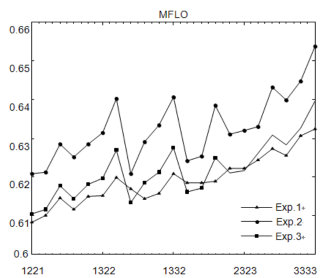
Experiment 2: Units are heterogeneous and have different weights

In some cases, weights are different for the business models units which are heterogeneous. According to definition 4, we divided the units into two types according to their positions. One set of units became the intermediate units and the other set the extremity units. We then performed experiments for five different trials using two different cases.

① $\alpha_1=\alpha_4<\alpha_2=$

For this case, we observed the evolution of business model innovation for three different trials: $\alpha_i = 1/4 (i=1,2,$; $\alpha_1=\alpha_4=$, $\alpha_2=\alpha_3=$; and $H_3: \alpha_1=\alpha_4=$, $\alpha_2=\alpha_3=$. The trial will be the reference trial.

As seen in Figure 3(a) and Figure3(b), we can see that the MFLO and MFGO trends are almost the same for all three trials. The curve of is smoothest but its value (the overall fitness) is lowest of the three trials. In contrast, the curve of fluctuates the most and its value is highest.



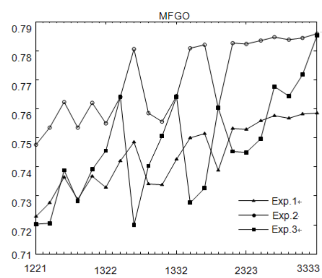


Fig. 3(a) $MFLO\ (H_1, H_2, H_3)$ **Fig. 3(b)** $MFGO\ (H_1, H_2, H_3)$

② $\alpha_2=\alpha_3<\alpha_1=$

In the second case, the intermediate units of the business model are more important than the extremity units, such that $\alpha_2=\alpha_3<\alpha_1=$. Again, we use as the reference to observe the evolution of business model innovation for three different trials: $\alpha_i = 1/4 (i=1,2,$; $\alpha_1=\alpha_4=1/$, $\alpha_2=\alpha_3=$; and $H_3: \alpha_1=\alpha_4=$, $\alpha_2=\alpha_3=$. When $\alpha_1=\alpha_4=$, $\alpha_2=\alpha_3=$ or $H_3: \alpha_1=\alpha_4=$ and $\alpha_2=\alpha_3=$, shows that MFLO and MFGO both increase with increasing business model complexity. When one unit becomes more complex as others stay the same, MFLO and MFGO increase slightly, consistent with the results of Experiment 1. The value of is larger than, but their fluctuations are almost in opposite directions. For example, when the business model complexity changes from $K=1$ to $K=2$, MFLO and MFGO decrease in the trial but increase in the trial.

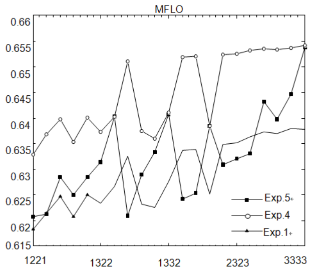


Fig. 4(a) $MFLO(H_1, H_4, H_3)$

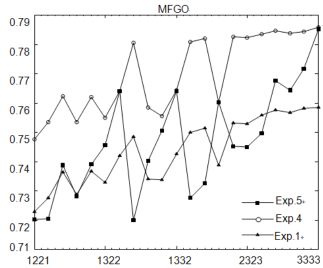


Fig. 4(b) $MFGO(H_1, H_4, H_3)$

5 Rules of Business Model Innovation

Our experiments show that business model evolution is different when the weights of the business model units changes.

Rules 1: MFLO increases with increasing complexity of the business model. That is, business benefits will be better when the units become more closely tied together. Actually, the closer the units get, the more successful the business becomes because the contact between each unit can create overall benefits for the business model far greater than the sum of the benefits of the individual units. This allows a company to gain a significant competitive advantage over other companies.

Rules 2: For business models of the same complexity scale, models which have intermediate units with more contact amongst themselves, rather than with extreme units, achieve a greater total profit. In business models, intermediate units always play a supporting role. Less contact with other units of the business model means less restriction by others so that business innovation will be easier. Many companies rely on product or technology innovation to gain more profit.

Rules 3: When business models are of the same complexity scale, if units of larger weight have less contact with other units within the business model, and units of lighter weight have more contact with other units, the total profit of the business model will be greater. This result is obtained because an adaptable value network unit and value maintenance unit can help the company gain a competitive advantage. Actually, an adaptable value network and value maintenance units with independent value propositions and value realization units can make a business model more successful. Similarly, those units which a company pays more attention to in terms of removing restrictions also bring success.

6 Conclusion

This paper constructed a new NK model for business innovation based on value theory. We determined typical business models of different complexity, and then simulated the evolution of those business models using a random number analysis, allowing us to explore the laws of business evolution in different scenarios.

However, to provide a more detailed understanding of business innovation, more work needs to be done to study the different disciplines for specific business cases.

Acknowledgements. This research was supported by the National Natural Science Foundation of China under grant 70971141.This work was also supported by the Ministry of Education Humanities and Social Sciences Planned Project under grant 09YGA630156, by the Fundamental Research Funds for the Central Universities and Soft science research projects in Guangdong grant 20090601.

References

1. Morris, M., Minet, S.: The Entrepreneur's Business Model: Toward a Unified Perspective. Journal of Business Research 58(1), 726–735 (2003)
2. Osterwalder, A., Yves, P., Chirstopher, L.T.: Clarifying Business Models: Origins, Present, and Future of the Concept. Communications of the Information Systems 15(5), 1–25 (2005)
3. Juneseuk, S., Yongtae, P.: On the creation and evaluation of e-business model variants: The case of auction. Industrial Marketing Management 38, 324–337 (2009)
4. Marcello, L., Marlon, D., et al.: Configurable multi-perspective business process models. Information Systems 36, 313–340 (2011)
5. Liliana, D., Marie, E.R.: What do business models do? Innovation devices in technology entrepreneurship. Research Policy 38, 1559–1570 (2009)
6. Weng, J.Y.: Business model Innovation. Economic Management Publishing House, Peking (2004)
7. Xu, D.: Complexity of Business Model Innovation. Economic Management Publishing House, Peking (2005)
8. Henning, K., Weiying, Z.: Business Models 2010. China Machine Publishing House, Beijing (2008)
9. Wang, Q.: Research on the Mechanism and Path of Business Model Innovation Based on IT Diving. Journal of Management 8(1), 126–132 (2011)
10. Lei, Y.: Commercial System Reconstruction. China Industrial Economy 6(6), 70–79 (2007)

Research on Emergency Crowd Behavior Based on Epidemic Dynamics

Yingchao Pei and Guangmou Wu

Abstract. Emergency crowd behavior has become an important factor affecting social stability. According to the crowd is whether affected by unexpected events, it is divided into susceptible group and infectious group. Based on the characteristics of crowd behavior, an epidemic dynamic model of emergency group behavior is built in this paper to investigate the conditions for crowd behavior as well as infection among individuals. It gives systematical quantitative analysis of group behavior. The results indicate that infection capability is the determinant of infection among individuals, while infection rate and cure rate are the necessary conditions for the group behavior. Furthermore, preventive measures, such as reducing infection capability, lowering infection rate and improving cure rate, are provided at last, which is strongly in practical.

1 Introduction

A large number of unexpected events have occurred in recent years. With the increase of emergencies, worldwide countries have grown concerns about them. It's all known that abruptness, destructiveness and durative are the key characteristics of unexpected events, while they will easily affect individual's decision-making in panic and further lead to crowd behavior. Emergency crowd behavior may cause large numbers of deaths or drastically destroy local social stability. For example, the severe stampede happened at Cambodian capital in 2010 eventually led to painful results that more than 400 people were killed and over 700 were injured. Therefore, study of emergency crowd behavior is worthy practically exploring. Typical research on emergency crowd behavior primarily focuses on evacuation and escape. A simple, rapid method for calculating evacuation time estimates that was compatible with research findings about evacuee's behavior in hurricanes was described by Lindell [1]. Ren et al presented an agent-based model to construct crowd evacuation for emergency response from an area under an explosion and tested the model effect by iterative simulations [2]. Oguz et al simulated virtual

Yingchao Pei · Guangmou Wu
Institute of Systems Engineering, Southeast University, Nanjing, China 210096
e-mail: ychpei@gmail.com

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 99–106.
springerlink.com

crowds in emergency situations caused by an incident, and employed an offline occlusion culling technique to speed up the animation and visualization [3]. Wendt et al provided a crowds behavior model to investigate the dispersal of crowds and proved some factors had an impact on the simulation [4]. Enrico et al focused on the human behavior during an evacuation of a motorway tunnel and an innovative simplified approach was been applied to study how people behave in case of emergency [5]. Capote et al used egress modeling to explore the impact that crew procedures had on evacuation two high-speed trains under different fire scenarios [6]. A scenario in which secondary users can access unused spectrum vacated by idle primaries was considered by Liu and Ding [7]. Cheng proposed an optimization method for dynamic evacuation route based on improved ant colony algorithm [8]. Hao et al studied the evacuation technology based on risk theory and how to rescue people in the fire situation [9]. Pérez et al proposed a dynamic escape route system for emergency evacuation of a naval ship, and built a mixed-integer nonlinear programming model to optimize the evacuation process[10].

Current research mainly pays attention to the characteristics of escape and evacuation, while it is lack of deep exploration to the group behavior generation process which is the basis of further study. Some essential problems are ignored though more scholars focus on this field. First of all, the crowd will generate infection and imitation under panic or surroundings in emergencies, then, what's the key factor to decide existence of these behaviors becomes an important problem. The second one, with the generation of imitation, conformity and infection among individuals, how is the crowd behavior to produce and develop in emergencies becomes a crux. This is the necessary to avoid crowd behavior. Finally, what effective preventive measures should be adopted to avoid infection and imitation as well as emergency crowd behavior is worthy of consideration in practical.

In this paper, we focus on a class of public emergencies that are difficult to predict in advance and will have a wide influence afterwards, such as influenza A/H1N1, Yushu earthquake, nuclear leakage in Japan and so on. The contribution in this paper can be unfolded as following aspects: we pay attention to the essential problems mentioned above, and propose an emergency crowd behavior model based on epidemic dynamics to analyze the generation and development process of emergency crowd behavior. Then, we consider the infection among individuals is how to produce after occurrence of unexpected events, that is, the key factor to decide the production of infection is given based on the model. Next, we emphasize on analyzing the conditions for emergency crowd behavior. Finally, preventive measures are suggested on the basis of combining dynamic model and practice.

The remainder of this paper is organized as follows. Section 2 analyzes feasibilities and proposes the behavior model. Section 3 introduces the conditions for infection and emergency crowd behavior based on the model. Then, preventive measures are given in Section 4. Finally, conclusions and future research directions are reported in Section 5.

2 Emergency Crowd Behavior Model

2.1 Feasibilities

Lux proposed a relevant model to describe the price changes in stock market with dynamics[11]. It believed that emergency of bubbles was resulted from equilibrium prices which deviated from fundamental values and were led by imitation traders. Hence, it emphasized on the imitation among traders was how to influence the price changes in stock market.

Though its research focused on imitation traders in stock market, this paper has some certain similarities in research ideas with his. The emergency crowd behavior refers that individuals affected by the behavior strategy of others are producing conformity, imitation or over-reliance so as to adopt the same behavior strategy as others. The nature of crowd behavior is the impact of others' behavior on individual behavior decision-making. American sociologist Park proposed the definition of group behavior most early, that is, "Group behavior is a kind of individual behavior under the collective promotion and influence, and it's also a mood fluctuation" [12]. The research philosophy of emergency crowd behavior is mainly analyzing the formation and propagation conditions quantitatively to reveal the development of crowd behavior. It also exists imitation or infection among individuals in the development of crowd behavior, and it's no doubt that the two behaviors will importantly affect the process of crowd behavior.

According to the research object classification and characteristics of emergency crowd behavior, the crowd can be divided into susceptible group and infectious group to reveal if an individual is affected by the impact of others or environment around. While in Lux's study, the crowd is divided into optimistic and pessimistic investors to indicate people's attitude to stock prices. Because of imitation among investors, optimistic investors may become the opposite ones, so do the pessimistic investors. While in our research, the two crowds will transform mutually.

Based on the above comparative analysis, we can see that both have a certain consistency in research ideas and classification of research object. Therefore, applying the research thoughts of Lux in emergency crowd behavior has certain feasibility and be consistent with the research purpose of crowd behavior.

2.2 Emergency Crowd Behavior Model

We'll build the following model under Lux's thoughts.

Assume the population is a constant $2N$. The crowd is divided into two groups: $S(t)$ is the susceptible number at t, $I(t)$ means the infected number at t. We suppose the infection probability among individuals is equal, then the conversion rate between two states is:

$$S'(t) = -\alpha S(t) + \beta I(t) \tag{1}$$

$$I'(t) = \alpha S(t) - \beta I(t) \tag{2}$$

Where, α is infection rate and β is cure rate. Eq. (1) represents the net change rate of population from infectious crowd to susceptible crowd, and Eq. (2) represents the opposite net change rate. We set the average state index $p = k/N$, where $k = \frac{1}{2}\left(S(t) - I(t)\right)$. We know that: $-1 < p < 0$ means the number of infectious crowd is more than that of susceptible crowd, i.e., most of people have been infected; $0 < p < 1$ is the opposite; $p = 0$ means the crowd's state is in equilibrium; $p = -1$ or $p = 1$ refers to the extreme condition that the crowd is completely affected or unaffected respectively. According to the definitions of k and p, we can get:

$$p'(t) = \frac{1}{2N}\left(S'(t) - I'(t)\right) \tag{3}$$

$S(t), I(t)$ can be expressed with N, k:

$$\begin{cases} S(t) = N + k \\ I(t) = N - k \end{cases} \tag{4}$$

Eq. (3) can be described as followings according to (1), (2), (4):

$$\begin{aligned} p'(t) &= \frac{1}{N}\left(\beta I(t) - \alpha S(t)\right) \\ &= \beta(1-p) - \alpha(1+p) \end{aligned} \tag{5}$$

The imitation and infection lead to mutual conversion between susceptible crowd and infectious crowd, that is, the generation of infection rate α and cure rate β largely depend on the mutual conversion. While p is the quantitative reflection of mutual conversion. So it's obvious that α and β are decided by the actual distribution of p.

$$\alpha = \alpha(p) = \alpha\left(k/N\right) \tag{6}$$

$$\beta = \beta(p) = \beta\left(k/N\right) \tag{7}$$

Considering the definition of α and β as well as practice, we suppose they are only related to infection capability γ and changing speed η between two crowds. So we denote that

$$d\beta/\beta = \gamma dp, d\alpha/\alpha = -\gamma dp \tag{8}$$

The above equations respectively indicate the relative change probability of population from infectious crowd toward to susceptible crowd increases linearly with p , and the relative change probability of population from susceptible crowd to infectious crowd increases linearly in the opposite direction with p . Then

$$\beta = \eta e^{\gamma p} \tag{9}$$

$$\alpha = \eta e^{-\gamma p} \tag{10}$$

Therefore, the dynamic model of emergency crowd behavior based on epidemic can be described as followings:

$$p'(t) = 2\eta(\sinh(rp) - p\cosh(rp)) \tag{11}$$

3 Emergency Crowd Behavior Based on Epidemic Model

3.1 Infection Conditions among Individuals

Considering practice, the infection capability γ is mainly subjected to the following factors: influence of others γ_1, result of group behavior γ_2 and the initiative infection led by deviation γ_3. The relationship among them is as shown below:

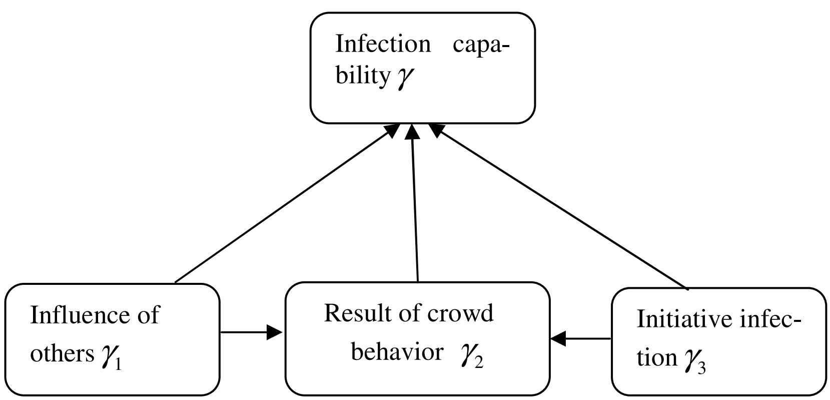


Fig. 1 Releationship of Infection Capability

(1) If $\gamma = \gamma_1 + \gamma_2 + \gamma_3 < 1$ Eq. (11) has the unique stable solution $p = 0$, and $\beta = \alpha = \eta$. The crowd behavior is in stable dynamic equilibrium and the infection among individuals is weak and will eventually disappear, which indicate the three parts that influence of others, result of crowd behavior and initiative infection will not show the cumulative amplification and increase twofold. It may exist infection among individuals in the crowd at first, but the result of group behavior would not change much so the infection is weaker and will disappear in the end.

(2) If $\gamma = \gamma_1 + \gamma_2 + \gamma_3 > 1$ it will have accumulated infection among individuals, because $p = 0$ is unstable. That is, the infection rate will be higher with the growth of γ. The crowd dynamic equilibrium is unstable at this moment, and the three parts will show cumulative amplification. Crowd behavior would change a lot since partial individuals imitate others. Then the other individuals are inspired to imitate each other and the mutual imitation will spread in crowd.

Therefore, the superposition of three factors that influence of others, result of crowd behavior and initiative infection is the basic condition for the infectious in crowd.

3.2 Conditions for Emergency Crowd Behavior

We set α' as the number of individuals infected by each infectious person per unit time. Then Eq. (1) can be described as

$$S'(t) = -\alpha' S(t) I(t) + \beta I(t) \tag{12}$$

And we have the followings

$$\begin{cases} S'(t) = -\alpha' S(t) I(t) + \beta I(t) \\ S(t) + I(t) = 2N \\ S(0) = S_0 \end{cases} \tag{13}$$

According to equations (13), we can get

$$S'(t) = \alpha'(2N - S)\left(\frac{\beta}{\alpha'} - S\right) \tag{14}$$

We suppose $\psi = \dfrac{\beta}{\alpha'}$. Eq. (14) will have a unique balance point $S = 2N$ when $\psi \geq 2N$. For any given solution belonged to $(0, 2N]$, $S(t)$ increases monotonously and tends to $2N$ while $I(t)$ decreases monotonously and tends to zero, which shows that it will not form emergency crowd behavior. When $\psi < 2N$, $S(t)$ will tend to ψ while $I(t)$ tends to $2N-\psi$ with t. Then emergency crowd behavior will be formed in this case. Therefore, $\frac{\beta}{2\alpha' N} = 1$ is the threshold of emergency crowd behavior.

4 Preventive Measures for Emergency Crowd Behavior

With the model analysis, we know that $\gamma = 1$ and $\frac{\beta}{2\alpha' N} = 1$ respectively are the thresholds of infection and emergency crowd behavior. To do well the preventive work of crowd behavior in emergencies, the followings are suggested.

(1) Lowering the infection capability γ. This value is directly related to the infection degree, and it will lead to emergency crowd behavior when the degree reaches a certain level. We know that the infection capability is affected by influence of others, result of crowd behavior and initiative infection. According to the relationship in Fig.1, lowering the infection capability should start with reducing the influence of others as well as initiative infection.
The corresponding measures are suggested as followings: One is to do well psychological propaganda and popularize psychological knowledge through a variety of ways to improve the psychological quality. And it should also consciously exercise the crowd's psychological sensitivity facing pressures as well as do psychological counseling work in time to avoid bad mood. The other one is to strengthen the learning of counter measures for emergencies and do necessary emergency drill to exercise crowd's reaction and coping ability.

(2) Reducing the infection rate α. Due to emergencies after the occurrence of unexpected events, the crowd easily generates blind conformity behavior, and either emotional infection or language infection easily leads to crowd behavior. The main reason for the crowd blind conformity is that the crowd can't obtain the accurate information timely and exactly. Therefore, the officials should investigate the nature and sources of unexpected events in time and determine response measures. At the same time, they should public relative information through effective way to stabilize the crowd's mentality and to keep their sense of trust in government.

(3) Improving the cure rate β. The average infection period can be expressed as $1/\beta$. The improvement of cure rate will greatly reduce the infection cycle and the number of crowd infected. After the occurrence of emergencies, on one hand the officials should immediately take emergency measures to minimize contact among individuals. On the other hand, they should timely make appropriate psychological counseling to eliminate the panic and to avoid long-term harm to people's mental caused by emergencies.

5 Conclusions

This paper proposes a dynamic model to reflect the average state of emergency crowd behavior combined with characteristics of crowd behavior. The result shows that infection capability decides the infection behavior among individuals, while infection rate and cure rate are the keys to the emergency crowd behavior by the dynamic analysis of model. Under the development process of emergency crowd behavior revealed by the model, preventive measures are suggested at last, namely, from the key aspects of infection capability, infection rate and cure rate, it respectively gives the corresponding measures, which is significant in practical.

There are some limitations in this paper. The crowd is divided into susceptible crowd and infectious crowd without considering the recovery crowd as well as embryonic crowd. Therefore, taking the two crowds into consideration for the model is worthy of further study.

References

1. Lindell, M.K.: EMBLEM2: An empirically based large scale evacuation time estimate model. Transp. Res. 42(1), 140–154 (2008)
2. Ren, C., Yang, C., Jin, S.: A multi-agent simulation of escape panic. In: CSO 2009, vol. 1, pp. 383–385 (2009)
3. Oguz, O., AkaydIn, A., YIlmaz, T., et al.: Emergency crowd simulation for outdoor environments. Comput. Graph 34(2), 136–144 (2010)
4. Wendt, J., Qu, G., Niu, J.: Modeling the behavior of crowds in panicked conditions. Act Media Technol. 6890, 288–299 (2011)
5. Briano, E., Mosca, R., Revetria, R., Testa, A.: A Simplified Human Cognitive Approach for Supporting Crowd Modeling in Tunnel Fires Emergency Simulation. In: Mehrotra, K.G., Mohan, C.K., Oh, J.C., Varshney, P.K., Ali, M. (eds.) IEA/AIE 2011, Part II. LNCS, vol. 6704, pp. 58–67. Springer, Heidelberg (2011)
6. Capote, J.A., Alvear, D., Abreu, O., et al.: Analysis of evacuation procedures in high speed trains fires. Fire Saf. J. 49(1), 35–46 (2012)
7. Liu, X., Ding, Z.: ESCAPE: A channel evacuation protocol for spectrum-agile network. In: 2007 2nd IEEE International Symposium on New Frontier in Dynamic Spectrum Access Network, pp. 292–302 (2007)
8. Cheng, N.W.: An optimization method for dynamic evacuation route programming based on improving ant colony algorithm. In: Proceedings 2010 International Conference on ISDEA 2010, vol. 1, pp. 265–267 (2010)
9. Hao, S.Q., Huang, H.W., Yuan, Y.: Fire evacuation of underground tunnel based building EXODUS. In: 2009 WRI World Congress on Computer Science and Information Engineering, vol. 7, pp. 612–615 (2009)
10. Pérez, V.F., Salmerón, J., Wood, K.: Dynamic evacuation routes for personnel on a naval ship. Nav. Res. Logis 55(8), 785–799 (2008)
11. Lux, T.: Herd behavior, bubbles and crashes. Econ. J., 881–896 (1995)
12. Yu, H.: A history of western sociological thought. Fudan University, Shanghai (1993) (in Chinese)

Strategies for Promoting Tourism Competitiveness Using a Hybrid MCDM Model

Kua-Hsin Peng and Gwo-Hshiung Tzeng

Abstract. Numerous studies have focused on measuring tourism destination competitiveness (TDC). However, few studies have attempted to identify strategies for improving TDC, preventing decision makers from obtaining valuable cues for making accurate decisions to improve competitiveness. This study thus explores strategies for improving tourism competitiveness using a new hybrid MCDM model combined with DANP (DEMATEL-based ANP). An empirical case is presented to demonstrate the effectiveness of a new hybrid MCDM model combining DANP and VIKOR for evaluating tourism competitiveness to identify competitiveness gaps and explore strategies for improving tourism competitiveness based on the influential relation map. Decision makers should increase the priority of the cause criteria in advance, to successfully create a high TDC to achieve the aspired/desired levels.

Keywords: Tourism destination competitiveness (TDC), improvement strategy, MCDM (multiple criteria decision making), DEMATEL, DANP (DEMATEL-based ANP), VIKOR.

1 Introduction

Tourism has undoubtedly become the leading leisure activity [1]. UNWTO [2] reported that international tourist arrivals reached 935 million globally in 2010, increasing 6.7% from 2009, with positive growth reported in all regions of the world. International tourism has recovered faster than expected from the global financial crisis and economic recession of late 2008 and 2009. Consequently, the

Kua-Hsin Peng · Gwo-Hshiung Tzeng
Institute of Project Management, Kainan University, No. 1, Kainan Road, Luchu, Taoyuan 338, Taiwan
e-mail: jenny19830514@yahoo.com.tw

Gwo-Hshiung Tzeng
Institute of Management of Technology, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu 300, Taiwan
e-mail: ghtzeng@mail.knu.edu.tw, ghtzeng@cc.nctu.edu.tw

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 107–115.
springerlink.com

tourism industry has contributed significantly towards economic growth. The recent literature is increasingly focused on tourism industry and its related issues, especially competitiveness issue.

Hovinen [3] indicated that the success of tourism destinations depends on their competitiveness. According to other researchers, destination competitiveness is associated with the economic prosperity of the residents of a country [4], [5]. Consequently, the recent literature is increasingly focused on measuring tourism destination competitiveness (TDC) [6], [7]. For example, Zhang *et al.* [8] used TOPSIS and information entropy to assess TDC in the Yangtze River Delta of China. Moreover, Kozak and Rimmington [9] presented a method of establishing the competitiveness criteria of international tourist destinations based on both quantitative and qualitative data. Furthermore, Cracolici and Nijkamp [10] used various multidimensional statistical techniques to estimate the competitive attractiveness of Southern Italy. Additionally, the World Economic Forum [11] presented the global Travel & Tourism Competitiveness Index (TTCI). However, few studies have explored strategies for improving TDC, and preventing decision makers from obtaining valuable cues for making accurate decisions to improve TDC.

A hybrid MCDM method has been developed and is widely used in numerous fields. Ou Yang *et al.* [12] combined DEMATEL and ANP to solve the dependence and feedback problems to suit the real world. Furthermore, Hung *et al.* [13] used the hybrid MCDM model to solve this knowledge management systems adoption problem. Additionally, Kuan *et al.* [14] used the hybrid MCDM model to assess the total performance of the new product development (NPD) process. Additionally, Yang and Tzeng [15] demonstrated how the DEMATEL method clarified the direct/indirect influential relationship of criteria. Based on the above discussion, this study attempts to explore the strategy for improving TDC using a new hybrid MCDM model that is combined with DANP. An empirical case is also presented to demonstrate the effectiveness of a new hybrid MCDM model combining DANP, and VIKOR is used to assess TDC to identify the competitiveness gaps and explore strategies for improving TDC based on the influential relation map.

2 Methodology

A list of criteria that can enhance TDC was gathered based on a report dealing with tourism competitiveness from World Economic Forum in 2009. These dimensions and criteria included: (1) regulatory framework(D_1)-policy rules and regulations(C_1), environmental sustainability(C_2), safety and security(C_3), health and hygiene(C_4), prioritization of Travel & Tourism(C_5); (2) business environment and infrastructure(D_2)-air transport infrastructure(C_6), ground transport infrastructure(C_7), tourism infrastructure(C_8), Information and Communication Technology (ICT) infrastructure(c_9), price competitiveness(C_{10});

and (3) human, cultural, and natural resources(D_3)-human resources(C_{11}), affinity for Travel & Tourism(C_{12}), natural resources(C_{13}), cultural resources(C_{14}).

The survey targeted experts or scholars engaged in travel and tourism areas. First, this study used a four-point scale ranging from 0 (no influence) to 4 (very high influence) to identify the criteria and their influence on one another. Five experts were then asked to assess the influence of the criteria on one another, and the consensus rates of the dimensions and criteria were 96.89% and 96.71% (both exceeding 96% in confidence), respectively. Finally, this study gathered secondary data on tourism competitiveness from the tourism competitiveness report published in 2009. Furthermore, this study used VIKOR to assess TDC, identify the gaps in competitiveness, and explore the strategy for improving tourism competitiveness based on the influential relation map.

DEMATEL is an analytical technique for building a structural model. DEMATEL is mainly used to solve complex problems to clarify their essential nature. DEMATEL uses matrix of direct influence-relation and indirect influence-relation to calculate the cause and effect relationships involved in each element. This technique is widely used to solve various complex studies, and particularly to understand complex problem structures and provide viable problem-solving methods [16]; [17]; [18]; [19]; [20]; [21]; [22]. Additionally, DEMATEL is based on the concept of influential relation map, which can distinguish the direct/indirect influential relationship of the criteria, allowing decision-makers to identify the key criterion for developing strategies for improving TDC.

This study not only uses the DEMATEL technique to confirm the interaction (interdependence and feedback) among the various factors/criteria, but also seeks the most accurate influential weights. This study found that ANP can serve this purpose. This study used the basic concept of ANP [23], which eliminates the limitations of Analytic Hierarchy Process (AHP) and is applied to solve nonlinear and complex network relations. ANP is intended to solve interdependence and feedback problems of cluster criteria. This study thus applies the characteristics of influential weights ANP and combines them with DEMATEL (call DANP, DEMATEL-based ANP) to solve these kind of problems based on the basic concept of ANP. This approach yields more practical results.

Opricovic and Tzeng [24] proposed the compromise ranking method (VIKOR) as a suitable technique for implementation within MCDM ([25], [26], [27], [28], [24], [29]). VIKOR uses the class distance function [30] based on the concept of the Positive-ideal (or the Aspired level) solution and Negative-ideal (or the Worst level) solution and puts the results in order. For normalized class distance function it is better to be near the positive-ideal point (the aspired level) and far from the negative-ideal point (the worst value) for normalized class distance function ([31]; [32]).

3 An Empirical Case of Taiwan

This section presents an empirical case involving Taiwan to explore strategies for improving TDC based on a new hybrid MCDM model.

This study finds expert cognition and opinion in three dimensions, and also identifies the relationship between the extents of the impact, which is compared with other dimensions, as shown in Table 1. According to the total influential prominence $(r_i + d_j)$, "Regulatory framework (D_1)" is the highest impact of the strength of relation that means the most important influencing factors; additionally, "Human cultural and natural resources (D_3)" is all the factors that affects the least degree of other factors. According to the influential relation $(r_i - d_j)$, "Regulatory framework (D_1)" is the highest degree of impact relationship that affect other factors directly. Otherwise, "Business environment and infrastructure (D_2)" is the most vulnerable to impact that compare with other dimensions.

Table 2 lists the relationship between the extents of the direct or indirect impacts and compares them with other criteria. "Prioritization of Travel and Tourism (C_5)" is the most important consideration criteria; additionally, "Safety and security (C_3)" is the impact of all criteria in the least degree of other criteria. Furthermore, Table 2 shows that "Policy rules and regulations (C_1)" is the highest degree of impact relationship in all the criteria. Otherwise, "Tourism infrastructure (C_8)" is the most vulnerable to impact of criteria that compare with other criteria.

This study builds the assessment model using DEMATEL, which is combined with the DANP (DEMATEL-based ANP) model to obtain the influential weights of each criterion, as shown in Table 2. Furthermore, the influential weights combine with the DEMATEL model to assess the priority of problem-solving based on the competitiveness gaps identified by VIKOR and the influential relation map.

Table 1 Total influential matrix of $\boldsymbol{T}$ and the sum of the effects on the dimensions

Dimensions		D_1	D_2	D_3	r_i	d_i	$r_i + d_i$	$r_i - d_i$
D_1	Regulatory framework	0.305	0.825	0.782	1.912	0.916	2.828	0.996
D_2	Business environment and infrastructure	0.321	0.237	0.332	0.891	1.497	2.388	-0.606
D_3	Human cultural and natural resources	0.290	0.435	0.208	0.932	1.322	2.254	-0.389

Note: average gap = $\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\left|g_{ij}^{n} - g_{ij}^{n-1}\right|}{g_{ij}^{n}}\times 100\%$ =3.11% < 5%, n denotes the samples of 5 experts and the consensus rate is 96.89%.

Table 2 The sum of the effects, weights and rankings of each criterion

Criteria	r_i	d_j	$r_i + d_j$	$r_i - d_j$	Degree of importance (Global weight)	Ranking
D_1					0.2866	3
C_1	1.750	0.882	2.633	0.868	0.0544	3
C_2	0.865	0.933	1.798	-0.068	0.0546	2
C_3	0.716	0.846	1.562	-0.131	0.0500	5
C_4	0.764	0.886	1.651	-0.122	0.0537	4
C_5	1.857	1.192	3.048	0.665	0.0739	1
D_2					0.3803	1
C_6	0.726	0.935	1.661	-0.209	0.0744	3
C_7	0.735	0.936	1.670	-0.201	0.0739	4
C_8	0.754	1.020	1.774	-0.266	0.0809	1
C_9	0.734	0.884	1.618	-0.150	0.0717	5
C_{10}	0.690	1.014	1.704	-0.325	0.0794	2
D_3					0.3332	2
C_{11}	1.103	0.778	1.881	0.325	0.0769	4
C_{12}	0.729	0.930	1.659	-0.202	0.0837	3
C_{13}	0.884	0.896	1.780	-0.013	0.0841	2
C_{14}	0.803	0.977	1.781	-0.174	0.0885	1

A real case involving Taiwan is used to assess the total competitiveness using the VIKOR method, as listed in Table 3. The scores of each criterion and the total average gap (S_k) of Taiwan are obtained, using the relative influential weights from DANP to multiply the gap (r_{kj}). Consequently, this study obtains the total performance gap of Taiwan based on the scoring value. The DEMATEL technique (Fig. 1) can obtain valuable cues for making accurate decisions. This system structure model illustrates that Taiwan suffers significant gap in the "Human cultural and natural resources (D_3)" dimensions, making it necessary to pay significant attention to the "Regulatory framework (D_1)" dimensions for improving TDC of Taiwan. Furthermore, for improving the regulatory framework (D_1) dimension, this study finds that the criterion of "Health and hygiene (C_4)" prioritizes improving the maximal competitiveness gap. Figure 1 shows that the criteria of "Policy rules and regulations (C_1)", "Prioritization of Travel & Tourism (C_5)" and Environmental sustainability (C_2) are the most important and influential criteria because they are most closely related to other criteria in the (D_1) dimension.

Table 3 The performance evaluation of the case study by VIKOR

Dimensions / Criteria	Local weight	Global weight (by DANP)	Case study of Taiwan	
			Score	Gap (r_{kj})
D_1	**0.2866(3)**		**4.40**	**0.433**
C_1	0.1898	0.0544(3)	4.80	0.367
C_2	0.1905	0.0546(2)	4.20	0.467
C_3	0.1745	0.0500(5)	5.50	0.250
C_4	0.1874	0.0537(4)	3.30	0.617
C_5	0.2579	0.0739(1)	4.20	0.467
D_2	**0.3803(1)**		**4.90**	**0.357**
C_6	0.1956	0.0744(3)	3.80	0.533
C_7	0.1943	0.0739(4)	5.70	0.217
C_8	0.2127	0.0809(1)	4.40	0.433
C_9	0.1885	0.0717(5)	5.30	0.283
C_{10}	0.2088	0.0794(2)	5.10	0.317
D_3	**0.3332(2)**		**3.90**	**0.517**
C_{11}	0.2308	0.0769(4)	5.70	0.217
C_{12}	0.2512	0.0837(3)	4.60	0.400
C_{13}	0.2524	0.0841(2)	2.40	0.767
C_{14}	0.2656	0.0885(1)	2.90	0.683
Total performances			**4.40**	**-**
Total gap (S_k)			**-**	**0.437**

Additionally, for improving the human cultural and natural resources (D_3) dimension, this study finds that the criterion of "Natural resources (C_{13})" is the maximal performance gap. Furthermore, the criteria of "Human resources (C_{11})" is the most important and influential criteria, and thus can be considered the critical criteria for improving natural resources. Thus, the criteria of "Human resources (C_{11})" can be considered the critical criterion for improving the regulatory framework. Additionally, the comprehensive indicator (R_k) can be obtained, which value of v can make decisions by the expert that is defined as $v = 0.5$ in this paper. This study obtains the result of the comprehensive indicator (R_k) as 0.602, representing that the Taiwanese government must improve the gap of TDC. Furthermore, the government can find the problem-solving points according to the DEMATEL technique combined with DANP and VIKOR (called the hybrid MCDM model).

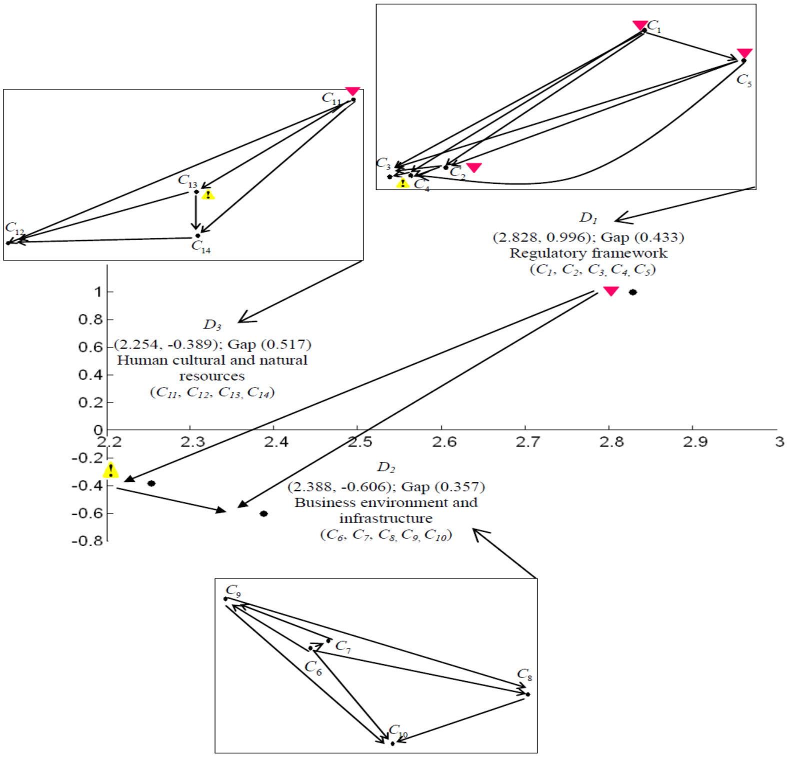


Fig. 1 The influential relation map of each dimension and criteria

4 Conclusions

This study can help decision-making to improve TDC. Furthermore, this study uses the DEMATEL method to develop network relationships, then, calculates the importance degree using DANP. Finally, this study uses VIKOR to assess total competitiveness.

The human cultural and natural resources dimensions exhibit a significant gap, and the regulatory framework dimensions may need to be considered to improve TDC. Furthermore, to improve the regulatory framework dimension, this study finds the criterion of health and hygiene is the maximal gap. Therefore it is necessary to improve the priorities of the cause criteria, namely policy and regulations, environmental sustainability, and prioritization of travel. Additionally, to improve the human, cultural and natural resources dimension, this study finds that the criterion of natural resources exhibits the maximal competitiveness gap. Furthermore, it is necessary to consider the need to improve human resources, to enhance natural environment quality (natural resources). Based on the above, the government should increase its prioritization of the cause criteria, allowing it to successfully increase TDC to achieve the aspired levels.

References

1. Claver-Cortes, E., Molina-Azorin, J.F., Pereira-Moliner, J.: Competitiveness in mass tourism. Annals of Tourism Research 34(3), 727–745 (2007)
2. World Tourism Organization, UNWTO (2010), Tourism Highlights, http://www.unwto.org/facts/eng/pdf/highlights/UNWTO_Highlights10_en_HR.pdf
3. Hovinen, G.: Revising the destination lifecycle model. Annals of Tourism Research 29(1), 209–230 (2002)
4. Crouch, G.I., Ritchie, J.B.R.: Tourism, competitiveness, and societal prosperity. Journal of Business Research 44(3), 137–152 (1999)
5. Buhalis, D.: Marketing the competitive destination in the future. Tourism Management 21(1), 97–116 (2000)
6. Zhang, H., Gu, C.L., Gu, L.W., Zhang, Y.: The evaluation of tourism destination competitiveness by TOPSIS and information entropy-A case in the Yangtze River Delta of China. Tourism Management 32(2), 1–9 (2010)
7. Cracolici, M.F., Nijkamp, P.: The attractiveness and competitiveness of tourist destinations: A study of Southern Italian regions. Tourism Management 30(3), 336–344 (2008)
8. Zhang, H., Gu, C.L., Gu, L.W., Zhang, Y.: The evaluation of tourism destination competitiveness by TOPSIS and information entropy-A case in the Yangtze River Delta of China. Tourism Management 32(2), 1–9 (2010)
9. Kozak, M., Rimmington, M.: Measuring tourist destination competitiveness: conceptual considerations and empirical findings. International Journal of Hospitality Management 18(3), 273–283 (1999)
10. Cracolici, M.F., Nijkamp, P.: The attractiveness and competitiveness of tourist destinations: A study of Southern Italian regions. Tourism Management 30(3), 336–344 (2008)
11. World Economic Forum (WEF), The Travel & Tourism Competitiveness Report (2009)
12. Ou Yang, Y.P., Shieh, H.M., Leu, J.D., Tzeng, G.H.: A novel hybrid MCDM model combined with DEMATEL and ANP with applications. International Journal of Operations Research 5(3), 1–9 (2008)
13. Hung, Y.H., Chou, S.C.T., Tzeng, G.H.: Knowledge management adoption and assessment for SMEs by a novel MCDM approach. Decision Support Systems (2011), (forthcoming) (available online February 5, 2011)
14. Kuan, M.J., Hsiang, C.C., Tzeng, G.H.: Probing the innovative quality system for NPD process based on combining DANP with MCDM model. International Journal of Innovative Computing. Information and Control (special issue) (2011), (forthcoming)
15. Yang, J.L., Tzeng, G.H.: An integrated MCDM technique combined with DEMATEL for a novel cluster-weighted with ANP method. Expert Systems with Applications 38(3), 1417–1424 (2011)
16. Chiu, Y.J., Chen, H.C., Shyu, J.Z., Tzeng, G.H.: Marketing strategy based on customer behavior for the LCD-TV. International Journal and Decision Making 7(2-3), 143–165 (2006)
17. Tzeng, G.H., Chiang, C.H., Li, C.W.: Evaluating intertwined effects in e-learning programs: A novel hybrid MCDM model based on factor analysis and DEMATEL. Expert Systems with Applications 32(4), 1028–1044 (2007)

18. Huang, C.Y., Shyu, J.Z., Tzeng, G.H.: Reconfiguring the innovation policy portfolios for Taiwan's SIP mall industry. Technovation 27(12), 744–765 (2007)
19. Chen, Y.C., Lien, H.P., Tzeng, G.H.: Measures and evaluation for environment watershed plans using a novel hybrid MCDM model. Expert Systems with Applications 37(2), 926–938 (2010)
20. Yang, J.L., Tzeng, G.H.: An integrated MCDM technique combined with DEMATEL for a novel cluster-weighted with ANP method. Expert Systems with Applications 38(3), 1417–1424 (2011)
21. Hung, Y.H., Chou, S.C.T., Tzeng, G.H.: Knowledge management adoption and assessment for SMEs by a novel MCDM approach. Decision Support Systems (2011), (forthcoming) (available online February 5, 2011)
22. Shen, Y.C., Lin, G.T.R., Tzeng, G.H.: Combined DEMATEL techniques with novel MCDM for the organic light emitting diode technology selection. Expert Systems with Applications 38(3), 1468–1481 (2011)
23. Saaty, T.L.: Decision Making with Dependence and Feedback: The Analytic Network Process. RWS Publications, Pittsburgh (1996)
24. Opricovic, S., Tzeng, G.H.: Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European Journal of Operational Research 156(2), 445–455 (2004)
25. Tzeng, G.H., Teng, M.H., Chen, J.J., Opricovic, S.: Multicriteria selection for a restaurant location in Taipei. International Journal of Hospitality Management 21(2), 171–187 (2002)
26. Tzeng, G.H., Lin, C.W., Opricovic, S.: Multi-criteria analysis of alternative-fuel buses for public transportation. Energy Policy 33(11), 1373–1383 (2005)
27. Opricovic, S., Tzeng, G.H.: Multicriteria planning of post-earthquake sustainable reconstruction. Computer-Aided Civil and Infrastructure Engineering 17(3), 211–220 (2002)
28. Opricovic, S., Tzeng, G.H.: Fuzzy multicriteria model for post-earthquake land-use planning. Natural Hazards Review 4(2), 59–64 (2003)
29. Opricovic, S., Tzeng, G.H.: Extended VIKOR method in comparison with outranking methods. European Journal of Operational Research 178(2), 514–529 (2007)
30. Yu, P.L.: A class of solutions for group decision problems. Management Science 19(8), 936–946 (1973)
31. Lee, W.S., Tzeng, G.H., Cheng, C.M.: Using novel MCDM Methods Based on Fama-French three-factor model for probing the stock selection. APIEMS, December 14-16, 1460–1474 (2009)
32. Ho, W.R.J., Tsai, C.L., Tzeng, G.H., Fang, S.K.: Combined DEMATEL technique with a novel MCDM model for exploring portfolio selection based on CAPM. Expert Systems with Applications 38(1), 16–25 (2011)

Symbolic Hierarchical Clustering for Pain Vector

Kotoe Katayama, Rui Yamaguchi, Seiya Imoto, Keiko Matsuura,
Kenji Watanabe, and Satoru Miyano

Abstract. We propose a hierarchical clustering in the framework of Symbolic Data Analysis(SDA). SDA was proposed by Diday at the end of the 1980s and is a new approach for analysing huge and complex data. In SDA, an observation is described by not only numerical values but also "higher-level units"; sets, intervals, distributions, etc. Most SDA works have dealt with only intervals as the descriptions. We already proposed "*pain distribution*" as new type data in SDA. In this paper, we define new "*pain vector*" as new type data in SDA and propose a hierarchical clustering for this new type data.

Keywords: Visual Analogue Scale, Distribution-Valued Data.

1 Introduction

Conventional data analysis usually can handle scalars, vectors and matrices. However, lately, some datasets have grown beyond the framework of conventional data analysis. Most statistical methods do not have sufficient power to analyse these datasets. In this study, we attempted to extract useful information from such datasets.

Symbolic data analysis (SDA) proposed by Diday [3] is an approach for analysing new types of datasets. "Symbolic data" consist of a *concept* that is described by intervals, distributions, etc. as well as by numerical values. The use of SDA enriches data description, and it can handle highly complex datasets. This implies that complex data can be formally handled in the framework of SDA. However, most SDA works have dealt with only intervals as the descriptions and are very few studies

Kotoe Katayama · Rui Yamaguchi · Seiya Imoto · Satoru Miyano
Human Genome Center, Institute of Medical Science, The University of Tokyo,
4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
e-mail: k-kata@ims.u-tokyo.ac.jp

Keiko Matsuura · Kenji Watanabe
Center for Kampo Medicine, Keio University School of Medicine, 35 Shinano-machi,
Shinjuku-ku, Tokyo 160-8582, Japan

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 117–124.
springerlink.com

based on this simple idea. The case that *concept* is described by intervals is simple, but ignores detailed information in the intervals. We already proposed a hierarchical clustering for the visual analogue scale (VAS) in the framework of Symbolic Data Analysis (SDA)[8].

In this paper, we define "*pain vector*" and propose a hierarchical clustering for this vector. The "*pain vector*" is consist of distributions and categories.

2 Transform the Visual Analogue Scale into "*Patient Distribution*"

We already proposed a hierarchical clustering for the visual analogue scale (VAS) in the framework of Symbolic Data Analysis (SDA)[8]. In the paper, we transformed the Visual Analogue Scale into distribution valued data. The VAS is a method that can be readily understood by most people to measure a characteristic or attitude that cannot be directly measured. VAS is of most value when looking at change within a same people, and is of less value for comparing across a group of people because they have different sense. It could be argued that a VAS is trying to produce interval/ratio data out of subjective values that are at best ordinal. Thus, some caution is required in handling VAS. We described VAS as distribution and handle it as new type data in SDA.

2.1 The Visual Analogue Scale

The visual analogue scale (VAS) has developed to allow the measurement of individual's responses to physical stimuli, such as heat. The VAS is a method that can be readily understood by most people to measure a characteristic or attitude that cannot be directly measured. It was originally used in the field of psychometrics, and nowadays widely used to assess changes in patient health status with treatment.

A VAS consists of a line on a page with clearly defined end points, and normally a clearly identified scale between the two end points. For guidance, the phrase "no pain" and "worst imaginable pain" are placed at the both side of the line, respectively. Minimum value 0 of the VAS means "no pain" and maximum value 100 means "worst imaginable pain".

These scales are of most value when looking at change within patients, and are of less value for comparing across a group of patients because patient have a different sense of pain. It could be argued that a VAS is trying to produce interval/ratio data out of subjective values that are at best ordinal. Thus, some caution is required in handling such data. Many researchers prefer to use a method of analysis that is based on the rank ordering of scores rather than their exact values, to avoid reading too much into the precise VAS score.

2.2 *Transform the Visual Analogue Scale to Distribution-Valued Data*

We transform the VAS to distribution-valued data to compare across a group of patients . VAS varies according to patients, because sense of pain varies a great deal depending on people. Changing VAS score within patients means their sense of pain. If they have big change of VAS score, their expression of sense of pain is rough. On the contrary, if they have small change, their expression is sensitive. We suggest that these sense of pain is described by normal distribution and call it "*pain distribution*(PD)".

Let VAS score of patient's first time be x_1 and second time be x_2. We define the middle point of x_1 and x_2 as mean of PD μ, and $(\mu - x_1)^2 = (\mu - x_2)^2$ as variance. We describe PD as $N(\mu, \sigma^2)$. In case that the number of VAS score is d, PD is d-dimensional normal distribution. In this case, a diagonal matrix is used as a variance-covariance matrix of d-dimensional normal distribution.

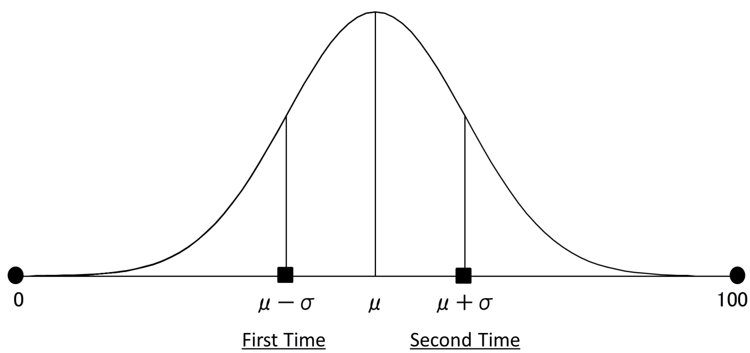


Fig. 1 Transform the Visual Analogue Scale to Distribution-Valued data

3 Medical Questionnaire in Keio University School of Medicine

Center for Kampo Medicine, Keio University School of Medicine, have a questionnaire to patients to help medical decision. The questionnaire includes one set of questions about their subjective symptoms. There are 244 yes-no questions, for example, "Are you constipate?", and 118 visual analogue scale questions, for example, "How do you feel pain with urination?". Patients answer these questions every time when they come to Keio University. Doctors can understand patients' fluctuate in severity.

4 *Pain Vector*

The Medical Questionnaire in Keio University School of Medicine is consist of "*Yes-No*" questions and "*VAS*" questions. To compare "*VAS*" questions among patients, we already proposed “*pain distribution*(PD)”. In the case, we only use ”*VAS*” questions and didn’t use ”*Yes-No*” questions.

Now we propose *Pain Vector* by using both questions. Let ”*Yes-No*” question be YN_a $(a = 1,\ldots,A)$. If patient answer ”Yes”, $YN_a = 1$, otherwise 0. Let ”*VAS*” question be PD_b $(b = 1,\ldots,B)$. The i-th patient’s *Pain Vector* is

$$\boldsymbol{PV}_i = [YN_{i1},\ldots,YN_{iA},PD_{i1},\ldots,PD_{iB}]'.$$

5 Hierarchical Clustering for Pain Vector

Cluster analysis groups data objects only on the bases of information found in the data that describes the objects and their relationships. The goal is that the objects within a group should be similar (or related) to one another and different from the objects in other groups.

In this section, we propose a hierarchical clustering for *Pain Vector*.

5.1 The Clustering Algorithm

We extend the idea of a hierarchical clustering in the framework of conventional data analysis. Let n be the number of PV and K be the number of cluster.

<Step1> Begin with K clusters, each containing only a single PV, $K = n$. Calculate distance between PV.
<Step2> Search the minimum distance in K clusters. Let the pair the selected clusters. Combine PVs into a new cluster, it is described by new Vector. Let K be $K-1$. If $K > 1$, go to Step3, otherwise Step4.
<Step3> Calculate the distance between new cluster and other cluster, and go back to Step2.
<Step4> Draw the dendrogram.

5.2 Distance between PVs

In our method, PVs consist of binary and distribution valued data. Let i-th patient’s vector be $\boldsymbol{PV}_i$ and j-th be $\boldsymbol{PV}_j$.

$$\boldsymbol{PV}_i = [\boldsymbol{YN}'_i,\boldsymbol{PD}'_i]' = [YN_{i1},\ldots,YN_{iA},PD_{i1},\ldots,PD_{iB}]',$$
$$\boldsymbol{PV}_j = [\boldsymbol{YN}'_j,\boldsymbol{PD}'_j]' = [YN_{j1},\ldots,YN_{jA},PD_{j1},\ldots,PD_{jB}]'.$$

The sum of distance between $\boldsymbol{YN}$ and distance between $\boldsymbol{PD}$ is distance between $\boldsymbol{PV}_i$ and $\boldsymbol{PV}_j$.

5.2.1 Distance between YNs

YNs are binary data. We define distance between YN_{ia} and YN_{ja} as

$$||YN_{ia} - YN_{ja}||.$$

Distance between $\boldsymbol{YN}_i$ and $\boldsymbol{YN}_j$ is

$$\sum_{a=1}^{A} ||YN_{ia} - YN_{ja}||.$$

5.2.2 Distance between PDs

We use symmetric KL-divergence as distance between PDs. Kullback-Leibler divergence is the natural way to define a distance measure between probability distributions [9], but not symmetry. We would like to use the symmetric Kullback-Leibler (symmetric KL) divergence as distance. The symmetric KL-divergence between two distributions s_1 and s_2 is

$$\begin{aligned} D(s_1(\boldsymbol{x}), s_2(\boldsymbol{x})) &= D(s_1(\boldsymbol{x})||s_2(\boldsymbol{x})) + D(s_2(\boldsymbol{x})||s_1(\boldsymbol{x})) \\ &= \int_{-\infty}^{\infty} s_1(\boldsymbol{x}) \log \frac{s_1(\boldsymbol{x})}{s_2(\boldsymbol{x})} d\boldsymbol{x} + \int_{-\infty}^{\infty} s_2(\boldsymbol{x}) \log \frac{s_2(\boldsymbol{x})}{s_1(\boldsymbol{x})} d\boldsymbol{x}, \end{aligned} \tag{1}$$

where $D(s_1||s_2)$ is KL divergence from s_1 to s_2 and $D(s_2||s_1)$ is one from s_2 to s_1.

Let PDs be d dimensional $N(\boldsymbol{\mu}_{ib}, \boldsymbol{\Sigma}_{ib})$ and $N(\boldsymbol{\mu}_{jb}, \boldsymbol{\Sigma}_{jb})$. Symmetric KL-divergence is

$$\begin{aligned} &D(p(\boldsymbol{x}|\boldsymbol{\mu}_{ib}, \boldsymbol{\Sigma}_{ib}), p(\boldsymbol{x}|\boldsymbol{\mu}_{jb}, \boldsymbol{\Sigma}_{jb})) \\ &\quad = tr(\boldsymbol{\Sigma}_{ib}\boldsymbol{\Sigma}_{jb}^{-1}) + tr(\boldsymbol{\Sigma}_{jb}\boldsymbol{\Sigma}_{ib}^{-1}) \\ &\quad\quad + tr((\boldsymbol{\Sigma}_{ib}^{-1} + \boldsymbol{\Sigma}_{jb}^{-1})(\boldsymbol{\mu}_{ib} - \boldsymbol{\mu}_{jb})(\boldsymbol{\mu}_{ib} - \boldsymbol{\mu}_{jb})^T) - 2d. \end{aligned} \tag{2}$$

Let PDs be $d = 1$,

$$\begin{aligned} &D(p(x|\mu_{ib}, \sigma_{ib}), p(x|\mu_{jb}, \sigma_{jb})) \\ &= \frac{1}{2}\left\{\log\frac{\sigma_{jb}^2}{\sigma_{ib}^2} + \frac{\sigma_{ib}^2 + (\mu_{ib} - \mu_{jb})^2}{\sigma_{jb}^2}\right\} + \frac{1}{2}\left\{\log\frac{\sigma_{ib}^2}{\sigma_{jb}^2} + \frac{\sigma_{jb}^2 + (\mu_{jb} - \mu_{ib})^2}{\sigma_{ib}^2}\right\} - 1. \end{aligned} \tag{3}$$

Distance between $\boldsymbol{PD}_i$ and $\boldsymbol{PD}_j$ is

$$\sum_{b=1}^{B} D(p(x|\mu_{ib}, \sigma_{ib}), p(x|\mu_{jb}, \sigma_{jb})).$$

5.3 *New PV in Clustering Algorithm Step 2 and Their Distance*

In Clustering Algorithm Step 2, we combine $\boldsymbol{PV}_i$ and $\boldsymbol{PV}_j$ into a new cluster and it is described by new vector. This new vector, $\boldsymbol{NPV}$, is described by using distance between YN_{ia} and YN_{ja} and Gaussian mixture distributions of PD_{ib} and PD_{jb}.

$$\boldsymbol{NPV} = [YN_{i1} - YN_{j1}, \ldots, YN_{iA} - YN_{jA}, mgd(PD_{i1}, PD_{j1}), \ldots, mgd(PD_{iB}, PD_{jB})]',$$

where $mgd(PD_{i1}, PD_{j1})$ means mixture distribution of PD_{i1} and PD_{j1}, and mixture weight equal 0.5.

After Section 5.1 Step2, we need symmetric KL-divergence between Gaussian mixture distributions. However, it cannot be analytically computed. We can use, instead, Monte-Carlo simulations to approximate the symmetric KL-divergence. The drawback of the Monte-Carlo techniques is the extensive computational cost and the slow converges properties. Furthermore, due to the stochastic nature of the Monte-Carlo method, the approximations of the distance could vary in different computations.

In this paper, we use unscented transform method proposed by Goldberger, *et al*[5].

We show approximation of $D(s_1||s_2)$ in (1). Let cluster c_1 contains d-dimensional distribution $N_d(\boldsymbol{\mu}_m^{(1)}, \boldsymbol{\Sigma}_m^{(1)})$, $(m = 1, \ldots M)$. Expression formula of c_1 is $s_1(\boldsymbol{x}) = \sum_{m=1}^{M} \omega_m^{(1)} p(\boldsymbol{x}|\boldsymbol{\theta}_m^{(1)})$, where $\omega_m^{(1)}$ is a mixture weight, $p(\boldsymbol{x}|\boldsymbol{\theta}_m^{(1)})$ is m-th probability density function of $N_d(\boldsymbol{\mu}_m^{(1)}, \boldsymbol{\Sigma}_m^{(1)})$ and $\boldsymbol{\theta}_m^{(1)} = (\boldsymbol{\mu}_m^{(1)}, \boldsymbol{\Sigma}_m^{(1)})$. Simmilary, cluster c_2 contains d-dimensional distribution $N_d(\mu_l^{(2)}, \Sigma_l^{(2)})(l = 1, \ldots L)$. Expression formula of c_2 is $s_2 = \sum_{l=1}^{L} \omega_n^{(2)} p(\boldsymbol{x}|\boldsymbol{\theta}_l^{(2)})$.

Approximation of KL-divergence from s_1 to s_2 by using unscented transform method is

$$D(s_1||s_2) \approx \frac{1}{2d} \sum_{m=1}^{M} \omega_m \sum_{k=1}^{2d} \log \frac{s_1(\boldsymbol{o}_{m,k})}{s_2(\boldsymbol{o}_{m,k})}, \tag{4}$$

where $\boldsymbol{o}_{m,t}$ are sigma points. They are chose as follows:

$$\begin{aligned} \boldsymbol{o}_{m,t} &= \boldsymbol{\mu}_m^{(1)} + \left(\sqrt{d\boldsymbol{\Sigma}_m^{(1)}}\right)_t, \\ \boldsymbol{o}_{m,t+d} &= \boldsymbol{\mu}_m^{(1)} - \left(\sqrt{d\boldsymbol{\Sigma}_m^{(1)}}\right)_t, \end{aligned} \tag{5}$$

such that $\left(\sqrt{\boldsymbol{\Sigma}_m^{(1)}}\right)_t$ is t-th column of the matrix square root of $\Sigma_m^{(1)}$. Then,

$$\begin{aligned} \boldsymbol{o}_{m,t} &= \boldsymbol{\mu}_m^{(1)} + \sqrt{d\lambda_{m,t}^{(1)}}\boldsymbol{u}_{m,t}^{(1)} \\ \boldsymbol{o}_{m,t+d} &= \boldsymbol{\mu}_m^{(1)} - \sqrt{d\lambda_{m,t}^{(1)}}\boldsymbol{u}_{m,t}^{(1)}, \end{aligned} \tag{6}$$

where $t = 1, \ldots, d$, $\boldsymbol{\mu}_m^{(1)}$ is mean vector of m-th normal distribution in s_1, $\lambda_{m,t}^{(1)}$ is t-th eigenvalue of $\boldsymbol{\Sigma}_m^{(1)}$ and $\boldsymbol{u}_{m,t}^{(1)}$ is t-th eigenvector. If $d = 1$, the sigma points are simply

$$\mu_m^{(1)} \pm \sigma_m^{(1)}.$$

We can calculate approximation of $D(s_2||s_1)$. Substituting these approximations into (1), we obtain the symmetric KL-divergence. We set the divergence as distance between PD in cluster c_1 and PD in c_2.

6 An Application

In this section, we apply our proposal method to real data from Keio University School of Medicine. This is masked data and is not be tied to any information that would identify a patient. For our analysis, we use the 2316 patients' result of medical questionnaire. There are 244 yes-no questions, and 118 visual analogue scale questions.

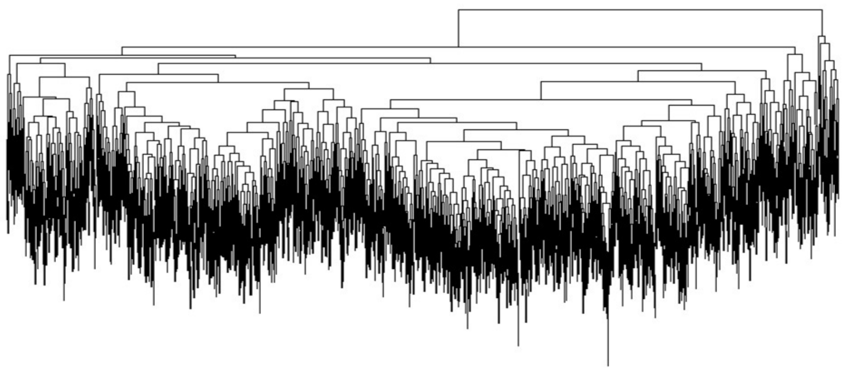


Fig. 2 Dendrogram for 2316 patients

7 Concluding Remarks

We already proposed a hierarchical clustering for the visual analogue scale (VAS) in the framework of Symbolic Data Analysis (SDA). To compare "*VAS*" questions among patients, we already proposed "*pain distribution*(PD)". In the case, we only used "*VAS*" questions and didn't use "*Yes-No*" questions. In this paper, we use both questions and define new "*pain vector*" as new type data in SDA. The "*pain vector* consist of binary and distribution valued data. We also propose a hierarchical clustering for this new type data. Through the simulation, we verified our model.

References

1. Billard, L., Diday, E.: Symbolic Data Analysis. Wiley, NewYork (2006)
2. Bock, H.-H., Diday, E.: Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data. Springer, Berlin (2000)

3. Diday, E.: The symbolic approach in clustering and related methods of Data Analysis. In: Bock, H. (ed.) Proc. IFCS Classification and Related Methods of Data Analysis, Aachen, Germany. North-Holland (1988)
4. Diday, E.: The symbolic approach in clustering and related methods of Data Analysis. In: Bock, H. (ed.) Classification and Related methods of Data Analysis, pp. 673–684. North-Holland, Amsterdam (1988)
5. Goldberger, J., Gordon, S., Greenspan, H.: An efficient image similarity measure based on approximations of KL-divergence between two gaussian mixtures. In: Proceedings of CVPR, pp. 487–494 (2006)
6. Gowda, K.C., Diday, E.: Symbolic clustering using a new dissimilarity measure. Pattern Recognition 24(6), 567–578 (1991)
7. Katayama, K., Suzukawa, A., Minami, H., Mizuta, M.: Linearly Restricted Principal Components in *k* Groups. In: Electronic proceedings of Knowledge Extraction and Modeling, Villa Orlandi, Island of Capri, Italy (2006)
8. Katayama, K., Yamaguchi, R., Imoto, S., Matsuura, K., Watanabe, K., Miyano, S.: Clustering for Visual Analogue Scale Data in Symbolic Data Analysis. Procedia Computer Science 6, 370–374 (2011)
9. Kullback, S.: Information theory and statisticsh. Dover Publications, New York (1968)

Temporal Perspective for the Software Testing Decision Support Framework

Masoud Mohammadian, Bala M. Balachandran, and Deane Larkman

Abstract. The development of the Decision Support Framework (DSF) for software testing provides a solid basis for assisting the software test manager in assessing the risk of achieving successful software testing. To further enhance the DSF, the DSF has been expanded beyond its static and dynamic analysis perspectives.

It has been observed that any activity or process has some type of inherent temporal perspective. Thus the adding of a temporal perspective to the DSF static and dynamic analysis perspectives is a step forward. The temporal information is derived and integrated utilizing the test manager's experience and his/her software test plan. This provides an additional analysis on successful software testing and the consequences of influences that affect software testing. An explanation of why the framework needs a time perspective and other associated changes is provided. The details of the temporal perspective and the other relevant changes to the framework are described. The total decision support framework, with an integrated temporal perspective, is presented and discussed.

1 Introduction

It has long been recognized that the software testing industry is and will continue to grow (Gelperin, D, and Hetzel, B. 1988). Software testing is still emerging into a healthy industry and an established discipline. Over recent years software testing has been examined from various perspectives – such as market size, growth potential, the cost of errors, and the number of testing staff employed. The global software testing market in 2006 was said to be worth US$13 billion (Amitysoft, 2006). The market opportunity for offshore software testing companies in India alone was estimated in 2006 to rise from US$2 billion to US$8 billion by 2008 (Amitysoft, 2006). The 2009 Ovum report projected that the worldwide market for

Masoud Mohammadian · Bala M. Balachandran · Deane Larkman
Faculty of Information Science and Engineering
University of Canberra
ACT, Australia
e-mail: {Bala.Balachandran,Masoud.Mohammadian}@Canberra.edu.au,
Deane.Larkman@uni.canberra.edu.au

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 125–135.
springerlink.com

software and systems testing services would reach US$56 billion by 2013, despite taking a hit from the global economic crisis (Physorg.com, 2009; Lesnikov, S. 2009). The Ovum report also predicted that software testing services will grow at a compound annual growth rate of 9.5% from 2008 to 2013. This is considered to be much faster than other information technology services (Physorg.com, 2009; Lesnikov, S. 2009).

Inadequate software testing infrastructure is estimated to cost the US economy an estimated US$59.5 billion annually (National Institute of Standards and Technology [NIST], 2002). This cost represented about 0.6% of the US`s US$10 trillion dollar GDP (NIST, 2002). Software developers accounted for about 40% of the total impacts on the testing infrastructure, whereas software users accounted for the about 60% (NIST, 2002).

Research conducted by Technavio indicates that the software testing services market is expected to grow at an annual growth rate of 11% per year from 2010 to 2014. "The software testing industry is witnessing significant changes both technology-wise as well as in the business process. Technology changes in testing have been continuous and are being driven by business needs such as cost reduction and improvement of quality and time-to-market. Further, evolving software technology is also compelling enterprises to implement testing right from the product development stage which is increasing the adoption of software testing services," reports Technavio analyst (Infiniti Research Limited, 2011).

2 Software Testing Decision Support Framework (DSF)

The DSF consists of three primary directly influence the goal of successful software testing. The primary elements are: Test Management, Test Information, and Test Environment. There is a secondary element is: Technical Support, that transitively influences the goal via the Test Environment.

The software test manager assumes the decision making role of filling in the information needed in influence weightings in the software testing decision support framework. He/She are expected to have an understanding of the type and extent of the software that is to be tested, and they expected to do element evaluation in accordance within the organisational context and the type of software to be tested. The software test manager needs to apply that information to the primary and secondary elements in the framework in arriving at the assigning of influence weights. Each of these is briefly discussed as follows.

Test Management considers which test approach, method, test tool or technique does the software test manager have experience in using, and can be used to support the type of software to be tested.

Test Information considers the test strategy that supports the test requirements and is an integral component in the testing process. An understanding of the basic elements of test design underpins the design of test cases; irrespective of the way those test cases are generated (manually or automatically).

The Test Environment is the physical and logical set up configured to support the full functionality of the software under test, and to support the tests to be executed on that software. This support includes any element required to ensure that the testing of the software provides accurate results, based on the software operating within a specific customer environment.

Technical Support is the underlying foundation necessary to develop and maintain the test environment. The efforts of technical support have a central focus; to minimise unforeseen side effects and disruptions to the testing process, while aiming to improve the overall quality, stability, and fitness for purpose of the software being tested. To construct a suitable test environment a set of technical support elements is needed. These elements entail relevant technical support for the test infrastructure – the test environment, testers' workstations and any facilities dedicated to testing.

Each element in the DSF consists of three or more supporting factors. The factor contribution percentages are as follows:

- for each factor, its contribution percentage is in the range of 0% to 100%;
- for each element, its total factor contributions percentage must initially be equal to 100%; and
- factor contribution percentages are derived by the software test manager as the best estimate for the given type of software to be tested.

The DSF model would look like the one in Figure 1 below before the influence weightings and factor contribution percentages have been added.

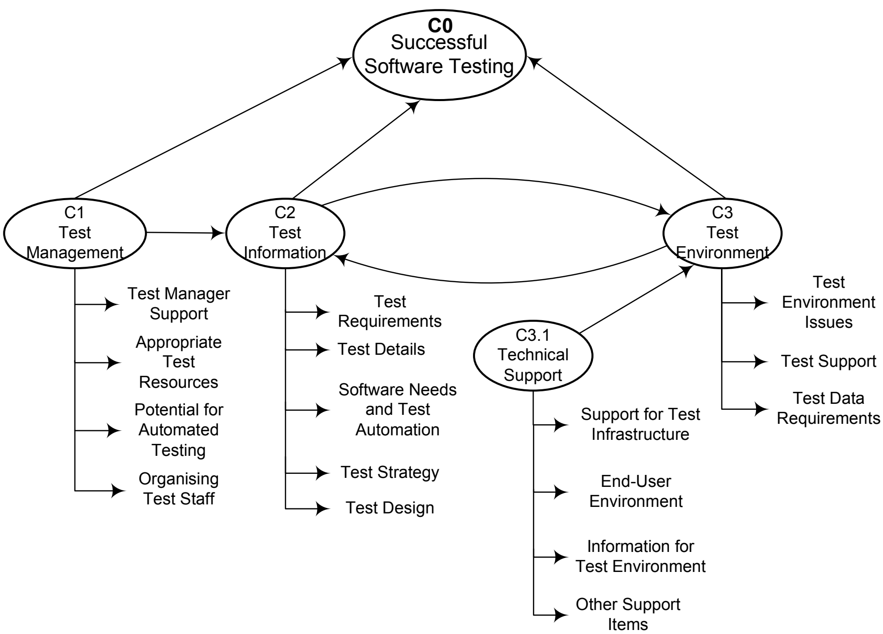


Fig. 1 Decision Support Framework Illustrated

Influence weightings and directional signed relationships work together. For each relationship there is assigned one and only one influence weighting. For each element a relationship is drawn to the goal or other elements it supports and influences. Influence weightings are stated as percentages. Influence weights are determined by the software test manager based on their knowledge of the type of software to be tested, underpinned by their software testing experience, and within the organisational context.

Real world software testing can be complex and complicated in order to include all the possible elements in testing software (Borysowich, 2005). Research into software testing needs to be directed and the research area needs to be well defined. The DSF was developed based on extensive research and is directed at assisting the software test manager. The concepts and the foundation upon which the Decision Support Framework was developed have been well researched and defined. A more detailed discussion of the DSF has been done and is not included in this paper (Larkman et. al., 2010a, Larkman et. al., 2010b, Larkman et. al., 2011).

3 Additional Framework Component

There are several other components that could be added to the DSF. However, one of the primary additional analysis components for the DSF is the inclusion of a temporal aspect. Temporal consequences are critical to successful software testing and the release of the software. This needs to be added to the framework to assist the test manager in the decision support activities before and during software testing. The term framework has been loosely applied to research that addresses situations of related issues that do constitute a framework but rather are supportive to the framework, as defined for the DSF (Larkman et. al., 2011).

3.1 Temporal

The purpose of the DSF is to assist software test managers in their planning and risk management for successful software testing before and during software testing (Larkman et al., 2010b). Two evaluative perspectives (static and dynamic) were identified that could be used to evaluate the DSF (Larkman et al., 2011; Larkman et al., 2010a). The two evaluative perspectives include several analytical techniques. The static techniques are:

- risk path analysis (which includes total path weight analysis and critical path analysis) (Larkman et al., 2010a); and
- event path analysis (unpublished analytical technique).

The dynamic techniques are:

- fuzzy cognitive map analysis (Larkman et al., 2010a); and
- factor contribution analysis (Larkman et al., 2011).

Software testing involves a time dimension – estimates of time for each software testing task are important for project success (Rani, 2007), and periods of use for each test resource (such as facilities, tools and staff) have to be specified (Institute of Electrical and Electronics Engineers, 2008). Decision making also involves a time dimension – relevant aspects of the time dimension are: the outcomes of a decision potentially involve time delays (Alter, 2002), and decisions must be made by a certain time (Harris, 2009). The previous information suggests that the DSF needs to include a time dimension for the software testing domain. However, the current structure of the DSF infers that the time dimension is implicitly within the framework elements. Therefore, this implied temporal implication has directed two questions. First: How can the implicit temporal perspective of the DSF be made explicit? And once the first question is addressed then: How can the temporal perspective be analysed?

The DSF is a useful tool in analysing and evaluating the risk on meeting successful software testing; often conditions provide a successful outcome at the expense of temporal consequences. To address temporal consequences, the DSF approach needs to explicitly incorporate the temporal perspective such that it can be integrated with other dynamic analytical techniques. The dynamic analytical techniques are given the focus herein because they are used to identify the consequences of change in the DSF, such as risk consequences that accompany changes in total factor contributions of one or more elements. The verb *change* is used in the sense of to become different. The DSF has an initial state, defined by the properties of its components (irrespective of whether the properties are base properties or derived properties). As the software testing situation changes, the DSF undergoes corresponding changes, from its current state to a new state, reflected by change to one or more of its properties. Change to the DSF establishes the basis for analysis of the consequences of that change.

By adding a temporal perspective to the DSF it will provide additional information for the software test manager and any subsequent organisation reporting. The DSF, enhanced with a temporal perspective, provides the software test manager with a more encompassing tool that can be used to both support risk management of successful software testing, and to analyse the temporal consequences on successful software testing. The organisation can identify the consequences of reduced support or the temporal effects of decisions on successful software testing.

4 Framework with a Temporal Perspective

The DSF model can be used to evaluate different types of software testing. Software testing has the potential to affect many aspects in the organisation, customers, management decisions, and other projects. To this end, a temporal perspective added to the DSF can provide a great deal of additional information to assist the test manager and the organisation.

4.1 Temporal Perspective

The test manager is ultimately responsible for the test plan. During the software testing life cycle, step one is planning and step two is design. Included in the planning and design steps is the development of a test plan. The test plan, in general (comes in many formats), may include such information as: Scope and Objectives, Testing Approach / Methodology, Test Series / Phases / Cycles, Resources, Test Schedule, Roles and Responsibilities, Bug Tracking, Reporting Requirements, and other information (Borysowich, 2005: Test Plan 2011).

In analysing the business and / or software testing requirements, resources needed and the test schedule work together in creating the estimated number of hours to conduct the software testing. The test plan includes some type of series of tests to be conducted. The software test manager reviews the software (or system) requirements specifications and determines that the following test series are to be done:

Table 1 Testing Series

Test Series	Description
100	System Administration Testing
200	Functional Testing
300	Load and Stress Testing
400	User Interface Testing
500	Security Testing
600	Other types of testing (such as, installation, online help and documentation…)

This list is not meant to be a totally encompassing series list. Each list and series would depend on the type of software is being tested, if it is new, an upgrade, or a new version, system requirements, customer requirements, and many other factors.

The software test manager, based on his/her information and experience, provides an estimate of the number of person hours it will take to successfully complete each test series. Generally the estimate includes the number of testers, approximate number of hours each will spend each day doing the tests, how many days each will be involved in doing the testing for the particular software, test environment setup and operation, and reporting. Then the organisation can determine when the software can be released.

It is not unusual for a release date to be set and software testing then develops a schedule to fit within that release date. In many circumstances the test manger thus has to work backwards. The organisation or customer, sets a due date for the release of the software and then the software test manager has to determine how long all testing will take in order to meet any milestones and final release deadline.

This example is based on an ideal situation, i.e. the test manger makes his/her estimate and the organisation then computes a release date for the software application.

Using the DSF model in Figure 4 the test manager comes up with the following estimate:

> 4 people x 5 hours / day, working 5 days a week, for 8 weeks = 800 hours
> (assume that only the information in Table 1 will be tested)

In other words, in order to achieve C0 (Successful software testing), it is estimated to take 800 hours.

In applying the temporal condition to the DSF, only primary elements and their influences weightings are considered. Thus C3.0 and C3.1 are considered together with only the 80% being applicable to the overall analysis.

The estimate considers the factors for each element shown in Figure 4. Thus:

Table 2 Normalised Weightings

	C1	**C2**	**C3**
Influence Weighting	75%	65%	80%
Normalised Percentage *	34%	29.5%	36.5%

* Normalised to 100%.

Thus:

Table 3 Element Hours Calculation

Element	**Calculation**	**Hours**
C1	34% * 800	272
C2	29.5% * 800	236
C3 w/C3.1	36.5% * 800	292

With the temporal considered, the completed DSF model now looks like Figure 2.

4.2 Temporal What-If Analysis

Let us begin with C1 and see what temporal affects a factor has on the overall successful software testing, i.e. the goal.

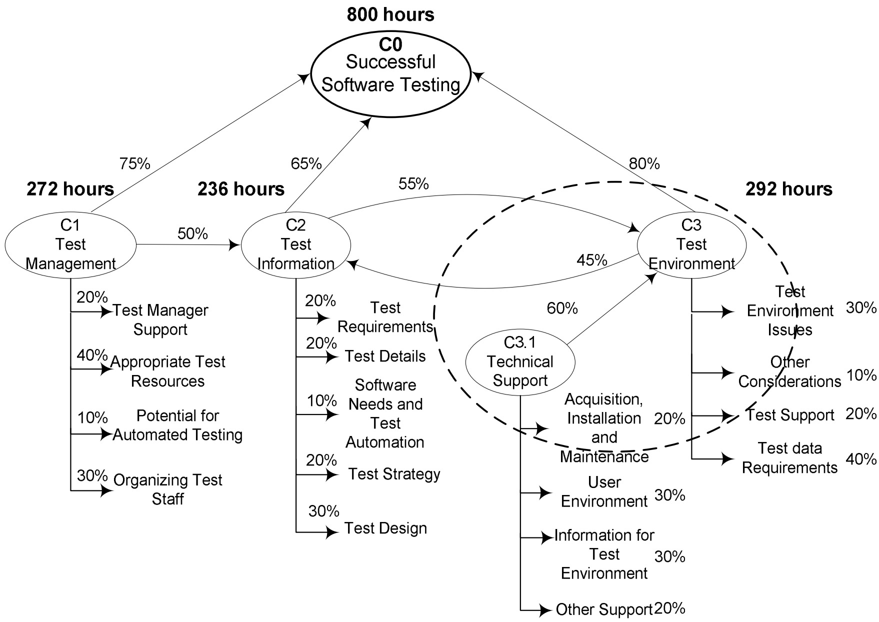


Fig. 2 Temporal DSF Model

4.2.1 Example 1

Something happens and the amount of test manager support goes from 20% to 15% - a reduction of 5%. The factors of C1 equal 272 hours. To achieve the continued influence weighting of 34% based on C1 → C0 something has to account for the 5% loss. 5% in this case is equal to 13.6 hours (.05 * 272). Thus 800 + 13.6 = 813.6 hours is needed to meet the goal due to the decrease in test manager support for test management.

Rounding up to the nearest time divisible by 5, (rounding up = 15 hours) 3 additional days are needed to achieve the goal. Another way of putting it, the goal of successful software testing will be delayed by 3 days. This is based on the original estimate of a person testing 5 hours per day.

There are a number of ways that the test manager can use to deal with the potential delay of 3 days. They could try to find additional resources, they could look at what other adjustments in other element factors could be used to offset or make up the delay, or they could apportion the delay to all elements.

For example, the number of hours distributed among the elements can be done in various ways. For example, each element could have an appropriately portioned number of increased hours as shown in Table 4.

Table 4 Delay Distribution

Element	Percentage	Est. Hours	Variance
C1	34%	276.5	4.5
C2	29.5%	240.0	4.0
C3 w/C3.1	36.5%	297.1	5.1

4.2.2 Example 2

Suppose that the technical support of setting up the user environment has some issues. These issues cause that factor to decrease from 30% to 15%. Since C3 & C3.1 together equal 292 hours of the 800, the factors for the two must be taken together. As follows:

Table 5 Delay Example 2

Factors	Before Normalisation	After Normalisation	With 15% decrease in support
Acqu & Maint.	20	10	10
User Environ.	30	15	7.5
Infor T E	30	15	15
Other Support	20	10	10
T E Issues	30	15	15
Other	10	5	5
Test Support	20	10	10
Test Data	10	20	20
Totals	**200**	**100**	**92.5**

A reduction in User Environment Technical support equates (.075 * 292) to 21.90 hours delay in the project. Again, rounding up 21.90 hours would be considered a 5 day delay. Remember this is based on a person testing 5 hours per day.

5 Decision Support Framework Software Testing Life Cycle

Figure 5 illustrates the evaluation process for the DSF. This process includes the Temporal Evaluation step as defined in this paper. Furthermore this process shown in Figure 5 now provides the flow between the software testing life cycle and the DSF steps.

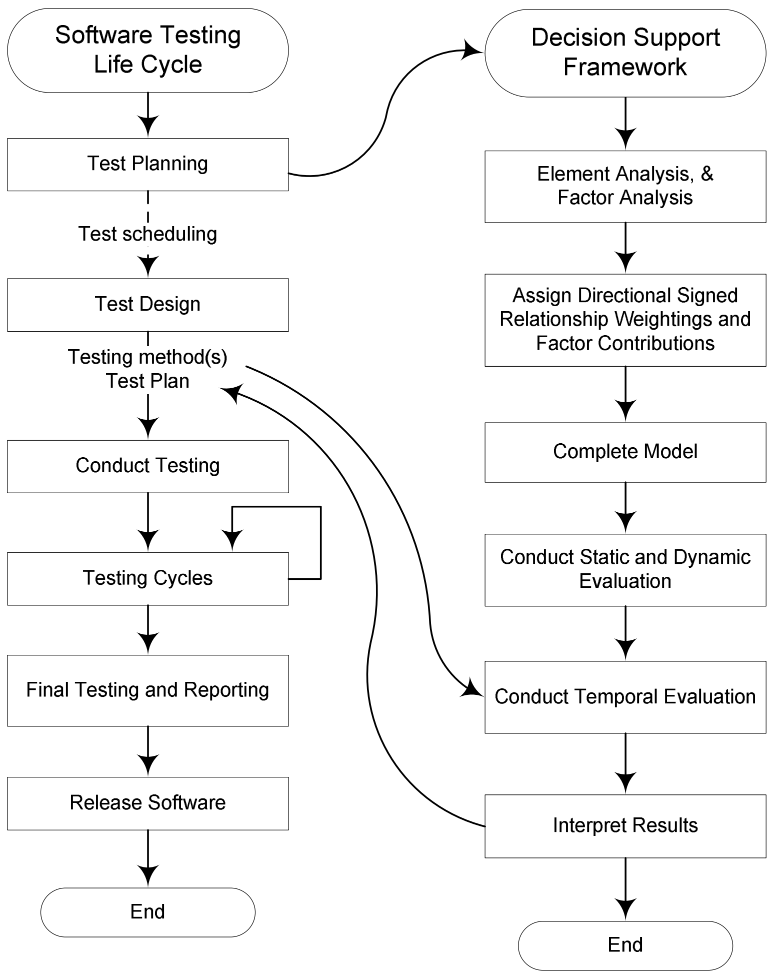


Fig. 4 Revised DSF process with temporal evaluation

6 Summary

The DSF has been expanded to include a temporal perspective as well the analytical static and dynamic perspectives for analysing the DSF's model. This now provides the software test manager with additional information to assessing the risk and its components of successful software testing.

References

Alter, S.: Information systems: Foundation of e-business, 4th edn. Prentice Hall, Upper Saddle River (2002)

Amitysoft, Vibrant trends in software testing (2006), http://www.amitysoft.com/press27.aspx (retrieved from the Amitysoft Technologies website)

Borysowich, C.: Testing: A Sample Test Plan. Toolbox.com. (2005), http://it.toolbox.com/blogs/enterprise-solutions/testing-a-sample-test-plan-3283 (February 24, 2005)

Gelperin, D., Hetzel, B.: The Growth of Software Testing. Special Edition. Communications of the ACM 31(6) (June 1988); The need for and potential growth needed is the software testing industry has long been recognized

Harris, R.: Introduction to decision making (2009), http://www.virtualsalt.com/crebook5.html (retrieved from the VirtualSalt website)

Infiniti Research Limited. Global Software Testing Services Market 2010-2014, pages 23 (June 2011)

Institute of Electrical and Electronics Engineers. IEEE standard for software and system test documentation (IEEE Std 829-2008) (2008), doi: 10.1109/IEEESTD.2008.4578383

Larkman, D., Mohammadian, M., Balachandran, B., Jentzsch, R.: General Application of a Decision Support Framework for Software Testing Using Artificial Intelligence Techniques. In: Phillips-Wren, G., Jain, L.C., Nakamatsu, K., Howlett, R.J. (eds.) IDT 2010. SIST, vol. 4, pp. 53–63. Springer, Heidelberg (2010), doi:10.1007/978-3-642-14616-9_5

Larkman, D., Mohammadian, M., Balachandran, B., Jentzsch, R.: Fuzzy Cognitive Map for Software Testing Using Artificial Intelligence Techniques. In: Papadopoulos, H., Andreou, A.S., Bramer, M. (eds.) AIAI 2010. IFIP AICT, vol. 339, pp. 328–335. Springer, Heidelberg (2010), doi:10.1007/978-3-642-16239-8_43

Larkman, D., Jentzsch, R., Mohammadian, M.: Software testing – factor contribution analysis in a decision support framework. In: Watada, J., Phillips-Wren, G., Jain, L.C., Howlett, R.J. (eds.) IDT 2011. SIST, vol. 10, pp. 897–905. Springer, Heidelberg (2011)

National Institute of Standards and Technology, The economic impacts of inadequate infrastructure for software testing (Planning report 02-3) (2002), http://www.nist.gov/director/planning/upload/report02-3.pdf (retrieved)

Physorg.com (2009), Software testing market resilient despite crisis: Report, http://www.physorg.com/news155992141.html (retrieved)

Lesnikov, S.: Software testing market will growth despite the crisis (March 12, 2009), http://software-testing-ukraine.blogspot.com/2009/03/software-testing-market-will-growth.html

Rani, N.S.: Software test estimation – 9 general tips on how to estimate testing time accurately [Web log post] (2007), http://www.softwaretestinghelp.com/software-test-estimation-how-to-estimate-testing-time-accurately/ (retrieved)

Toward Learning Support for Decision Making: Utilization of Library and Lecture Data

Toshiro Minami and Yoko Ohura

Abstract. Supporting students' learning is very important for a university as an educational organization. It is highly expected to give supports individually according to each student's situation such as his or her knowledge, study skills, learning history, preferences, etc. Due to the development of information and communications technology including Web and Internet, it becomes popular and easy to automatically collect data, and to analyze them and extract knowledge and information about the users' behavior and preferences. For example, the net-companies, such as the ones proving e-commerce services, utilize the customers' behavior data for extracting marketing information. Quite a lot of universities that provide institutional repository (IR) service try to analyze the log data in order to understand and evaluate what the service means for them. The data relating to library services are also useful for library marketing, which aims to provide better user, or patron, services and to improve management. In this paper, we discuss the importance of data analysis of the data relating to lecture data of university classes together with library data.

1 Introduction

Due to the popularization of ICT (Information and Communications Technology) our society is rapidly changing, mainly by the Internet and mobile terminals. The Internet has been well accepted in our society because of good-to-see and easy-to-use of Web. The popularization of Web started with net-surfing as an enjoyment

Toshiro Minami
Kyushu Institute of Information Sciences, 6-3-1 Saifu, Dazaifu,
Fukuoka 818-0117 Japan
e-mail: minami@kiis.ac.jp
Kyushu University Library
e-mail: minami@lib.kyushu-u.ac.jp

Yoko Ohura
Kyushu Institute of Information Sciences, 6-3-1 Saifu, Dazaifu, Fukuoka 818-0117 Japan
e-mail: ohura@kiis.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 137–147.
springerlink.com

of net-surfing. With Web 2.0, even ordinary people can submit information to the world quite easily. We can communicate with other people not only in face-to-face but also online with PCs, mobile phones and other such equipment whenever we want. At the same time, our society is getting more and more complicated. We have to keep learning as long as we live. We live in a life-long-learning age.

Have universities changed themselves sufficiently with using the ICT, changing the curricula and the lectures? Mission of university is to educate students so that they are to become well-trained working personnels in the society as they graduate. Therefore, most universities have already introduced courses for knowledge and skills relating to computer technology in the name of information literacy. Also class lecturers are learning how to use the digitized lecture tools such as word processors and presentation softwares so that they can deliver well-ICTized lectures.

It is a matter of course for many companies to collect customers' behavior data and analyze them for marketing [1]. For example, the net-companies utilize the customers' behavior data for extracting marketing information [11]. It is well known that convenience stores collect POS data and utilize them for predicting the sales of the next day. In these years quite a lot of universities that provide institutional repository service try to analyze the log data in order to understand and evaluate the services. The data relating to library services are also useful for library marketing, which aims to provide better user, or patron, services and to improve management.

In this paper we emphasize on a new approach to utilization of ICT with data analysis for better lectures. In Section 2, we demonstrate the potential usefulness of data analysis by showing some case studies in library data analysis. Circulation records are quite useful and have been analyzed from various points of views [2][3][12]. We also take circulation records for analysis. In Section 3, we present the concept of learning support system, which utilizes the lecture data in addition to library data. We also show some analysis examples which are considered to be useful for extracting tips and knowledge for better lectures as well as providing supports to students in their learning. Finally in Section 4, we conclude our discussions.

2 Library Data Analysis

Library data analysis methods [4][5] are applicable to lecture data analysis in two ways. First, the extracted knowledge/information is applicable as additional knowledge/information to those extracted from lecture data analysis, thus both types can be used for better support of students' learning. It is because supporting students' learning is an important service for libraries and is the essential part of lecturing. In this point of view, libraries and lecturers share a lot in their missions.

Second, the lecture data do not directly link to the knowledge and tips for better learning support in the similar meaning as the library data. In order to utilize these data for effective improvements in their eventual goal, we have to investigate deeply into mining the data. So we might need a lot of trial and error processes before finding something valuable. Thus the lecture data analysis and library data analysis should share a lot in their methodologies.

Book and patron profiling by circulation data analysis
We use the term "profiling" as a description of the target, or a series of attributes together with values for some of them. What important in profiling is not just describing the target book and patron but with showing its position among other targets. Book profiling, or characterization of book, is also quite useful in learning assistant, not only for library marketing with data analysis [9]. We enumerate some of their attributes that are considered to be useful.

Usage Type: As typical patterns on usage of books we consider two types; reference and textbook. A reference book is one that is used occasionally when needed, whereas a textbook is one that is used continuously in a period of time.

Usage Season: In what season the book is borrowed much? Is it borrowed in the period when the students take classes, in vacation seasons like summer, winter, or spring, or during the examination periods? Such seasonal variation might be interesting to know and might be usable for library marketing.

Expertise Level: The feature that describes if the book is used for introductory purpose in the subject area or it is read only by those who know much about the subject field; i.e. experts. We deal with this feature in the next section.

Borrowed Length of Days: Is the book borrowed for a long period of time, or just for a short period? Such difference may indicate some characteristic features of books and provides with some interesting information and tips in library marketing.

Co-Usage: Is the book used alone or is used together with other book(s)? What books are used together? This is a typical market basket analysis on library's circulation records [2] and it provides us with information about books to be read or to be studied together with the borrowed one.

These profile information can be used to help the library with choosing the books to be purchased by the library, with comparing the libraries' collection policy making, recommendation services to patrons.

Book Expertise Level Estimation
Estimation of the expertise level of books has a very high usefulness in library services and for library management. Possible applications include:

(i) Recommendation and Information Provision to Students
It is quite interesting for students to know about themselves in comparison with other students such as what are the books they are reading in terms of the levels of the books and differences from other books, how useful the books might be for the students, the appropriateness of the books, and so on. It must be very helpful for students if their library provides them with services that recommend learning materials, estimate the day-length needed to learn the material, what are the performance of the students in comparison with other students, and something like these.

(ii) Helping Academic Staff with Teaching
Professors should be very pleased if they are able to get information that helps with recognizing the students' knowledge level and thus they are able to let their students catch up with their classes. If the students' level does not match the assumed level, lecturers will adjust the lectures' level with students'. For example they decide to explain more precisely about the material if the students' level is lower than their expectation. Also professors are able to recommend reference books to the class students by considering the students' knowledge level.

(iii) Providing Reference Information to Faculties
The information about the knowledge of students is valuable for faculties as well. Such information might be very helpful for them when they (re-)consider the appropriateness of the courses they provide to students. By considering the knowledge levels of students in accordance with the subjects they are teaching, they can modify syllabuses, add new courses, delete inappropriate courses, and other such actions.

Experiment on book expertise level estimation from circulation data
The circulation records we use are obtained in the Central Library of Kyushu University, Japan from April 2007 to March 2008. The whole data contain 67,304 circulation records. A record item consists of the book ID, book's classification number, call number, borrower's patron ID, borrower's affiliation, borrower's type (undergraduate student, master's student, Ph.D. student, professor, staff, others), and the timestamps for borrowing and returning, etc.

We investigate the expertise level of a book in two values; t-value and a-value [6][7]. The assumption on the definition of t-value is, if a book is borrowed more by the borrowers with high expertise knowledge then the book has a high expertise level value. We assign expertise level from 1 to 10 to borrowers according to the borrower types; from 1 to 6 to the undergraduate students from the 1st year to 6th, respectively, and 8 to the master's students, 9 to the Ph.D. students, and 10 to professors.

The frequencies of books according to the expertise level in t-value have two peaks; one is at 3 (year 3 undergraduate level) and another is at 8 (master's level). Considering the borrower's ratio for the 3rd year student is the maximum among all the student borrowers, the peak value of 3 seems to come from the books borrowed mainly by undergraduate students. On the other hand, the peak value 8 probably comes because these books are borrowed by the graduate students and professors.

For another definition of expertise level of a book, i.e. a-value (affiliation based expertise level), we assume that if a book is borrowed only by those in a limited area, the book's expertise level is high, and if a book is borrowed by patrons in a wide range of expertise fields, the book's expertise level is low. We choose the faculties as the expertise fields. Kyushu University consist of the faculties for undergraduate students and some number of research centers, library, communications center, and others. We will take 12 faculties together with the graduate schools for graduate students relating fields and research organizations for professors; namely SC for sciences, AG for agriculture, TE for engineering, MD for medicine, DD for dental, PS for pharmaceutical, LA for law, LT for letter, EC for economy, ED for education, DS for design, and 21 for special faculty called 21st century program.

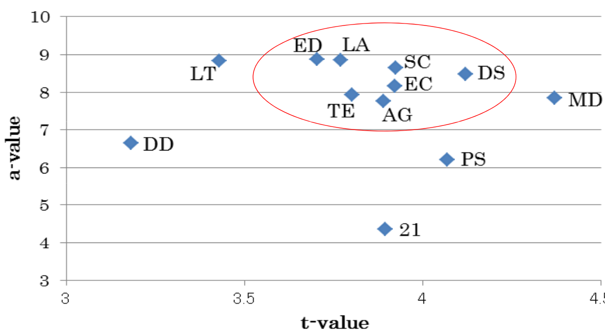


Fig. 1 Distribution of Faculties with t- and a-values

The frequencies of a-value, which varies from 0 to 10, have two peaks; at 1 and 10. Actually all the values belong to 1 are 0, i.e. these 5,686 books (26%) are borrowed by the patrons not affiliated with the nominated faculties. 13,194 books (61%) have value 10, i.e. they are borrowed only those belonging to one faculty group. The average a-value of all data is 6.8 and the average without zero is 9.2.

We define the expertise level of a faculty as the average value of that of books borrowed by the members of the target faculty group in the circulation record. Fig. 1 shows how faculties are distributed in terms of t- and a-values. It is easy to see that 7 faculty groups are located in the range from 3.7 to 4.2 in t-value and from 7 to 9 in a-value. The faculty group MD (medical) is located at the right side of this central region, which probably means more graduate students and maybe more professors affiliated in MD group borrow books from the central library than other faculty groups. PS (pharmaceutical) is located under the central region, which indicates that the members of PS borrow more books that are borrowed by more faculty members than other faculty groups. The faculty group 21 (21st century program) has the similar character more clearly. It is located far lower from the central region, which means the borrowers of this faculty borrowed books that are borrowed by many faculty members. This comes from that the students in this faculty have no specific subject field at the beginning. They can choose lectures from any faculties. The faculty groups LT (letter) and DD (dental) locate left side of the central region, which indicate more undergraduate students borrow books in these groups. Furthermore the a-value for DD is lower than those in the central region like PS, which means the borrowers in these faculties prefer to choose the books that are common to some other faculties than to the books specific to their faculty matters.

3 Learning Support System

3.1 Outline of the Learning Support System

The aim of the learning support system (LSS) we consider in this paper is to provide the user with environment for (automatically) collecting various data relating to the

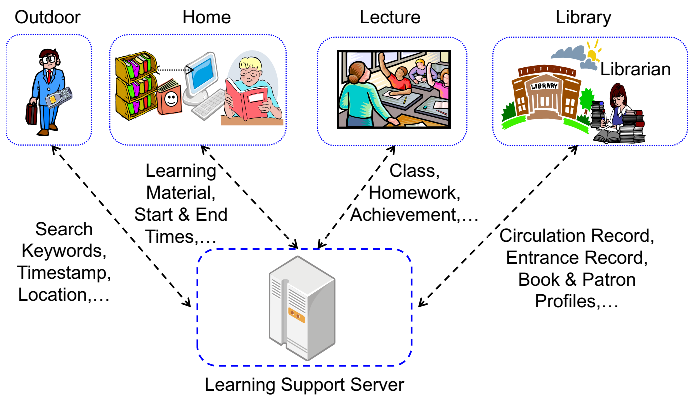


Fig. 2 Overview of Learning Support System

user's learning including classes, self studies, and library usage, analyzing them, and feedbacking the results for better learning. We are expecting that the libraries provide such a learning support service because they have been playing the role of helping the patrons with learning.

Fig. 2 shows the general idea of LSS in this paper. Let, for example, the user be a university student. The LSS server collects his learning data in various ways. A data source is the class he is taking. The class data include syllabus information such as aim of the class, textbook, learning schedule, etc., attendance record, exercise problems during the class and the student's achievements, score of the term-end examination, and so on. The achievement data for other students are expected to be included in the class data because they are valuable to recognize the position of the target student in the class by comparing his data with other students' data.

The student makes connection to the LSS at home, too. He manages his schedule under the support from LSS server, and provides it with the data about his learning. He can access the LSS server as he needs any kinds of information with his mobile equipment such as his smart phone. The server can collect access data such as search keywords, access log data with access timestamp and location data from the user's mobile environment. Library will also play an important role in LSS. As has been shown in the previous sections, library is able to collect useful data for helping its patrons with studying. It can provide the LSS server with circulation record, entrance/exit data, profile information about books and the patron user of the system. The LSS server may collect such data, analyze them, and utilize the result as a support to the users in their study planning, evaluating the study performance, advise the users based-on the comparisons with other users, and many others.

Our LSS is similar to LMS (Learning Management System) [10] such as Moodle [8] in the sense that their goal is to provide the learners with better learning environment, and the teachers with better teaching environment as well. They are somewhat different in the sense the information provided by LMS are rather about the overall status of the learners and each student, whereas the major intention of LSS in this paper is to provide with more individual, or personalized, information based on the learner's personal profile data. Thus these two approaches are not competitive each other but complementary for providing better learning/teaching environment.

3.2 Lecture Data Analysis for Learning Support

There are a lot of data in the lecture that are potentially useful for making each class more effective in learning for students.

Grade Record: The class lecturer has to evaluate the achievements of the class students by giving a score to each student at the end of the term. The score could be the result of the term-end examination or a combination of such a score and the rating score for the student's class participation, the amount of efforts, etc.

Achievement Test Score: The lecturer might give tests for evaluating the achievements of students. The frequencies may be rarely, occasionally or frequently according to the lecturer. Some lecturers give achievement test in every lecture not only for recognizing the class students' achievements but also for taking attendance data. The results of such test will provide a very good source for the lecturer in order to capture the strongs and weaks of students in learning the lecture's subject.

Homework Score: Homeworks are another good source for recognizing the class students' achievements. Not only they provide with students' learning levels but also how serious they are in studying the materials taught in the lectures.

Notebook Score: Students who are poorly achieved in the class often do not willing to take notes during lectures. Thus the ratings if a student keeps a notebook and tries to record what are taught in the lectures, and how good are the records give good source for evaluating the performance of him/her in the class.

Others: Some classes may give extra exercises depending on the class's special purpose. One example of such exercise is typing lessons. In our "information literacy" class, every lesson start with typing exercise using a type-lesson software. The achievement scores of such special exercises also gives information how eagerly the students are participating in the class.

3.3 Analysis on Information Literacy Course's Score Data

We take up the score data for the "information literacy" (IL) class in the second semester (from October 2006 to March 2007) of the fiscal year 2006. Fig. 3 shows the numbers of students according to the score ranges among the 64 students who took this class. The average score of all students is 58.4, which is slightly lower than the pass score of 60. There are two peaks in the class. The largest one is at 80 (i.e. from 80 to 89) and another one is at 0 (i.e. 0 to 9). Quite a number 26 of students (41%) got more than 79, which is the highest grade in this class. On the other hand 13 students (20%) got score 0, which means that 1 out of 5 students did not attend

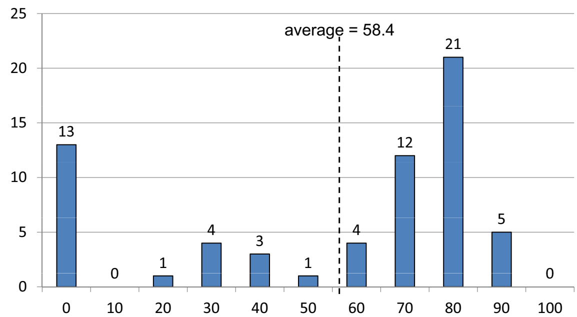


Fig. 3 Number of Students for the Information Literacy Class in 2006 According to Grades

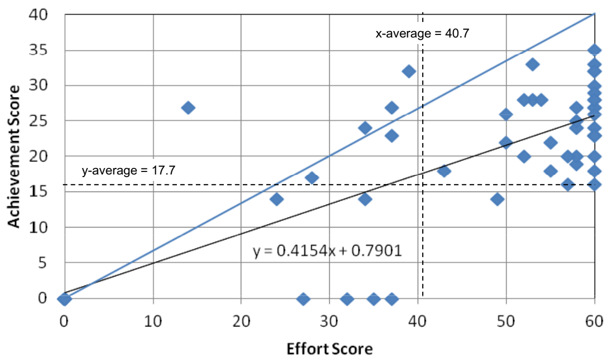


Fig. 4 Correlations between Effort and Achievement Grades

the classes at all and they also did not take the term-end examinations. This is a very high ratio and thus it is a big issue for not only the class but for the university.

We show in detail in Fig. 4 for further investigation. A total score of the IL class consists of that of term-end examination and that of class participation evaluation. In the term-end examination, several topics for writing are announced in advance and each student chooses two of them and he/she writes the answers in the one hour examination. So the students are able to get high scores if they prepare for it sufficiently. The two answers are evaluated with maximum score 20 for each and thus the total score for term-end examination is of maximum of 40. Therefore the term-end examination score can be called the achievement score (a-score), which reflects the preparedness and furthermore ability to plan and carry out.

On the other hand, the class participation score reflects the student's effort in his/her diligence; regular attendance of the class, doing exercises, doing homeworks, and others. Thus we call this score as effort score (e-score) in Fig. 4, and it is evaluated by 60 as maximum. In such a way, the total scores in Fig. 3 is divided into two components, i.e. e- and a-scores, in Fig. 4. The average value 58.4 for the total scores is divided into 40.7 (68% of full score) for e-score and 17.7 (44%) for a-score.

The big difference of these percentages inspires that many students just attend the classes without thinking of learning as much as they can. Thus it suggests that we need more efforts in letting them recognize that they are to study in order to learn, and just to attend the class is a waste of time. If the scores of a student locate on diagonal line from (0,0) to (60,40), it means the student takes proportional scores, or the percentages of the full e- and a-scores). The students plotted in the right triangular area of this line are those who have lower achievements in comparison with efforts. Most students are located in this area.

On the other hand, the ones in the left triangular area have achieved more than their efforts. We can see just 4 students (6%) are located in this area. The left-most one has 14 (35%) in e-score and 27 (45%) in a-score. Even though the student has higher potential than average students in terms of learning abilities including preparing and planning, he/she did not do much in every class. Lecturers have to put more efforts on inspiring them the importance of diligence in learning and doing everyday exercises seriously, and keep putting efforts continuously.

It is also a big problem for us lecturers to encourage the students in the low-achievement area. Actually the linear approximation line from the data locates at much lower position than the diagonal, i.e. proportional scores, line, Comparing the slopes, the approximation line has 0.42 and the diagonal line has 0.67, thus the approximation value is about 0.63 of the proportional value. So quite a lot of students seem to attend and (seemingly) study but not learn sufficiently in the classes.

This finding from the data does support our observations in our classes. Quite a lot of students just hear what are told in the class. Often they are operating their PCs when they are supposed to listen carefully what the lecturer is talking about. So they might believe they can do such different things quite well at the same time. But they are wrong because they can not remember what are told when the lecturer asks them about what he/she have talked about in the previous classes.

Improving such class situations is quite a big and very important issue in order to let them learn as much as possible. However, it is a pity indeed, that we have not an effective idea yet. It must be a key to let them recognize what situation they are in and start thinking seriously that they need to change their way of studying, and let them start improving their attitudes to learning. We have to keep our efforts in exploration on finding a solution or even a tip toward solving this problem.

4 Concluding Remarks

We have discussed about the concept of learning assistance of students with data analysis in this paper. Especially we put special emphasis on the lecture data analysis, which intends to help lecturers with various decision making in the class such as choosing learning materials, deciding lecture level, making homeworks, and many others. In order to make appropriate decisions it is necessary to capture exactly about the students and the materials used in the class together with their matching in various ways.

In this paper we showed a case study of lecture data analysis by taking the similar approach of that for library data. The approach taken in this paper is in its very beginning stage and thus we have to pursue more case studies at the moment, find effective methods, and automate them. Our LSS and LMSs are common in their eventual goal; to maximize the effectiveness and efficiency of the users' learning processes. They are different in the sense our LSS puts more emphasis on recognizing the personal profile, whereas most LMSs provide facilities to capture overall status of the class and summarized information about each learner such as study time, scores of exams, access time logs, etc. Thus these two systems are rather complementary in their functions.

We have to keep on going toward this direction in order to improve our ability in learning assistant of students. Our future plans on this direction include, (1) to estimate the students' knowledge and skill levels more precisely by giving more sophisticated expertise level estimation tools by analyzing the relationship between leaning materials and students in the similar way to the analysis method of expertise level estimation using the relationship between books and borrowing patrons with libraries' circulation records, (2) to install a support function in the learning assistant system that suggests the students with creating study groups and project teams, and (3) to recommend the students of learning materials using the collaborative filtering method in such a way like "the students similar to you take such textbooks."

References

1. American Marketing Association (AMA): Definition of Marketing, `http://www.marketingpower.com/AboutAMA/Pages/DefinitionofMarketing.aspx`
2. Cunningham, S.J., Frank, E.: Market basket analysis of library circulation data. In: Proceeding of 6th International Conference on Neural Information Processing, pp. 825–830. IEEE Computer Society, Perth (1999)
3. Littman, J., Connaway, L.S.: A Circulation Analysis of Print Books and e-Books in an Academic Research Library. Library Resources & Technical Services 48(4), 256–262 (2004)
4. Minami, T., Kim, E.: Seat Usage Data Analysis and its Application for Library Marketing. In: Nguyen, N.T., Kim, C.-G., Janiak, A. (eds.) ACIIDS 2011, Part I. Lecture Notes in Computer Science (LNAI), vol. 6591, pp. 238–247. Springer, Heidelberg (2011)
5. Minami, T.: Potentials of Circulation Data Analysis for Library Marketing: A Case Study in a University Library. In: Kim, T.-h., Adeli, H., Cuzzocrea, A., Arslan, T., Zhang, Y., Ma, J., Chung, K.-i., Mariyam, S., Song, X. (eds.) DTA / BSBT 2011. Communications in Computer and Information Science, vol. 258, pp. 90–99. Springer, Heidelberg (2011)
6. Minami, T.: Book Profiling from Circulation Records for Library Marketing — Beginning from Manual Analysis toward Systematization —. In: International Conference on Applied and Theoretical Information Systems Research (ATISR 2012) (to appear 2012)
7. Minami, T.: Expertise Level Estimation of Library Books by Patron-Book Heterogeneous Information Network Analysis — Concept and Applications to Library's Learning Assistant Service. In: The Fifth International Symposium on Frontiers of Information Systems and Network Applications, FINA-2012 (to appear 2012)
8. Moodle: `http://moodle.org/`

9. Online Computer Library Center, Inc. (OCLC): WorldCat Collection Analysis, `http://www.oclc.org/collectionanalysis/`
10. Romeo, C., Ventura, S., Garcia, E.: Data Mining in Course Management Systems: Moodle Case Study and Tutorial. Computers & Science 15(1), 368–384 (2008)
11. Silverstein, B.: Business-to-Business Internet Marketing. Maximum Press (1999)
12. Yamada, S.: Analysis of Library Book Circulation Data: Turnover of Open-shelf Books. Journal of College and University Libraries 69, 27–33 (2003) (in Japanese)

Using a Hybrid MCDM Model Combining Fuzzy DEMATEL Technique to Examine the Job Stress of Coach Driver

Chui-Hua Liu, Gwo-Hshiung Tzeng, Ming-Huei Lee,
Hao-Lin Tseng, and Po-Yan Lee

Abstract. In the competitive travel-transport field of coach tourism, coach drivers are far more commercially sensitive and susceptible to job stress than common bus drivers or professional drivers. However, few studies have discussed this issue or proposed strategies for alleviating job stress for coach drivers. The purpose of this study is to address the problem of job stress, using fuzzy MCDM (multiple criteria decision making) to examine the relationships between the various aspects of coach driver job stress and, ultimately, to suggest an optimal improvement plan according to the network relationship map (NRM).

Chui-Hua Liu
Department of Tourism & Hospitality, Kainan University, No. 1, Kainan Road,
Luchu Shiang, Taoyuan 33857, Taiwan
e-mail: liu32lee@mail.knu.edu.tw

Gwo-Hshiung Tzeng
Institute of Project Management, Kainan University, No. 1, Kainan Rd., Luchu,
Taoyuan 33857, Taiwan, 3 Institute of Management of Technology,
National Chiao Tung University 1001, Ta-Hsueh Road, Hsinchu 300, Taiwan
e-mail: ghtzeng@cc.nctu.edu.tw, ghtzeng@mail.knu.edu.tw

Ming-Huei Lee
Department of Leisure Management, Taiwan Hospitality & Tourism College,
No 268 Chung-Hsing ST., Feng-Shan Village, Shou-Feng County, Hualien 974, Taiwan
e-mail: mhlee@mail.thtc.edu.tw

Hao-Lin Tseng
Institute of Logistics and Shipping Management, Kainan University,
No. 1, Kainan Road, Luchu Shiang, Taoyuan 33857, Taiwan
e-mail: haolin.tseng@gmail.com

Po-Yan Lee
IMBA Candidate, Schulich School of Business, York University, Canada
e-mail: yenlee1023@gmail.com

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 149–156.
springerlink.com

Keywords: coach driver, coach tourism, job stress, Fuzzy DEMATEL (Decision Making Trial and Evaluation Laboratory), network relationship map (NRM).

1 Introduction

Coach tourism, with its core elements of high mobility and location change, is an effective way for travellers to visit the largest number of sites on a trip of given duration (Lue et al., 1993; Enoch 1996: 601). The coach tour is therefore the most popular travel mode, and coach drivers are significant because they are responsible for the safety of coach tour travellers (Lee, 2009; Wang, 2011).

According to the Tourism Bureau of Taiwan, the number of coaches and coach drivers is rapidly growing to meet the increasing demand for domestic and inbound tourism in Taiwan. However, such growth cannot satisfy the demands of the current tourism market. Coach drivers are reportedly rushing day and night on journeys and loading and unloading tourist groups at almost the same time (China Times, 2011). A series of deadly accidents has occurred as a result of human error – most notably driving while fatigued, which has been linked to job stress (Wang, 2011; Kao et al., 2010). According to the Directory of General of Highways of Taiwan (2010), from 2002 to Nov of 2010, there were 79 serious coach accidents, resulting in 87 deaths and 818 injuries. Nevertheless, very little literature has examined coach drivers and job stress or proposed strategies for improvement.

Based upon the Occupational Stress Index (OSI) of Cooper and Baglioni (1988), the pressure is measured with regard to eight job stressors: workload, relationships, home-work balance, managerial role, personal responsibility, daily hassles, recognition, and organisational climate. Using these actual work variables, Wang (2011) successfully developed a self-reported measurement of 22 items, using a five-point scale to measure 300 coach drivers' job stress. Finally, the research successfully extracted three stressor factors – job nature, social-economic problems, and relationship – with highly reliable Cronbach's α values (respectively, 0.880, 0.903, 0.856). However, this research does not go on to propose any improvement models. Further research by Huang (2012) has produced DEMATEL models, but this research lacks the management of linguistic uncertainty.

The purpose of the present study is to address this problem, using fuzzy MCDM (fuzzy multiple-criteria decision making, FMCDM), which takes fuzziness into consideration (Wu, 2007) and also using the Decision-Making Trial and Evaluation Laboratory (DEMATEL) method (Fontela and Gabus, 1976) to justify the interrelationship between the different coach driver job stress criteria and to present the improvement models for alleviating coach driver job stress.

2 Methodology

2.1 Establishing MCDM Evaluation System

Tzeng & Huang (2011) indicate that MCDM is a methodology that considers multiple criteria at the same time and helps the decision maker to estimate the best case by sorting cases according to the characteristics or criteria of each of the available cases. Based upon the job stress theories (Cooper & Baglioni, 1988; Brief et al., 1988; Mackay et al., 2004), a coach driver evaluation system, including four dimensions and twelve criteria within the context of the coach tour, The dimensions are:(1)Job intrinsic nature*(D_1)*, including the criteria of Insufficient family interaction(C_1), Fatigue driving*(C_2)*, and On alert*(C_3)*;(2)Socioeconomic problems*(D_2)*, including the criteria of little extra incomes*(C_4)*, Meager salary*(C_5)*, and Lack of social recognition*(C_6)*;(3)Unbounded Relationship*(D_3)*, including Shallow company relationships*(C_7)*, Lack of social & recreational activities*(C_8)*, and Conflict with tour partners*(C_9)*.

2.2 Constructing the Fuzzy DEMATEL Questionnaire and Clarifying Interrelationship

The Fuzzy DEMATEL is used to clarify the interrelations between the criteria and confirms the criteria of the evaluation system (Yang & Tzeng, 2011). The DEMATEL technique is an analytical method that uses a structural model to solve complex problems, utilising a matrix and related mathematical theories to calculate the cause and effect of each element (Huang, Shyu & Tzeng, 2007; Yang & Tzeng, 2011; Hung, Chou & Tzeng, 2011). Combined with the fuzzy theory, the fuzzy DEMATEL method is applied in following specific steps:

1. Evaluation criteria are selected and assigned using the fuzzy rating scale.
A survey was conducted utilising questionnaires distributed to an expert group comprised of 30 coach drivers, each with at least 15 years' experience in coach tourism. The survey collected the respondents' ratings for each criterion using a 5-point scale ranging from 0 (no effect) to 4 (extremely influential). To handle the uncertain/vague property of human assessment, fuzzy set theory is incorporated into the original DEMATEL. Five linguistic terms – VH (very high), H (high), M (medium), L (low), and VL (very low) – are adopted to assign the impact of the specific attribute on the corresponding aspect.

2. Linguistic information in interpreted into a fuzzy linguistic scale. Using linguistic information to convert fuzzy numbers into a crisp may score, the fuzzy assessments are defuzzied and aggregated as a crisp value.

3. The criteria are analysed into a cause-and-effect diagram. The crisp value is composed within the initial direct-relation matrix. The normalised direct-relation matrix and a cause-and-effect diagram can then be constructed.

3 Empirical Case Analysis for Coach Driver Job Stress

This section assesses the overall coach driver job stress to propose improving strategies using an empirical case, expert coach drivers in Taiwan. The data collected from these drivers are analysed by a fuzzy DEMATEL method, and the results are presented in useful models for decision-making.

3.1 Computing Fuzzy DEMATEL to Obtain Total Influential Effect

According to the opinions of expert drivers, the research used the steps of Fuzzy DEMATEL in subsection 2.2 to compute the data. Based upon the initial matrix A, the defuzzied matrix and the total effect matrix T of criteria (Table 1) was obtained with reliable agreed rate of 0.048 % (< 5%), and then serve to derive the influential relation (r_i-s_i) in Table 2. It can be seen that Unbounded relationship (D_3) with larger influential value of 0.035 has the strongest direct effect on other dimensions. Job intrinsic nature (D_1) with smallest influential value of -0.058 is the most vulnerable to impact.

3.2 Constructing the Network Relation Map by Fuzzy DEMATEL

This effect is further illustrated in Figure 1, Network Relationship Map (NRM). The priority of influence can thus be sequenced as D3 _ D2 _ D1.

Table 1 Total influential effect matrix *T* of criteria

T	*C1*	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	*C9*
C1	0.11	0.19	0.14	0.14	0.16	0.18	0.15	0.22	0.14
C2	0.16	0.14	0.21	0.19	0.15	0.2	0.22	0.15	0.19
C3	0.18	0.19	0.11	0.15	0.13	0.16	0.18	0.14	0.21
C4	0.25	0.24	0.18	0.16	0.23	0.26	0.25	0.24	0.24
C5	0.22	0.17	0.14	0.15	0.11	0.19	0.22	0.21	0.15
C6	0.13	0.17	0.12	0.16	0.12	0.1	0.13	0.12	0.13
C7	0.2	0.21	0.15	0.22	0.21	0.21	0.14	0.16	0.19
C8	0.17	0.21	0.17	0.14	0.13	0.15	0.14	0.1	0.14
C9	0.15	0.25	0.21	0.22	0.12	0.23	0.16	0.12	0.12

When considering improvement, the expert-drivers all regarded interpersonal relationships as most important and agreed that the first priority for improvement should be the Unbounded relationship (D_3), which can have an influential effect on the remaining dimensions, Job nature (D_1) and Socioeconomic problems (D_2). These results suggest that the drivers' primary concern is interpersonal relationships, including their relationships with the company bosses, tour guides and escorts, and social and recreational interaction. Experts believe that improving these relationships would have a direct effect on improving work conditions and socioeconomic status.

Table 2 The sum of the effects, weights and rankings of each criterion

Dimensions/ criteria	r_i	s_i	r_i+s_i	r_i-$s_{i\ /ranking}$
(D_1)	**0.475**	**0.533**	**1.008**	**-0.058 (3)**
(C_1)	0.445	0.449	0.895	-0.004
(C_2)	0.513	0.528	1.040	-0.015
(C_3)	0.479	0.460	0.938	0.019
(D_2)	**0.527**	**0.504**	**1.031**	**0.023 (2)**
(C_4)	0.654	0.468	1.123	0.186
(C_5)	0.458	0.464	0.922	-0.007
(C_6)	0.383	0.562	0.946	-0.179
(D_3)	**0.506**	**0.472**	**0.978**	**0.035 (1)**
(C_7)	0.491	0.442	0.934	0.049
(C_8)	0.381	0.383	0.764	-0.003
(C_9)	0.409	0.455	0.864	-0.046

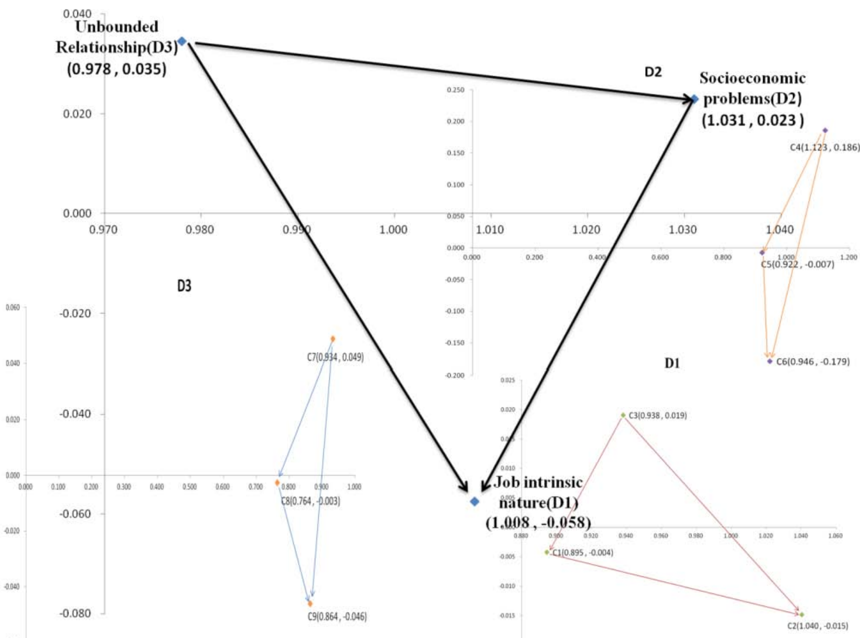


Fig. 1 The influential relation map of each dimension and criteria

The network relation can also be observed as influencing each dimension. For example, within the category of Unbounded relationship (D_3), Shallow company relationships (C_7) has a direct effect on the remaining criteria, Lack of social & recreational activities (C_8) and Conflict with tour partners (C_9). Drivers agree that improved company relationships, including increased personal interaction and a more considerate leadership style, would allow them to participate more in social and recreational activities and would ease the tension with their tour guides and escorts. Thus, the improvement priority in this dimension is (C_7) - (C_8) - (C_9). These findings are valuable to decision makers for improving coach drivers' job stress. For the decision makers, this solution is not only intelligent, but it makes improvement priorities easily identifiable based on complex criteria.

4 Conclusions and Implications

For the empirical case study, the dimensions and criteria of influence are calculated and illustrated using an NRM (Fig. 1). According to the degree of influence shown in Figure 1, the improvement priorities are sequenced as Unbounded relationship, Socioeconomic problems and Job intrinsic nature. This priority sequence is an important point for decision makers. The expert-coach drivers recognise that interpersonal relationships must come first. Efforts in that direction will produce network effects on the remaining dimensions and will spontaneously resolve multiple issues. This finding is consistent with previous research that has identified coach drivers as skilled but lonely outsiders who are focused on mechanical operation and short-term goals, lacking interpersonal interaction and opportunities to develop relationships (Huang, 2012; Wang, 2011). Moreover, this new finding improves upon previous causal findings with regard to the linear relationships between Unbounded Relationship, Socoeconomic problems and Job intrinsic nature (e.g., Wang, 2011; Kuo et al., 2010). The strength of the NRM presented here is that it allows us to illustrate influential networks beyond a linear relationship from the perspective of the dimensions or the criteria.

The stressor criteria, such as Shallow company relationships (C_7), Little extra incomes (C_4) and On alert (C_3), are confirmed to have a more influential effect on the other criteria in the individual dimension (Figure 1). Notably, improved company relationships would produce more organisational commitment and is the first priority because the lack of communication between employees or between employees and employer results in negative commitment to the organisation and a high level of frustration for employees (e.g., Cooper & Baglioni, 1988; Brief et al., 1988; Mackay et al., 2004). Little extra incomes (C_4) also deserves further attention. As mentioned, because of a lack of security and protection of the coach driver's income, benefits, and employment status (Kuo et al., 2010), a little extra income could significantly impact drivers' life expenses and cause deep distress. A reasonable share of commissions, tips and broker fees should be negotiated through constructive dialogue with employers (Huang, 2012). Furthermore, more attention should be given to On alert (C_3) because an appropriate break from continuous driving is necessary and helpful for reducing tension and fatigue. Both the authorities of the coach companies and the government should keep this point in mind to require scheduled breaks to refresh the mind, body and spirit of drivers and ultimately achieve safe driving.

References

Bartlett, D.: Stress: perspectives and processes. Open University Press, UK (1998)

Brief, A.P., Schuler, R.S., Van'sell, M.: Managing job stress. Little Brown, (1981); as cited in Umstot, Boston, D.D.: Understanding organizational behavior, 2 edn. West Publishing Company, St.Paul (1988)

Chen, F.H., Hsu, T.S., Tzeng, G.H.: A balanced scorecard approach to establish a performance evaluation and relationship model for hot spring ho-tels based on a hybrid MCDM model combining DEMATEL and ANP. International Journal of Hospitality Management 30(4), 908–932 (2011)

China Times Coach driver has got 50% salary (August 12, 2011), `http://news.chinatimes.com/focus/50109356/112011081200523.html` (retrieved on August 15, 2011)

Cooper, C.L., Baglioni, A.J.: A structural model approach toward the development of a theory of the link between stress and mental health. British Journal of Medical Psychology 3, 87–102 (1988)

Enoch, Y.: Contents of tour packages. A Cross-Cultural Comparison. Annals of Tourism Management 23(3), 599–616 (1996)

Fontela, E., Gabus, A.: The DEMATEL Observer, DEMATEL 1976 Report. Battelle Geneva Research Center, Switzerland (1976)

Huang, C.Y., Shyu, J.Z., Tzeng, G.H.: Reconfiguring the innovation policy portfolios for Taiwan's SIP mall industry. Technovation 27(12), 744–765 (2007)

Huang, C.K.: The job stress of tourist bus driver: using DEMATEL model. Unpublished master thesis, Institute of Tourism and Leisure Management. Chinese Culture University, Taiwan (2012)

Hung, Y.H., Chou, S.C.T., Tzeng, G.H.: Knowledge management adoption and assessment for SMEs by a novel MCDM approach. Decision Support Systems 51(2), 270–291 (2011)

Kao, H.T., Hsieh, H.S., Chen, C.H.: The study on the relationship between driving safety perception and aberrant driving behavior for tourist bus drivers. Journal of Taiwan Hospitality & Tourism, Special Issue of Asia Tourism Forum Proceeding, 133–144 (July 2010)

Lee, F.L.: A study in the safety management of the "License Leasing Practice" of the tour buses, unpublished master thesis, Graduate Institute of Traffic and Transportation Engineering and Management, Feng-Chia? University, Taiwan (2009)

Lee, W.S., Tzeng, G.H., Cheng, C.M.: Using novel MCDM methods based on Fama–French three-factor model for probing the stock selection. Paper presented at the 10th Asia Pacific Industrial Engineering and Management Systems Conference, APIEMS, December 14-16, pp. 1460–1474 (2009)

Liu, C.H., Tzeng, G.H., Lee, M.H.: Improving tourism policy implementation－The use of hybrid MCDM models. Tourism Management 33, 413–426 (2012)

Lue, C.C., Crompton, J.L., Fesenmaier, D.R.: Conceptualisation of mul-tidestination pleasure trips. Annals of Tourism Research 20, 289–301 (1993)

Mackay, C., Cousins, R., Kelly, P.J., Lee, S., Mccaig, R.H.: Managemwnt Standaeds' and work-related stress in the UK: policy background and science. Work and Stress 18(2), 91–112 (2004)

Opricovic, S.: Multicriteria optimization of civil engineering systems. Bel-grade: Faculty of Civil Engineering (1998)

Ou Yang, Y.P., Shieh, H.M., Leu, J.D., Tzeng, G.H.: A novel hybrid MCDM model combined with DEMATEL and ANP with applications. International Journal of Operations Research 5(3), 160–168 (2008)

Tsai, W.H., Hsu, J.L., Chen, C.H., Lin, W.R., Chen, S.P.: An integrated approach for selecting corporate social responsibility programs and costs evaluation in the international tourist hotel. International Journal of Hospitality Management 29(3), 385–396 (2010)

Tzeng, G.H., Huang, J.J.: Multiple attribute decision making: Methods and applications. Taylor & Francis Group, Boca Raton (2011)

Tzeng, G.H., Lin, C.W., Opricovic, S.: Multi-criteria analysis of alternative-fuel buses for public transportation. Energy Policy 33(1), 1373–1383 (2005)

Yang, J.L., Tzeng, G.H.: An integrated MCDM technique combined with DEMATEL for a novel cluster-weighted with ANP method. Expert Systems with Applications 38(3), 1417–1424 (2011)

Wang, S.H.: Coach diver job stress and aberrant driving behavior. Unpublished master thesis, Institute of Tourism and Leisure Management. Chinese Culture University, Taiwan (2011)

Wu, W.W.: Choosing knowledge management strategies by using a combined ANP and DEMATEL approach. Expert Systems with Applications 35(3), 828–835 (2008)

Using Fuzzy MCDM Method to Exploring the Influence Degree of Project Team Effectiveness Maturity

YaoFeng Chang and Hiroaki Ishii

Abstract. The Project Team Effectiveness Maturity (PTEM) refers to the goal achievement level of a project team to achieve its anticipated results on time and within budget. This including supports the project team or organization through evaluate the key processes and implementation of team. Therefore, the purpose of this paper is Using Fuzzy MCDM method to exploring the Influence Degree of Project Team Effectiveness Maturity. The results found that there were interactive relations between all the criteria, where the dimension of "Capacity management" was the most influential dimensions; on the contrary the "Organization management" was the least dimensions. Furthermore, criteria "Authority", "Capacity nurturing", "Conflict management" and "Performance management" have the higher influences among each dimension, so we suggest to consider them as the major steps to promote the capacity of project team.

Keywords: PTEM (project team effectiveness maturity), Fuzzy DEMATEL (fuzzy decision making trial and evaluation laboratory), ANP (analytical network processes).

1 Introduction

Now, more and more organizations and enterprises have begun to rely on project teams to fulfill assignments which are hard to be accomplished by a single branch or by branches in hierarchical organizational structure. The major manner to transform an organization is "Team", especially means project team. The "project team" already became the most effective and universal organization pattern to integrate knowledge and capacity under the indefinite environment.

Besides, more and more business operations have begun to focus on projects instead of organizational structures. However, the high failure rate of projects remains high even till now. Through analysis of the causes of failure of projects, it has been found out that the key of success or failure of a project lies in the project

YaoFeng Chang · Hiroaki Ishii
School of Science and Technology, Kwansei Gakuin University, 2-1gakuen,
Sanda-shi, Hyogo 669-1337, Japan

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 157–167.
springerlink.com

team. Therefore, the dissertation puts forward the concept of Project Team Effectiveness Maturity (PTEM) and makes some explorations into it theoretically and practically.

PTEM refers to the goal achievement level of a project team to achieve its anticipated results on time and within budget. Taking PTEM as its object; the coupling relationship between system composing factors of PTEM as its basis and the advancement system of PTEM as its approach, through the implementation of effectiveness advancement process corresponding to different maturity grades, the research on PTEM aims to realize the development of project team effectiveness from lower maturity grades to higher ones so as to improve the success rate of projects. The research flowchart is shown as Figure1 on the basis of above statement.

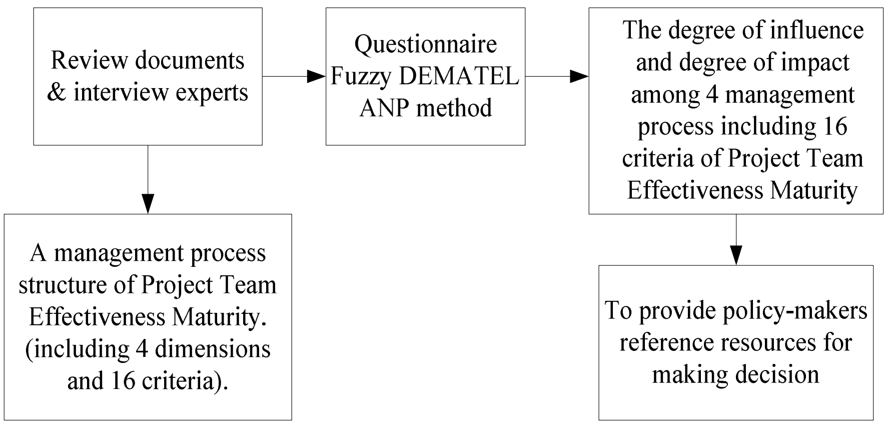


Fig. 1 Research flowchart

2 The Components of Project Team Effectiveness Maturity Research

Taking PTEM as its object; the coupling relationship between system composing factors of PTEM as its basis and the advancement system of PTEM as its approach, through the implementation of effectiveness advancement process corresponding to different maturity grades, the research on PTEM aims to realize the development of project team effectiveness from lower maturity grades to higher ones so as to improve the success rate of projects.

Most businesses are at level one grades, the ad hoc/chaos stage, where there is no standard project management process and no defined project steering process. (The project steering process is the process that the management team uses to select, prioritize, allocate resources and oversee the portfolio of projects.) In order to move to level two, Standardization, the project team needs to do the following:

1. Chose a standard PM method.
2. Make sure project leaders are using basic project management skills.
3. Make sure sponsors are writing charters for all new projects.
4. Have project teams use basic project management templates.

Level two grades create the basis for developing a core competency in project management across the organization. As you begin to move to level three, Core Competency, you'll to need spread the training process for basic project management skills to project team members, and continue the development of your project leader's skills. In addition, you'll need to establish the rudiments of a project steering process, which should be run by the project steering council (a subset of the management team). At level three, every manager in the organization accepts accountability for the effectiveness of the organization's portfolio of projects.

Figure 2 shows the content of theses 4 management process dimensions including all 16 criteria of project team capacity maturity. Besides, it fits to the system operation structure of project team capacity maturity. We discuss each management process dimension as following.

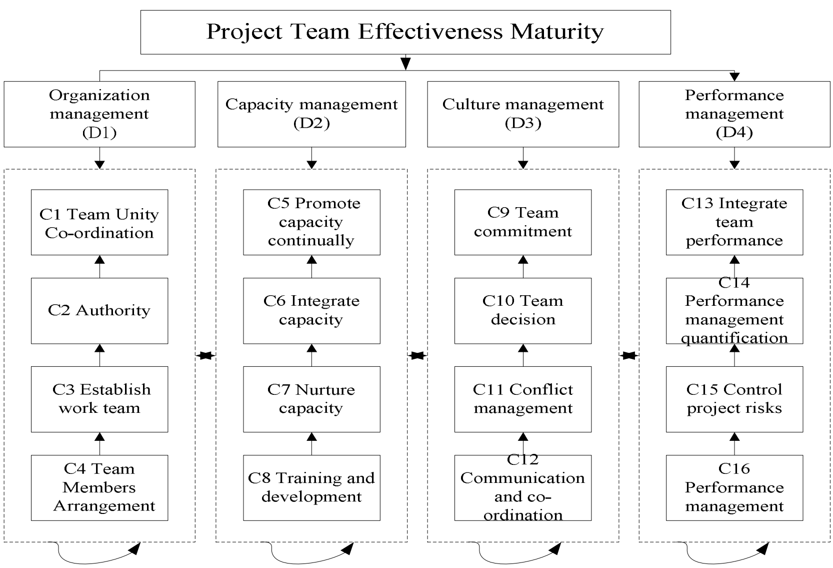


Fig. 2 Structure of project team effectiveness maturity

3 Research Method

A hybrid MCDM model combined with fuzzy DEMATEL, FDANP and VIKOR, for evaluating and improving problems is more suitable in the real world than the previously available methods. This study used the fuzzy DEMATEL technique to acquire the structure of the MCDM problems.

3.1 Fuzzy DEMATEL Method

The application of the fuzzy DEMATEL method expresses different degrees of influences or causalities obtained from crisp DEMATEL using five linguistic terms (Very high, High, Low, Very low, No) and their corresponding positive triangular fuzzy numbers (Lin &Wu, 2004). The linguistic terms and corresponding fuzzy numbers are shown in Table 1.

Table 2 Fuzzy DEMATEL linguistic terms and fuzzy numbers

Linguistic terms	Fuzzy number
Very High Influence (VH)	(0.75, 1.0, 1.0)
High Influence (H)	(0.5, 0.75, 1.0)
Low Influence (L)	(0.25, 0.5, 0.75)
Very Low Influence (VL)	(0, 0.25, 0.5)
No Influence (No)	(0, 0, 0.25)

Tzeng *et al*. (2007) indicate that DEMATEL can help understand special problems, collaborate with problem groups, and provide feasible ideas. The method can be applied as follows:

Step 1*: Calculate the initial average matrix* A *by using scores.*

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix} \tag{1}$$

Step 2*: Calculate the initial influence matrix X.*

$$X = m \times A \tag{2}$$

$$m = \min\left[\frac{1}{\max\limits_i \sum\limits_{j=1}^{n} |a_{ij}|}, \frac{1}{\max\limits_j \sum\limits_{i=1}^{n} |a_{ij}|}\right] \tag{3}$$

Step 3*: Derive the full direct/indirect influence matrix* ***T****.*

$$\boldsymbol{T} = \boldsymbol{X} + \boldsymbol{X}^2 + \boldsymbol{X}^3 + \cdots + \boldsymbol{X}^q = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{X})^{-1}$$

Proof:

$$\begin{aligned}\boldsymbol{T} &= X(I + X + X^2 + \cdots + X^{q-1})(I - X)(I - X)^{-1} \\ &= \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{X}^q)(\boldsymbol{I} - \boldsymbol{X})^{-1}\end{aligned}$$

when $q \to \infty$, $\boldsymbol{X}^q = [0]_{n \times n}$, then

$$\boldsymbol{T} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{X})^{-1} \tag{4}$$

where $\boldsymbol{T} = [t_{ij}]_{n \times n}$, $i, j = 1, 2, \ldots, n.$

*Step 4: Build the Network Relation Matrix based on the vectors **r** and **c**.*

$$\boldsymbol{r} = [r_i]_{n \times 1} = \left[\sum_{j=1}^{n} t_{ij} \right]_{n \times 1} \tag{5}$$

$$\boldsymbol{c} = [\boldsymbol{c}_j]_{n \times 1} = \left[\sum_{i=1}^{n} \boldsymbol{t}_{ij} \right]_{n \times 1} \tag{6}$$

where *r* denotes the sum of the row “i” in matrix T ,and *c* denotes the sum of the column “j” in matrix T.

3.2 Combining Fuzzy DEMATEL and ANP for Calculating the Weights of Criteria

ANP is the general form of the analytic hierarchy process (AHP) (Saaty, 1980), which has been used in MCDM to relax the restriction of hierarchical structure. Within ANP, there is an outer dependence among clusters and an inner dependence within the criteria of clusters, as illustrated in Figure 3.

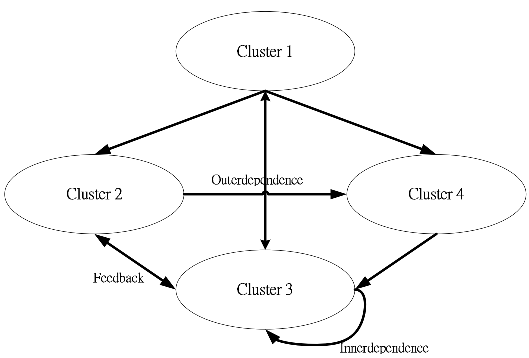


Fig. 3 Relation of clusters

According to Ou Yang *et al.* (2008), a supermatrix normalizes the matrix by assuming each pair of criteria has the same weight. Although such a method can easily normalize it, it also neglects the fact that different groups have a different degree of impact. In this study, the following steps are used:

Step 1: *Compare the criteria in the supermatrix.*

$$
\begin{array}{cc}
 & \begin{array}{cccc} C_1 & C_2 & \cdots & C_n \\ e_{11}e_{12}\cdots e_{1n} & e_{21}e_{22}\cdots e_{2n} & \cdots & e_{N1}e_{N2}\cdots e_{Nn} \end{array} \\
\boldsymbol{W} = \begin{array}{cc} C_1 & \begin{array}{c} e_{11} \\ e_{12} \\ \vdots \\ e_{1n} \end{array} \\ C_2 & \begin{array}{c} e_{21} \\ e_{22} \\ \vdots \\ e_{2n} \end{array} \\ \vdots & \vdots \\ C_n & \begin{array}{c} e_{N1} \\ e_{N2} \\ \vdots \\ e_{Nn} \end{array} \end{array} & \begin{bmatrix} \boldsymbol{W}_{11} & \cdots & \boldsymbol{W}_{12} & \cdots & \boldsymbol{W}_{1N} \\ \boldsymbol{W}_{21} & \cdots & \boldsymbol{W}_{22} & \cdots & \boldsymbol{W}_{2N} \\ \vdots & & \vdots & & \vdots \\ \boldsymbol{W}_{N1} & \cdots & \boldsymbol{W}_{N2} & \cdots & \boldsymbol{W}_{NN} \end{bmatrix}
\end{array} \tag{7}
$$

Step 2*: Obtain the weighted supermatrix by multiplying the normalized matrix, which is derived according to the DEMATEL method.*

$$
\boldsymbol{T}_\alpha = \begin{bmatrix} t_{11}^{\alpha} \cdots & t_{1j}^{\alpha} \cdots & t_{1n}^{\alpha} \\ \vdots & \vdots & \vdots \\ t_{i1}^{\alpha} \cdots & t_{ij}^{\alpha} \cdots & t_{in}^{\alpha} \\ \vdots & \vdots & \vdots \\ t_{n1}^{\alpha} \cdots & t_{nj}^{\alpha} \cdots & t_{nn}^{\alpha} \end{bmatrix} \tag{8}
$$

where if $t_{ij} < \alpha$, then $t_{ij}^{\alpha} = 0$ else $t_{ij}^{\alpha} = t_{ij}$, and t_{ij} is an ij criteria of the total-influence matrix T. Dividing by the following value makes the α-cut total-influence matrix $\boldsymbol{T}_\alpha$ normalized.

$$
d_i = \sum_{j=1}^{n} t_{ij}^{\alpha} \tag{9}
$$

Therefore, we normalize the α-cut of the total-influence matrix and denote it as T_n.

$$
\begin{aligned}
T_n &= \begin{bmatrix} t_{11}^{\alpha}/d_1 \cdots & t_{1j}^{\alpha}/d_1 \cdots & t_{1n}^{\alpha}/d_1 \\ \vdots & \vdots & \vdots \\ t_{i1}^{\alpha}/d_i \cdots & t_{ij}^{\alpha}/d_i \cdots & t_{in}^{\alpha}/d_i \\ \vdots & \vdots & \vdots \\ t_{n1}^{\alpha}/d_3 \cdots & t_{nj}^{\alpha}/d_3 \cdots & t_{nn}^{\alpha}/d_3 \end{bmatrix} \\
&= \begin{bmatrix} t_{11}^{n} \cdots & t_{1j}^{n} \cdots & t_{1n}^{n} \\ \vdots & \vdots & \vdots \\ t_{i1}^{n} \cdots & t_{ij}^{n} \cdots & t_{in}^{n} \\ \vdots & \vdots & \vdots \\ t_{n1}^{n} \cdots & t_{nj}^{n} \cdots & t_{nn}^{n} \end{bmatrix}
\end{aligned} \tag{10}
$$

where $t_{ij}^{n} = t_{ij}^{\alpha} / d_{i}$.This study adopts the normalized α-cut total-influence matrix T_{n}. The unweighted supermatrix W is changed in the weighted supermatrix W_{w} through Equation (11), which shows the level of influence values in the weighted supermatrix.

$$W_{w}=\begin{bmatrix} t_{11}^{n}\times W_{11}\cdots & t_{21}^{n}\times W_{12}\cdots & t_{n1}^{n}\times W_{1n} \\ \vdots & \vdots & \vdots \\ t_{12}^{n}\times W_{21}\cdots & t_{22}^{n}\times W_{22}\cdots & t_{ni}^{n}\times W_{in} \\ \vdots & \vdots & \vdots \\ t_{1n}^{n}\times W_{n1}\cdots & t_{2n}^{n}\times W_{n2}\cdots & t_{nn}^{n}\times W_{nn} \end{bmatrix} \tag{11}$$

Step 3: *Limit the weighted supermatrix by raising it to a sufficiently large power p, as shown in Equation (12),*

$$\lim_{p\to\infty} W_{w}^{p} \tag{12}$$

This formula is limited the weighted supermatrix, when $\boldsymbol{p} \to \infty$, the W_{w} has converged and become a long-term stable supermatrix.

4 Empirical Analysis

4.1 Constructing the Fuzzy DEMATEL

We discussed the effectiveness of project team through dividing them into 4 dimensions, totally 16 criteria (as shown in Fig.2), to analyze the interrelation among them by questionnaire survey. We request 25 experts, who have joined in projects over two years, to fill out the questionnaires. Then, accounting and analyzing the interrelatedness among each dimension and criterion according to the recycled 20 questionnaires.

We calculated with Fuzzy DEMATEL method to confirm the question structure then obtain above result. We got total-relation matrix T shown as Table 2, and we may understand, between each dimension all has its reciprocity.

By Table 3 demonstrated its influence, we may see Team Performance Management Process has the strongest relatedness with other processes. (value of the

Table 2 Project Team Effectiveness Maturity Total Relation Matrix

T	D1	D2	D3	D4
D1	0.51	0.50	0.51	0.53
D2	0.52	0.47	0.50	0.52
D3	0.54	0.47	0.48	0.55
D4	0.55	0.51	0.52	0.55

r+d to be highest); but Team Capacity Management Process has the weakest relatedness with other processes (value of the r+d to be lowest). Team Capacity Management Process has the greatest influence to other dimensions. (value of the r-d to be highest). On the contrary, Team Organization Management Process has the smallest influence to other processes (value of the r-d to be lowest).

Table 3 Project Team Effectiveness Maturity Influence Degree

Dimension	r	d	r+d	r-d
Team Organization Management Process	2.05	2.12	4.17	-0.07
Team Capacity Management Process	2.01	1.94	3.95	0.07
Team Culture Management Process	2.04	2.01	4.05	0.03
Team Performance Management Process	2.13	2.17	4.30	-0.04

We can draw up the dimension relationship mapping shown as Figure 4 based on Table 1 and Table 2, in which r+d (X axis) represents the relation degree among each dimension; r-d (Y axis) represents the influence degree among each dimension, r-d>0 means the dimension could be comparatively considered as a cause factor; r-d>0 means the dimension could be comparatively considered as an effect factor.

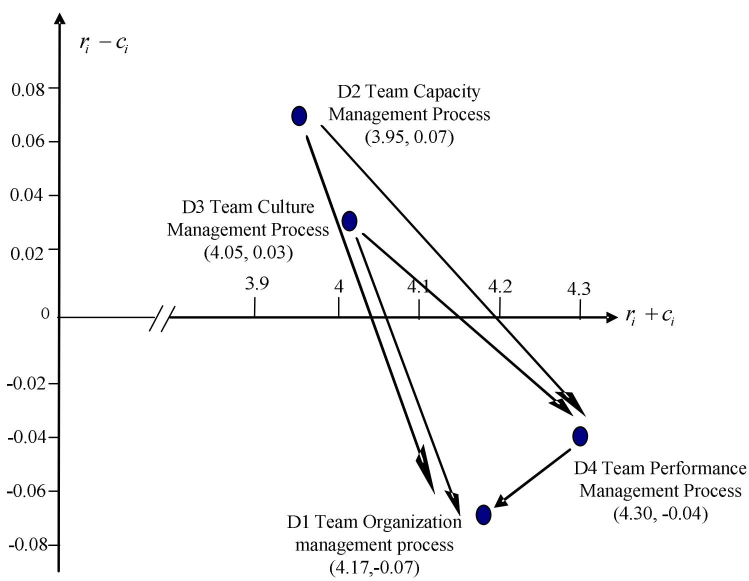


Fig. 4 The Dimensions impact-direction map of PTEM

4.2 Calculating Weights of Criteria by ANP

After getting Dimension Relationship Mapping, we follow the calculating steps of ANP method for acquiring Table 4 to explain the weight. It means the degree of impact of these dimensions and criteria. The primary survey experts included scholars of project and managers of project. The level of importance (global weights) of 16 criteria can be calculated by ANP shown as Table3.

Table 4 The weights ranking of each criteria

Dimensions	**Criteria**	**Weight of Dimensions**	**Weight ranking**	**Weight of criteria**	**Weight ranking**
D1 Team Organization Management Process	C1 Team Unity Co-ordination	0.26	2	0.260	2
	C2 Authority			0.217	4
	C3 Establish Work Team			0.252	3
	C4 Team Members Arrangement			0.271	1
D2 Team Capacity Management Process	C5 Promote Capacity Continually	0.24	3	0.253	2
	C6 Integrate Capacity			0.264	1
	C7 Nurture Capacity			0.236	4
	C8 Training and Development			0.246	3
D3 Team Culture Management Process	C9 Team Commitment	0.23	4	0.235	4
	C10 Team Decision			0.249	3
	C11 Conflict Management			0.252	2
	C12 Communication and Co-ordination			0.264	1
D4 Team Performance Management Process	C13 Integrate Team Performance	0.27	1	0.251	2
	C14 Performance Management Quantification			0.252	1
	C15 Control Project Risks			0.250	3
	C16 Performance Management			0.247	4
	Sum	1.00		4	

According to the data on Table 4, we find Dimension 4 (Team Performance Management Process) has the highest weight. The second one is Dimension 1 (Team Organization Management Process), Dimension 2 (Team Capacity Management Process) and Dimension 3 (Team Culture Management Process) have lower weight than above. However, Dimension 2 and Dimension 3, which have the lower weight, are both effect factors displaying on Figure 4 – Dimension

Relationship Mapping. For this reason, improving Dimension 2 and Dimension 3 to influence Dimension 4 and Dimension 1 for promoting all is the best way.

5 Conclusion and Discussion

This research adopts a Fuzzy MCDM method, including Fuzzy DEMATEL and ANP, for exploring project team effectiveness maturity. This research provides leaders the suggestions for improving team's management process and increasing team efficiency.

In evaluating the PTEM model, experts considered and recognize the dimension with the greatest influence among 4 dimensions is Team Capacity Management Process. And, Criterion A2（Authority）, Criterion B3 (Nurture Capacity), Criterion C3 (Conflict Management), and Criterion D4（Performance Management） are with the greatest influence on each dimension. For this reason, we suggest leaders to improve above dimension or criteria at first to have the advantage of promoting team capacity management process.

In the future, we advise to keep in researching about the dimension of external environment which is including some criteria like human, finance, resources, and organization supports. Then, to recognize the relative superiority and degree of maturity in the cross-sectional domestic enterprises after they adopt project team management method for providing a reference to domestic enterprises targets about team promoting and team management methods.

References

[1] Sundstrom, E., De Meuse, K.P., Futrell, D.: Work teams: Applications and effectiveness. American Psychologist 45(2), 120–133 (1990)

[2] Xiaoyi, D.C., Wells, G.W.: An exploration of project management office features and their relationship to project performance. International Journal of Project Management 22, 523–532 (2004)

[3] Vionkur-Kaplan, D.: Treatment Teams that Work (and those that don't): An Application of Hackman's Group Effectiveness Model to Interdisciplinary Teams in Psychiatric Hospitals. Journal of Applied Behavioral Science 31(3), 308 (1995)

[4] Huang, Y.Q.: Building high performance project teams. East China University of Science and Technology Press (2008) ISBN:978-7-5628-2301-8

[5] Yang, M.H.: Researching the Project Team Capacity Maturity. China Social Sciences Press (2008) ISBN:978-7-5004-7464-7

[6] Webber, S.S., Donahue, L.: Impact of Highly and Less Job-related Diversity on Work Group Cohesion and Performance: a Meta-analysis. Journal of management 27(2), 141–162 (2001)

[7] Gladstein, D.: Groups in context: A Model of Task Group Effectiveness. Administrative Science Quarterly 29, 499–517 (1984)

[8] Cohen, S.G., Ledford Jr., G.E., Spreitzer, G.M.: A Predictive Model of Self-Managing Work Team Effectiveness. Human Relations 49(5), 647 (1996)

[9] Tzeng, G.H., Lin, C.W., Opricovic, S.: Multi-criteria analysis of alternative-fuel buses for public transportation. Energy Policy 33(11), 1373–1383 (2005)

[10] Ou Yang, Y., Shieh, H., Leu, J., Tzeng, G.H.: A novel hybrid MCDM model combined with DEMATEL and ANP with applications. International Journal of Operations Research 5(3), 160–168 (2008)

[11] Tzeng, G.W., Chiang, C.H., Li, C.W.: Evaluating intertwined effects in e-learning programs: a novel hybrid MCDM model based on factor analysis and DEMATEL. Expert Systems with Applications 32(4), 1028–1044 (2007)

[12] Jassbi, J., Mohamadnejad, F., Nasrollahzadeh, H.: A Fuzzy DEMATEL framework for modeling cause and effect relationships. Expert Systems with Applications 38, 5967–5973 (2011)

[13] Belassi, W., Tukel, O.I.: A new framework for determining critical success/failure factors in projects. International Journal of Project Management 14(3), 141–151 (1996)

[14] Ding, J.F.: Fuzzy MCDM approach for selecting strategic partner: An empirical study of a container shipping company in Taiwan. International Journal of Innovative Computing, Information and Control 5(4), 1055–1068 (1999)

[15] Hsieh, T.Y., Lu, S.T., Tzeng, G.H.: Fuzzy MCDM approach for planning and design tenders selection in public office buildings. International Journal of Project Management 22, 573–584 (2004)

[16] Wu, W., Lee, Y.: Selecting knowledge management strategies by using the analytic network process. Expert systems with Applications 32(3), 841–847 (2007)

Part II
Other Related Topics

A Model Study on Emotional Communication in a Mono-brand Fashion Store Application of the Lens Model in the Fashion Industry

Zhiqing Jiang and Shin'ya Nagasawa

Abstract. The purpose of this paper is to propose a model that analyzes the visual efficiency of emotional communication from brand manager to consumers in a mono-brand fashion store. Before setting up the model, an investigation of relevant literatures in marketing and psychology was done in order to put forward a hypothesis of the model. Based on recommendations of the extant research, the scale used to measure emotional communication in this study mainly relied on measuring brand personality. Then, an empirical, statistical study was conducted of Tod's in China for verifying the feasibility of the emotional-communication model. The results show that the emotional-communication model works to explain the visual efficiency of emotional communication in a mono-brand fashion store; a set of regression equations of estimated expectation is also provided on both brand manager and consumers' part. This research has presented a new perspective on fashion brands and on managers, consumers, and brand stores as a single mechanism.

1 Introduction

In recent years, one important symbolic brand association in marketing research has been *brand personality*, which is defined formally as "the set of human characteristics associated with a brand", and human perceptions are the basis of individual behavior, attitudes and beliefs, as well as physical and demographic characteristics (Aaker, 1997). Unlike "product-related attributes" serving a utilitarian function for consumers, brand personality tends to serve a symbolic or self-expressive function (Keller, 1993), which contributes to the emotional effects of the brand. Emotion arguably provides the means to coordinate the diverse mental and physical components required to respond to the world in a coherent fashion (Cosmides and Toody, 2000). Meanwhile, on the part of consumers, it is the store rather than the brand

---

Zhiqing Jiang · Shin'ya Nagasawa
Waseda University, Graduate school of commerce
e-mail: `zhiqing-jiang@ruri.waseda.jp, nagasawa@waseda.jp`

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 171–179.
springerlink.com

that acts as the activation area. Different groups of consumers prefer different types of retail stores (Finn and Louviere, 1990). In the fashion industry, this phenomenon is extremely significant because the industry has undergone a slow, general shift in the last decade away from production towards retailing and services. This has been one of the most important aspects of the fashion sector's evolution (Tartaglione, 2005).

The communication of emotion in a mono-brand fashion store has been chosen as the subject of this thesis because the perceived *emotional essence* of a brand influences the degree of consumer brand loyalty; in other words, the emotional experience can contribute to the brand's development, while, conversely, brand personality can offer consumers various emotional experiences. Therefore, how consumers feel about a brand in a mono-brand fashion store is worth discussion.

2 Theoretical Framework

There is no universally accepted definition of emotional communication. In psychology, Bucci (2001) stated that the phenomena characterized clinically as "unconscious communication" may be accounted for systematically as emotional communication, which occurs both within and outside of our awareness. In marketing, Clow and Baack (2007) defined the communication process that is part of any advertising or marketing program as "sender→encoding→transmission device→decoding→receiver." They explained that encoding the message is the second step in the communication process, which first requires a creative idea. Decoding occurs when the message reaches one or more of the receiver's senses. And Henderson (1996) noted that environmental psychology draws from the stimulus–organism–response (S-O-R) paradigm, which finds its roots in psychology.

In order to explain the complex cognitive process of emotional communication in the context of a mono-brand fashion store, we introduce the "lens model" of Brunswik(1956), which indicates the possibilities available to any cognitive progress over time. Furthermore, particular attention was paid to the "lens" function, which acts as multi-mediation and is the most important part of the model, and to the sensory cues inside of it. After outlining a summary of the application of the lens model, we were inspired by Juslin (2000), who has generalized the model for the first time in research on emotional response to music. Juslin took the music itself as the "lens" and set up a theoretical model to explain the emotional-communication process of music. However, a mono-brand fashion store, despite the instability of human factors, involves all the sensory variables, and from an examination of the marketplace, it is clear that not all the cues outlined in the lens model will be necessary for all fashion stores.

Visual brand language is branding terminology for a unique "alphabet" of design elements which directly and subliminally communicate a company's values and personality through compelling imagery and design style. This "alphabet", properly designed, results in an emotional connection between the brand and the consumer (Brunner and Emery, 2009). Visual brand language is a key ingredient necessary to make an authentic and convincing brand strategy that can be applied uniquely and creatively in all forms of brand communications to both employees and customers (Lockwood and Walton, 2008). Thus, in this research authors focus on the visual variables, which act as the visual cues in mono-brand fashion store.

3 Hypothesis and Methodology

– *Hypothesis*

Using the lens model as our core idea, we put forward the hypothesis. This model has five components: initial focal variable (brand manager), terminal focal variable (consumer), lens: a multiple mediation (mono-brand fashion store), areas of non-specificity (surrounding environment) and functional arc (feedback from focal variables). (See Fig 1)

The hypothesis is:

- The dependent variable has an actual value; each brand has its own emotional content expressed in a certain way in a mono-brand store. (Brand value in context is shown as Y_e in Table 1.)
- The participant makes a judgment about an object given information through cues. (This judgment is represented as Y_s in Table 1.)

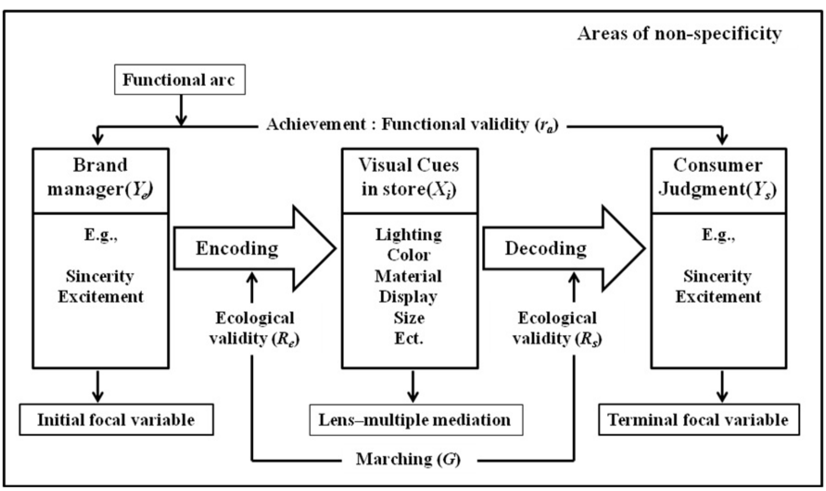


Fig. 1 The emotional-communication model in a mono-brand fashion store

- There are things in the environment that relate to our dependent variable. These things (independent variables) are called *cues* in lens-model terminology. (In our research, cues might be the objects in a mono-brand fashion store that can contribute to visual information. The cues are shown in the center of the diagram above, and labeled from X_1 to X_9.)

- There is a functional relationship between each cue and each dependent variable. (In the lens model, it is customary to talk about the linear correlation between each cue and the judgment, $\boldsymbol{Y_s}$, and also between each cue and the environment.)
- It is customary to talk about the correlation between the judgment and the environment as the achievement index, $\boldsymbol{r_a}$

Table 1 Variables and parameters of the emotional-communication model in a mono-brand fashion store

Mathematical symbols	Description
Y_e'	Predicted criterion value; the brand personalities that the brand manager expects to communicate
Y_e	Actual criterion value; the brand personality that the manager actually communicates
Y_s'	Estimated value of consumer or observer expectation of the brand's personality in a mono-brand fashion store
Y_s	Actual value of the brand's emotional content judged by consumers or observers in a mono-brand fashion store
X_i	Sensory cues carrying emotional content in a mono-brand fashion store
R_e	Correlation between actual and predicted criterion value, indicating how predictable the effect of emotional communication through the cues in a mono-brand fashion store is
R_s	Correlation between actual and predicted judgment of the brand's emotional content communicated through the cues in a mono-brand fashion store
r_a	Achievement of the organism which indicates how well consumer or observer judgment matches manager expectations
G	Matching index; correlation coefficient of Ye´ and Ys´
C	Configurable index

– ***Visual cues in the mono-brand fashion store***

After verifying the feasibility of the lens model by interviewing experts and professionals working in fashion and marketing, we committed to summarize all the visual cues in a mono-brand fashion store through study and interviews. With regard to the concept of visual cues, Berman and Evans (1995) divided atmospheric stimuli or elements into four categories: external variables, general interior variables, layout and design variables and point-of-purchase and decoration variables. Jianliu Yang (2008) discussed the visual merchandising variables of luxury brands in China. He divided visual variables into 11 categories: lighting, architectural style, size, design and allocation of space, color theme, materials, display, composition, identification marks, shop assistants, and audience. However, the human variables (the final two), especially audience, are too active and changeable for the present study, and may always differ between individuals and over time. As well, the shop assistant variable contains information mostly related to interpersonal communication and too complex for the present work. The remaining nine visual cues are broken down into 43 items (See Table 2.)

– ***Methodology***

Previous research has suggested that brand personality evokes emotions in consumers (Biel, 1993). The signal effect of a brand can be based on the image of a generalized user of the brand or on the personality of the brand itself. When using a brand, consumers come to associate it with these signals. A unique brand personality can help create a set of unique, favorable associations in the minds of consumers and thus build and enhance brand equity (Keller, 1993; Johnson et al., 2000; and Phau and Lau, 2000). Thus, in the present research, Aaker's (1997) "Brand Dimensions Scales" (BDSs) have been introduced to describe the emotional content communicated in a mono-brand fashion store. In order to describe brand personalities as concretely and accurately as possible, the semantic differential method has been utilized in the interview and consumer questionnaires. Each brand dimension scale is described by a set of traits which are measured by a range from 1 (strongly not agree) to 5 (strongly agree).

Table 2 Visual cues in mono-brand fashion store

Variable	Category	Item description
X_1	Lighting	Basic lighting, highlight lighting, decorative lighting
X_2	Architectural style	Shopping-district architecture, surrounding stores, exterior structure, internal structure
X_3	Size	Exterior walls, exterior signs, display windows, decorative props, POP (Point-of-Purchase) posters
X_4	Design and allocation of space	Height of ceiling, width of entrance, product density
X_5	Color theme	exterior walls, floor, display props, decorative props, accessory props, mannequins, products, POP posters
X_6	Materials	Exterior walls, floors, display props, decorative props, accessory props, mannequins, products, POP posters
X_7	Display	Display props, decorative props, accessory props, mannequins, products, POP posters
X_8	Composition	Division of store, window mix, organization of entrances and exits, product
X_9	Identification marks	Brand labels, functional signs (price labels, etc.)

Fashion is considered a general term for any currently popular style or practice, especially in clothing, footwear, or sartorial accessories (Cumming, 2004). And this theoretical model has been applied to a fashion brand—Tod's, which is an Italian brand dedicated to building up a strong brand identity in Asia. Questionnaires were utilized for interviews with Tod's brand managers, their assistants, and 67 consumers or observers in Shanghai to obtain relevant information.

4 Findings and Analysis

We gathered information on expected and actual value of the brand's emotional content from these interviews and questionnaires. Meanwhile, statistical methods

Table 3 Weight of Visual Variables

Variable	Category	Visual cues	Secondary index (W_{ij})	Primary index (W_i)
X_1	Lighting	X_{11} Basic lighting	0.023521	0.068845
		X_{12} Highlight lighting	0.023463	
		X_{13} Decorative lighting	0.021861	
X_2	Architectural style	X_{21} Shopping-district architecture	0.022049	0.092975
		X_{22} Exterior structure	0.024134	
		X_{23} Surrounding stores	0.021078	
		X_{24} Internal structure	0.025713	
X_3	Size	X_{31} Exterior walls	0.020577	0.130647
		X_{32} Exterior signs	0.020815	
		X_{33} Display windows	0.022148	
		X_{34} Display props	0.02301	
		X_{35} Decorative props	0.021684	
		X_{36} POP posters	0.022413	
X_4	Design and allocation of space	X_{41} Height of ceiling	0.020204	0.065413
		X_{42} Width of entrance	0.020813	
		X_{43} Product density	0.024396	
X_5	Color theme	X_{51} Exterior walls	0.023424	0.185899
		X_{52} Floors	0.023557	
		X_{53} Display props	0.021314	
		X_{54} Decorative props	0.022661	
		X_{55} Accessory props	0.020652	
		X_{56} Mannequins	0.022693	
		X_{57} Products	0.027298	
		X_{58} POP posters	0.0243	
X_6	Materials	X_{61} Exterior walls	0.020594	0.17684
		X_{62} Floor	0.020843	
		X_{63} Display props	0.021473	
		X_{64} Decorative props	0.021324	
		X_{65} Accessory props	0.021721	
		X_{66} Mannequins	0.021765	
		X_{67} Products	0.027295	
		X_{68} POP posters	0.021826	
X_7	Display	X_{71} Display props	0.022354	0.136087
		X_{72} Decorative props	0.021011	
		X_{73} Accessory props	0.020892	
		X_{74} Mannequins	0.022213	
		X_{75} Products	0.025883	
		X_{76} POP posters	0.023734	
X_8	Composition	X_{81} Division of store	0.02204	0.094119
		X_{82} Window mix	0.023696	
		X_{83} Organization of entrances and exist	0.022383	
		X_{84} Products display	0.026001	
X_9	Signs and cards	X_{91} Brand labels	0.024709	0.049174
		X_{92}Functional signs (price labels)	0.024465	

and tools (Microsoft Excel 2007 and SPSS 13.0) were used to analyze the data to form the estimating regression equation of consumer judgments and manager expectations and calculate the weight of visual cues (Table 3).

The value of emotional communication in the Tod's mono-brand stores is shown in Table 4.

Table 4 Value of emotional communication in a Tod's mono-brand store

Brand dimension scale	Y_e'	Y_e	Description	Y_s'	Y_s
(α) Sincerity	25	36	Down-to-earth, honest, wholesome, cheerful	42.31	43.66
(β) Excitement	33	48	Imaginative, spirited, up-to-date, daring	48.79	45.17
(γ) Competence	37	32	Reliable, successful, intelligent	50.74	45.21
(ε) Sophistication	54	38	Upper-class, charming	38.93	43.38
(μ) Ruggedness	60	64	Tough, outdoorsy	47.29	46.83

Then, according to the answers, emotional-communication content was assessed through consumer and observer interviews using the five categories above.

The results show that:

1. Color, material and style, especially color, are the most important visual cues;
2. The final assessed values are listed below:

3. Ecological validity (consumers/observers):

$$R_s = r_{Y_sY_s'} = 0.737$$

1. Ecological validity (managers):

$$R_e = r_{Y_eY_e'} = 0.558$$

1. Matching index:

$$G = r_{Y_s'Y_e'} = -0.168$$

1. Configurative index:

$$C = r_{[Y_s - Y_s'][Y_e - Y_e']} = -0.204$$

1. Functional validity:

$$r_a = GR_s\mathrm{R_e} + C\sqrt{1-R_s^2}\sqrt{1-R_e^2} = -0.184$$

1. From these results, it is obviously that R_s gains more value than R_e, which means that the correlation coefficient of consumers is higher than that of managers. Matching index and functional validity are both low, indicating that emotional communication through visual cues in the mono-brand store

2. involves imperfect communication between managers and consumers, which might be mostly due to managers' decision-making processes.

4. Furthermore, the regression equations of the expectation on both sides of "the lens" have been calculated below. With these equation any changes in the visual cues of Tod's mono-brand may lead to the different value of the expecation.

$$Y_e' = 7.94 - 0.259 f(Architecture) + 0.012 f(Style) - 0.116 f(Color) - 0.720 f(Material)$$

$$Y_s' = [Y_\alpha', Y_\beta', Y_\gamma', Y_\varepsilon', Y_\mu']$$

$$Y_\alpha' = 0.291 - 0.045 X_{\alpha 1} + 0.102 X_{\alpha 2} + 0.056 X_{\alpha 3} + 0.065 X_{\alpha 4} + 0.205 X_{\alpha 5} + 0.217 X_{\alpha 6} + 0.207 X_{\alpha 7} - 0.031 X_{\alpha 8} + 0.007 X_{\alpha 9}$$

$$Y_\beta' = 0.319 + 0.038 X_{\beta 1} + 0.055 X_{\beta 2} + 0.118 X_{\beta 3} - 0.006 X_{\beta 4} + 0.233 X_{\beta 5} + 0.160 X_{\beta 6} + 0.003 X_{\beta 7} + 0.112 X_{\beta 8} + 0.056 X_{\beta 9}$$

$$Y_\gamma' = -0.383 + 0.185 X_{\gamma 1} + 0.126 X_{\gamma 2} - 0.099 X_{\gamma 3} + 0.076 X_{\gamma 4} - 0.002 X_{\gamma 5} + 0.244 X_{\gamma 6} + 0.121 X_{\gamma 7} + 0.285 X_{\gamma 8} + 0.060 X_{\gamma 9}$$

$$Y_\varepsilon' = -0.575 + 0.05 X_{\varepsilon 1} + 0.127 X_{\varepsilon 2} + 0.006 X_{\varepsilon 3} - 0.181 X_{\varepsilon 4} + 0.119 X_{\varepsilon 5} + 0.362 X_{\varepsilon 6} + 0.247 X_{\varepsilon 7} + 0.233 X_{\varepsilon 8} + 0.060 X_{\varepsilon 9}$$

$$Y_\mu' = -0.673 + 0.063 X_{\mu 1} + 0.112 X_{\mu 2} + 0.016 X_{\mu 3} - 0.182 X_{\mu 4} + 0.116 X_{\mu 5} + 0.383 X_{\mu 6} + 0.265 X_{\mu 7} + 0.232 X_{\mu 8} - 0.059 X_{\mu 9}$$

5 Conclusion, Limitations and Future Directions

The use of a lens-based emotional-communication model in a mono-brand fashion store has been demonstrated here. This model focuses strongly on the need to respond to cognitive differences between managers and consumers and, for managers, emphasizes achieving efficient emotional communication with consumers. One fashion brand has been analyzed, and the feasibility of this emotional-communication model has been shown.

This research provided a new approach to analyzing the effectiveness of visual communication from both brand and consumer perspectives, and calculated the estimated regression equations of consumer expectation of fashion brand personality.

However, the future of fashion branding will be far fiercer than ever before. The emotional-communication model should be modified to include all five sensory channels and should be applied to other brands to get more general constants for the regression equation. In addition, the rhetoric used in this research was designed with a certain research group in mind, so the accuracy of their descriptions of brand personality might be influenced. And since in this research only one fashion brand was assessed, more brands should be analyzed so as to get a more general regression equation.

References

1. Aaker, J.L.: Dimensions of brand personality. Journal of Marketing Research 35, 347–356 (1997)
2. Berman, B., Evans, J.R.: Retail management: A strategic approach. Prentice-Hall, Engle-wood Cliffs (1995)
3. Brunner, R., Emery, S.: Do you matter? How great design will make people love your company. FT Press, Upper Saddle River (2009)
4. Brunswik, E.: The conceptual framework of psychology. The University of Chicago Press, Chicago (1952)
5. Brunswik, E.: Perception and representative design of experiments. University of California Press, Berkeley (1956)
6. Bucci, W.: Pathways of emotional communication. Psychoanalytic Inquiry 21, 40–70 (2001)
7. Clow, K.E., Baack, D.: Integrated advertising promotion and marketing communications, 3rd edn., Upper Saddle River (2007)
8. Cosmides, L., Toody, J.: Consider the source: The evaluation of adaptations for coupling and metarepresentation. In: Sperber, D. (ed.) Metarepresentations, pp. 53–115. Oxford University Press, New York (2000)
9. Cumming, V.: Understanding fashion history. Costume & Fashion Press, Hollywood (2004)
10. Finn, A., Louviere, J.: Shopping-center patronage models fashioning a consideration set and segmentation solution. Journal of Business Research 21, 259–275 (1990)
11. Hawkins, D.I., Coney, K.A.: Consumer behavior: Building marketing strategy. McGraw-Hill College, New York (2004)
12. Henderson, P.W., Spangenberg, E.R., Crowley, A.E.: Improving the store environment: Do olfactory cues affect evaluations and behaviors? Journal of Marketing 60(2), 67–81 (1996)
13. Simon, J., McMillan, J., Woodruff, C.: Why do Firms Hide. Journal of Public Economic 76, 495–520 (2000)
14. Juslin, P.N.: Cue utilization in communication of emotion in music performance: Relation performance to perception. Journal of Experimental Psychology: Human Perception and Performance 26, 1797–1813 (2000)
15. Keaveney, S.M., Hunt, K.A.: Conceptualization and operationalization of retail and store image. Journal of Marketing Science 20(2), 165–176 (1992)
16. Keller, K.L.: Conceptualizing, measuring, and managing customer-based brand equity. Journal of Marketing 57, 1–22 (1993)
17. Lockwood, T., Walton, T.: Building design strategy. Allworth Press, New York (2008)
18. McDaniel, S.W., Burnett, J.J.: Consumer religiosity and retail store evaluative criteria. Journal of the Academy of Marketing Science 18(2), 101–113 (1990)
19. Mehrabian, A., Russell, J.A.: An approach to environmental psychology. MIT Press, Cambridge (1974)
20. Phau, I., Lau, K.C.: Conceptualising brand personality: A review and research propositions. Journal of Targeting, Measurement, and Analysis for Marketing 9(1), 52–69 (2000)
21. Tartaglione, C.: Il sistema Moda verso il rinnovamento strategico in un mercato dove sono strutturalmente cambiate le leve della competizione, Mimeo, Rome (2005)
22. Yang, J.: Visual merchandising for luxury brand store based on visual communication. Donghua University, MA (2008)

A Proposal of a Japanese Input System for Mobile Terminals Enhanced by the Word Forecast Function Based on the Location Information

Ken Tarusawa, Jun Sawamoto, Eiji Sugino, and Norihisa Segawa

Abstract. Mobile terminals are now very familiar electronic devises for our daily life. Especially, the usage of the mobile phone keeps expanding year by year. And, the use of e-mail function came to the first place surpassing the call function on the present mobile phones. The importance of the Japanese input system increased in connection with it. In this paper, the speed of the character input is accelerated by increasing the predictive accuracy of the Japanese characters input function. The purpose of our research is to display words frequently input in the present location in the high rank of the forecast candidate's list, and improve the character input efficiency. The system is composed of two component systems. One is the Japanese-language input system "CocoIME" that operates on the user's Android terminal. Another one is the automatic dictionary generation system "KNDS" that operates on the server. We conduct a preliminary experiment to evaluate the efficiency of the proposed CocoIME input system and discuss analyses and remaining future works.

Keywords: Japanese input system, Mobile phone, Word prediction, Location information, GPS.

1 Introduction

Mobile terminals are now very familiar electronic devises for our daily life. Especially, the usage of the mobile phone keeps expanding year by year. The diffusion rate increased to 91.6% as of December, 2010 in Japan [1]. And, the use of e-mail function came to the first place surpassing the call function on the present mobile phones [2]. The importance of the Japanese input system increased in connection with it.

Ken Tarusawa · Jun Sawamoto · Eiji Sugino · Norihisa Segawa
Faculty of Software and Information Science, Iwate Prefectural University, IPU
Takizawa, Iwate, 020-0193, Japan
e-mail: g231j026@s.iwate-pu.ac.jp,
{sega,sawamoto}@iwate-pu.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 181–189.
springerlink.com

In this paper, the speed of the character input is accelerated by increasing the predictive accuracy of the Japanese characters input function. Up to now, the words input had been predicted from the past operation record and the text context. However, researches that change the prediction candidate according to the user's situations are carried out recently. These days, mobile phones became multi-functional and it became easy to understand user's situation by using various sensors. There are existing systems that generate forecast candidate words from location information [3][4]. However, it is not widely used because the dictionary creation is very time-consuming, and the size of the dictionary tends to be large. Moreover, the possibility of "words related to the location or place" is exactly "words that the user wants to input" is low.

The purpose of our research is to display words frequently input in the present location in the high rank of the forecast candidate's list, and improve the character input efficiency. The existing system [4] displays shop names and landmarks etc. in the surrounding of the present location as the forecast candidates. On the other hand, the words of the high frequency input in the present location in the past are displayed as the forecast candidates in this research. Moreover, the volume of data to read is suppressed to the minimum by efficiently managing huge dictionaries based on location information, and choosing selecting dictionaries to be read very selectively.

2 System Configuration

Japanese is a language that needs many kinds of characters compared with other languages. For example, 83 hiragana, 86 katakana, 10 figures, 52 Roman alphabets, and 6355 Kanji characters or more are used. When the user inputs a Kanji character, the pronunciation is input first. Next, the user pushes the conversion button. Then, words are displayed in the order of high possibility that fits the user's input request in the conversion candidate field. Finally, the user chooses a word from the conversion candidates. This work is very cumbersome. Candidates forecast are displayed even if all the pronunciations are not completed if the system has a smart capability. However, when the forecasting algorithm is poor, the word requested is not displayed in the high rank location of the candidate field.

In this research, the system tries to display words that user intend to input in higher rank of forecast candidate field. The system is composed of two component systems. One is the Japanese-language input system "CocoIME" that operates on the user's Android terminal. Another one is the automatic dictionary generation system "KNDS" that operates on the server.

2.1 The Japanese-Language Input System and the Automatic Dictionary Generation System

CocoIME is enhanced by adding the function that the forecast conversion candidates are displayed according to the present location, to the existing ordinary

Japanese-language input system. First, when the user inputs the character with CocoIME, "Location information" and "Input words" are output as a log. Logs are regularly up-loaded to the server on which KNDS is installed. KNDS collects logs, and examines whether there are relations in "Location information" and "Input words". The dictionary is automatically generated as a result. CocoIME regularly downloads the generated dictionary. Fig. 1 shows the system overview.

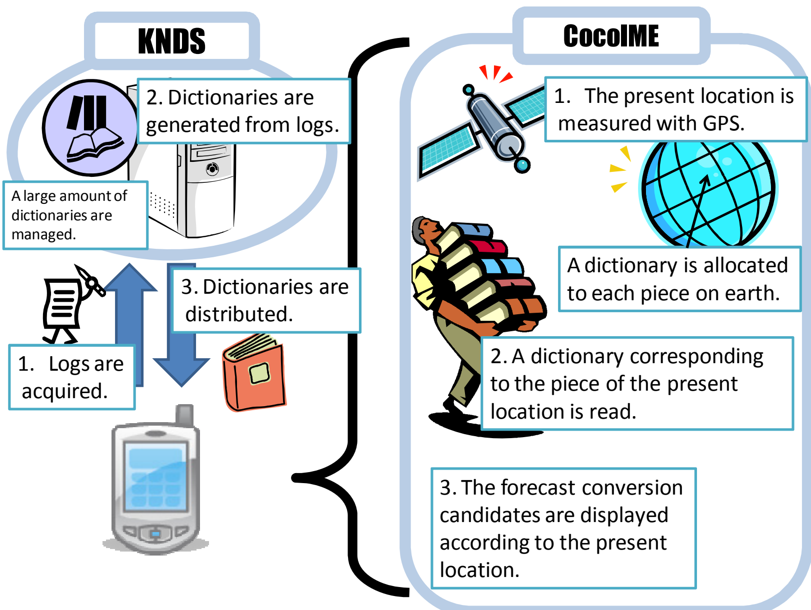


Fig. 1 CocoIME is enhanced by adding the function that the forecast conversion candidates are displayed according to the present location, and logs are regularly up-loaded to the server on which KNDS is installed. KNDS collects logs, and examines whether there are relations in "Location information" and "Input words". CocoIME regularly downloads the generated dictionary from KNDS.

2.2 Relation of Location Information and Input Words

We explain location and input words' relations. First, we draw lines on the earth by the spacing of latitude and the longitude 0.0005 degrees. We call this work "Gridding". Then, quadrangle and triangular shapes are made. These shapes look rather like trapezoidal figures. We call this figure "a piece" hereafter. This system can make dictionaries up to the number of pieces in the mobile phone. In the first state, the number of dictionaries in the terminal is 0. This system makes 532.9 billion dictionaries maximum in theory. However, the number of dictionaries is usually small, because a person's moving range is limited to some degree.

Next, we explain the flow that registers words to dictionaries. First, the system regularly measures location information with GPS. As a result, the system gets latitude and the longitude of the user's location continuously. The user inputs the character string through CocoIME interface, and the input character string is recorded in the dictionary of the corresponding peace which covers the present location. The system keeps location information and input character string as a log. The system regularly up-loads logs to the server KNDS.

The user can change the interval of acquisition of location information at any time. In the default state, the update interval is one minute. The reason to assume the default interval to be one minute is that it is thought that this value is appropriate considering the speed along which man walks and the system specification. The speed on foot is 4.8 kilometers per hour according to the real estate rule. The distance between the center of the piece and the center of the next piece is about 55 meters. The time necessary for the movement from a center to the next center is 41.25 seconds. And, the time the portable terminal is taken out, and the time the mail screen is started are added, the total is about one minute. Therefore, we thought that one minute interval for location update is reasonable.

2.3 Decision of Forecast Candidate's Priority

In the following, the flow of the system when characters are input is described. First, CocoIME downloads latitude and the longitude at the present location every minute by using GPS built in the mobile phone. It downloads, from KNDS server, dictionaries corresponding to pieces which are located within latitude and longitude ±0.01 degrees from the present location. Only when the update timestamp of the dictionary on KNDS is newer than that of the previously acquired dictionary, CocoIME acquires the new dictionary. As a result, CocoIME can keep dictionaries of pieces within about 1.1 kilometers in the radius on the map in each terminal at any time. When KNDS cannot be used, dictionaries from KNDS are just ignored and not used.

When the user inputs characters, the system uses dictionaries on the terminal. The system regularly measures the current location, and reads pieces' dictionaries corresponding to the present location and the surrounding areas. Authors are now examining optimal ranges of dictionaries of pieces that the system should read in. In this paper, the system reads 13 pieces. Words and the frequency values (value that shows how often the word is input) are registered in dictionaries. The system displays the forecast candidate words of 1.0 or more the value which is derived by multiplying the frequency value and the constant value determined by forecast candidate's priority as shown in Table 1. As shown in Figure 2, the priority value changes depending on the distance from the piece of the present location. Figure 2 and Table 1 show the image of priority assignment and the constant value assigned to each priority value.

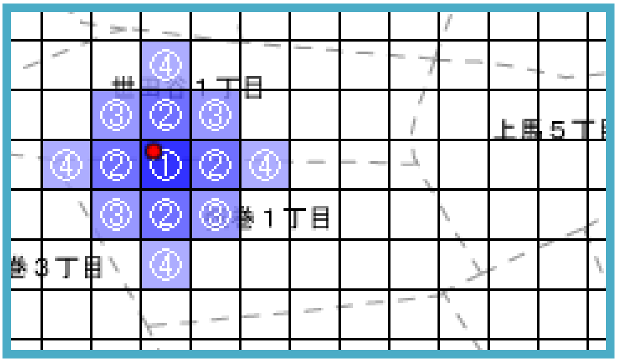


Fig. 2 An example view of the “Gridding”. The priority value changes depending on the distance from the piece of the present location.

Table 1 Forecast candidate's priority assignment and associated constant definition

Priority	Place of read dictionary	Constant
1	Piece at present location （①）	1.0
2	Pieces that are located one up and down and right and left from the present piece （②）	0.7
3	Pieces that are diagonally located from the present piece （③）	0.4
4	Pieces that are located two up and down and right and left from the present piece （④）	0.1

2.4 Automatic Dictionary Generation

KNDS is a system that automatically generates new dictionaries from logs of CocoIME. Dictionaries are generated using words that a lot of users input frequently. As a result, CocoIME can display words that other people have input as forecast candidates. In this paper, an experiment only on CocoIME is discussed and the detailed explanation of KNDS is omitted.

3 Accuracy of the Location

3.1 Accuracy of GPS

The portable terminal used by the experiment is "GALAXY S SC-02B" made by Samsung Electronics Co., Ltd. The career is NTT DoCoMo. Inside a room, the accuracy of GPS of the terminal is from about 10 to 50 meters and it is from about 2 to 10 meters outdoors. The accuracy of GPS depends on the terminal device and the environment, etc. Therefore, we consider the accuracy of GPS in this research.

3.2 Difference of Length of the Longitude

When the longitude is delimited at an interval of 0.0005 degrees, the width of a piece becomes about 55 meters on the equator, and the piece is almost square. However, when the absolute value of latitude increases, it becomes a trapezoidal figure because the length of longitude shortens when the absolute value of latitude increases. Present CocoIME disregards this phenomenon. The solution of this problem is left as one of the future tasks.

3.3 Error Margin by Centrifugal Force of the Earth

The earth is almost the same as the shape called the earth ellipsoid instead of a perfect sphere. Therefore, even if the latitude of the earth is simply delimited at an interval of 0.0005 degrees, it never becomes constant length. However, the cost performance worsens very much when the system changes the sizes of pieces by the location in consideration of the earth ellipsoid. Therefore, the error margin is understood, and the system doesn't reflect it in the specification in this research.

3.4 Error Margin by Elastic Deformation of Celestial Body Gravitation

The earth figure is not a globe (Refer to 3.3). However, even if the delimitation interval between latitude and the longitude is changed according to the earth ellipsoid, length is not constant because the earth is deformed by the gravitation of the moon and the sun. This problem can be solved by calculating the amount of the horizontal change on the surface of the earth called Shida’s number. However, this error margin is not reflected in the system in consideration of the cost performance just like 3.3.

4 Experiment

4.1 Experimental Method

In this experiment, KNDS is not used, and only CocoIME is used. The dictionary is not shared with other terminals, and the experiment is carried out with only one terminal.

Experimental conditions are as follows.

1) The test subject (testee) is one person.
2) The testee inputs Japanese mail of about 50 characters (error margin ± ten characters) written in kanji and kana.

3) A group of 100 mails; 50 of them are mails that the testee inputs during a trip and 50 of them the testee inputs at home.
4) First, learning function of the dictionary is initialized. Then, the testee inputs 100 mails without using the function to display the word forecast based on the location information. Afterwards, the testee inputs 100 mails by using the function to display the word forecast based on the location information.
5) The input mode is a character circulation specification method.
6) The terminal used to experiment is GALAXY S SC-02B.
7) The contract career of the terminal is NTT DoCoMo.

The testee inputs 100 mails of about 50 characters. Among 100 mails, 50 mails seem to be related to the location. The testee input the mails that had been input at the school (University), dining out (Miyagi Prefecture Sendai City), shopping (Tokyo Taito Ward, Akihabara Ward), job hunting (Aichi Prefecture Nagoya City), and the concert (Sapporo City, Hokkaido), etc. with the terminal used for the experiment. These mails are called "Travel mails". The other 50 mails seem not to be related to the location. The testee input the mail at home. These mails are called "Home mails".

The input mode uses the character circulation specification method that is the most general input mode in recent mobile phones. Location information is not a value from an actual GPS receiver. Location information was specified directly. Therefore, there is no margin of error by the GPS receiver in this experiment.

Here, we describe the method of counting the number of key strokes. First, the testee inputs the first character. If the word to be input is displayed as the forecast conversion candidate, it is selected. The number of key strokes changes by on what position of the forecast conversion candidate is displayed. If displayed on the Nth, N will be added to the number of key strokes. Otherwise, the testee inputs the second character, and looks up at the word from among the forecast conversion candidates. The testee counts the number of key strokes while repeating this operation. GALAXY S mounts a physical button (home button, back button, menu button, volume control button, and power-on/off button). However, these buttons are not used while experimenting. The testee chooses the word displayed as the forecast conversion candidates with the touch panel. And the testee actually does the tap once for choosing the candidate. However, it is important to be displayed on the higher rank of the forecast conversion candidate as the experiment. Therefore, the count of the number of key strokes is decided according to the candidate's turn. Signs, such as pictographs are not used in this experiment. A number or the change of input mode is counted as one key stroke. Authors experimented in the above-mentioned environment.

4.2 Experimental Results

Table 2 and Table 3 show the result of the experiment. Forecasting words by using location information decreases and the number of key stroke and the input time

Table 2 Comparison of Numbers of Key Strokes. Forecasting words by using location information decreases and the number of key stroke and the input time have improved. But the value of the standard deviation increased by using location information.

		Home mails	Travel mails	All mails
Average(times)	Not using location information	133.7	138.7	136.2
	Using location information	123.2	122.0	122.6
Standard deviation	Not using location information	13.50	16.49	15.20
	Using location information	21.24	27.97	24.72

Table 3 Comparison of Input Times. Forecasting words by using location information decreases and the input time have improved. But the value of the standard deviation increased by using location information.

		Home mails	Travel mails	All mails
Average(second)	Not using location information	73.9	77.9	75.9
	Using location information	62.0	56.1	59.0
Standard deviation	Not using location information	9.18	10.38	9.94
	Using location information	21.94	24.50	23.34

have improved. However, input efficiency did not improve in all mails, and input efficiency has decreased in some of the mails. That is the reason why the value of the standard deviation increased by using location information.

4.3 Considerations

It was expected that the proposed system is effective when the words that relate to the location was input. Travel mails correspond to it. On the other hand, it was expected that it has the opposite effects for home mails. However, the system was effective in both travel mails and home mails in the result of the experiment. It could be thought that the reason for the home mails is that there is a problem in the content of the home mail. The testee stays home very little and the contents of mails at home are restricted to some extent. In such a situation it is thought that a house is a one of the trip destinations. Therefore, it is thought that the system was able to forecast words that the testee inputs to some degree. It is expected that the contents of mails become varied, and input efficiency decreases if the tetee stays at home longer.

The number of key strokes has decreased on a large scale many times for travel mails. However, it has rarely happened for home mails. The reason why the number of input key strokes sometimes increased by the travel mails is that there are a lot of proper nouns. Therefore, when a proper noun is input for the second time and afterwards, the number of key strokes is decreased easily if the function to forecast the word from location information is used.

Moreover, the standard deviation was high when the proposal technique was used. Words forecast by the proposal technique were occasionally inappropriate. As a result, input efficiency improved greatly, or it fell and the value varied by that cause.

5 Conclusion

It has been understood that it is useful to use position information when the word is forecast in the Japanese-language input system. The system displays not only names of places and landmarks but also words frequently input as the forecast candidates. As a result, it is effective for E-mails which seem not to be related to the place either. However, the function to forecast words from location information can not necessarily put out the great result. It is expected that input efficiency improves further if the forecast function can be turned on and off according to the situation.

There is a problem with sizes of pieces generated from the gridding. To cope with this problem, it is necessary to change the way to delimit the longitude in proportion to latitude. A concrete solution is to lengthen the delimitation spacing of the width of the piece away from the equator. As a result, the size of a piece is always made the shape near the square.

In this experiment, only CocoIME is used, so the dictionary is not ready in the place visited for the first time. Therefore, input efficiency is the same as the usual input system in the place first visited. Other users' input histories are displayed by the forecast system by simultaneously using KNDS. As a result, it is expected that more efficient input can be achieved even in the place first visited.

As the future works, authors will aim at a further improvement of input efficiency by improving CocoIME, and at the same time conducting the experiment that introduces KNDS.

References

1. The Ministry of Internal Affairs and Communications, Information and Communications Statistics Database, `http://www.soumu.go.jp/johotsusintokei/field/data/gt01020101.xls` (access January 23, 2011)
2. Report of the research on the most frequently used functions on the mobile phone, iSHARE.Co. (2009), `http://release.center.jp/2009/12/2201.html` (in Japanese)
3. Tsuchida, M.: Portable Terminal Equipment and Character String Processing Method. Published patent application, Publication number: 2007-65906 (March 15, 2007)
4. Arakawa, Y., Suematsu, S., Tagashira, S., Fukuda, A.: Dynamic Dictionary Generation Method for Context-aware Input Method Editor. IPSJ Journal 52(3), 1033–1044 (2011)

Activity Recognition by Fuzzy Logic System in Wireless Sensor Network for Physical Therapy

Shu-Yin Chiang, Yao-Chiang Kan, Ying-Ching Tu, and Hsueh-Chun Lin

Abstract. The physical therapy for geriatrics training or stroke patients requires incessant and routine rehabilitation during the cure period. The physiatrists hereby refer feedback from clinical records to offer necessary assistant programs. The ubiquitous health care (UHC or u-healthcare) becomes the most concern of the successful treatment that needs to ensure patients following the therapeutic assignment continuously. This study proposes a facile activity recognition procedure to interact patients and computation for measuring essential movements of human body with privacy concern through wireless sensor network (WSN) body motion sensors that involves the accelerometer and gyroscope. At this initial stage, sensor data of static postures and dynamic motions are recognized by the fuzzy algorithm. According to the proposed process, the fuzzy parameters are calibrated by the adoptive feature sets and are verified by a blind test. The overall recognition accuracy for regularly steady activities achieves over 96%. Two simple rehab postures of physical therapy were discussed and the recognition rate can imply the threshold of specific rehab activity. The approach may support the interface to monitor privately remedy process for patients with non-imaged and non-invasive u-healthcare of physical therapy.

Keywords: Accelerometer, Gyroscope, Fuzzy Logic, Activity Recognition, Ubiquitous Health Care, WSN.

Shu-Yin Chiang · Ying-Ching Tu
Department of Information and Telecommunications Engineering,
Ming Chuan University
5 De Ming Road, Gwei-Shan, Taoyuan 333, Taiwan
e-mail: sychiang@mail.mcu.edu.tw

Yao-Chiang Kan
Department of Communications Engineering,
Yuan Ze University, Chung-Li, Tao-Yuan 320, Taiwan

Hsueh-Chun Lin
Department of Health Risk Management,
China Medical University, Taichung 404, Taiwan

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 191–200.
springerlink.com

1 Introduction

The activity detection is one of the most important issues of the ubiquitous health care (UHC) system, particular for monitoring daily motions of aged people or hemiplegia patients who could unintentionally fall down. Based on the physical therapy and medicine, routine and continuous repeats of the specific motions are typically assigned for geriatrics training or stroke patients during the remedy period of rehabilitation treatment. If the physiatrists can completely acquire the rehabilitation records of patients during the cure process through UHC (or u-healthcare) with ambulatory measurement, they must be more aware of designing assistant programs. The activity recognition therefore can be a prospective technique to achieve this scope. Human beings essentially take some typical movements such as “lie,” “sit,” “stand,” “walk” and so on. These activities can be captured and recognized in different circumstance by appropriate techniques [1]. In general, the methodology to identify movements of human body can be classified by image [2] and non-image [3] procedures. By considering the privacy of human right, the cost of equipments, and convenience in an open environment, it is preferred to adopt the non-image technology for u-healthcare. The u-healthcare related to rehabilitation requirement is conceptually extended from the homecare personal area network that provides intelligent monitoring and supporting hospital functions with ad hoc network [4-7]. In the past decade, due to rapid advances of technology in micro-electro-mechanical systems (MEMS), integrated circuit (IC), and radio frequency (RF), the wireless sensor network (WSN) [8] has been widely spread out for different types of healthcare because it can be applied for detecting various physiological signals such as blood pressure, impulse, motion as well as variation of the home environment [9]. Many studies approved this emerging technique on a variety of the home health care systems [10] since it provides wearable, portable, and mobile functionalities for usage to ubiquitously combine patients’ healthcare information with clinical data in hospital [11]. For this approach on the physical therapy and fundamental rehabilitation, the activity recognition procedure of WSN data can be studied to measure the basic motions such as “lie,” “sit,” “stand,” “walk,” and “run” for further application.

In this study, we design the wearable sensor by integrating the accelerometer and the gyroscope with the WSN devices to detect signals of body movements. For the measurement, the testers wear two self-developed WSN inertial sensors node (ISN) on chest and thigh to transport required signals [12]; thus the fuzzy algorithm is applied for advanced data calibration. The proposed procedure is further promoted for measuring two fundamental rehabilitation activities. Herein, the configuration of WSN devices will be designed in the next section followed by necessary methodology that can adopt feasible feature sets of body movement due to the data acquired by the sensors. Successively, the strategy of activity recognition based on the fuzzy algorithm is explained to calibrate sensed data and obtain output movement. Finally, the test results are discussed and conclusion remarks are addressed.

2 System Implementation

2.1 Hardware Implementation

One of the primary assistance of u-healthcare for the physical therapy is allowing physiatrists to continuously trace clinical data of rehabilitation for stroke patients or training aged people so that patients' disability can be surely rehabilitated in the cure. Essential motions of rehab patients usually include "lie," "sit," "stand," "walk," and "run" which imply spatial coordinates of the specific movement during a period thus the gyroscope and accelerometer can reach this scope. Both of the MEMS-enabled chips are increasingly embedded in the portable electronic device to detect the position of the user and make the ambulatory measurement possible by transporting data through the WSN to approve the u-healthcare for physical therapy.

The WSN mote shown in Fig. 1 (left) supports embedded micro control unit (MCU), radio frequency (RF) and antenna to process and deliver the sensor signals wirelessly. The accelerometer and gyroscopes shown in Fig. 1(right) can detect the acceleration and angular velocity, respectively, corresponding to the local coordinate system of sensor. In addition, a flash memory is installed to store sensed data temporarily prior to transportation. Based on this design, the proposed WSN sensor provides wireless modules that enable low energy consumption for surveillance safety of human body behavior. It is compliant with the TinyOS embedded system [13] and adapts typical network protocols to the backend interface for remote control and data analysis with convenience and efficiency.

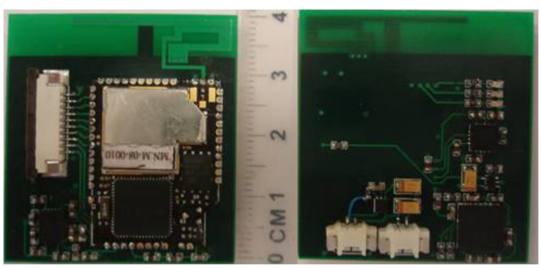


Fig. 1 WSN mote (left), Accelerometer and gyroscope modules (right)

2.2 Data Calibration

In this study, two WSN sensors are considered for measuring specific body motions. The tester can wear the sensor A and B on the right chest and the left thigh shown in Fig.2, respectively, to measure the relevant data including acceleration and angular velocity. In which, the global coordinate system defines z axis as the forward direction while the directions of vertical to ground and lateral to body are represented by x and y axes, respectively. Meanwhile, six data streams of acceleration and angular velocity along the 1, 2, and 3 axes of the sensor can be obtained corresponding to the local coordinate system of sensor. Herein the static posture "stand" is set as the initial status, i.e. the axis-1 of sensor A and B are approximately vertical and horizontal, respectively, to the ground in the beginning. Let g_A, g_B, ω_A, and ω_B denote the vectors of acceleration and angular velocity measured by the sensor A and B, respectively, at an arbitrary moment. In which, each vector consists of three components in three axes, i.e. $g_A = (g_{A1}, g_{A2}, g_{A3})$, $g_B = (g_{B1}, g_{B2}, g_{B3})$, $\omega_A = (\omega_{A1}, \omega_{A2}, \omega_{A3})$ and $\omega_B = (\omega_{B1}, \omega_{B2}, \omega_{B3})$. The variation of acceleration and angular velocity can be retrieved as diagrams in Fig.3.We hence gather a set of sensed data distributed during a period, e.g. one second, and take their mean value as the acceleration or angular velocity per unit period.

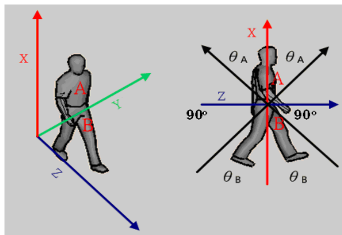


Fig. 2 Location of the sensors at the body and relationship of tilt angles

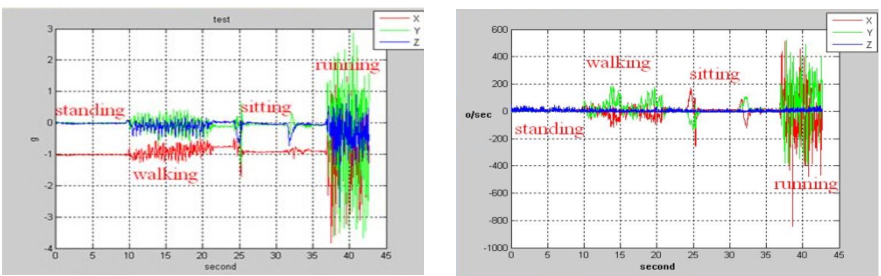


Fig. 3 Variations of acceleration in gravity (left) and angular velocity in degree/sec (right) detected by mote A

2.3 Feature Extraction

The parameters for feature extraction may not be the original raw data from the accelerometer and gyroscope. For recognizing the activity variation, the modified parameters below are extracted for the fuzzy activity recognition [14].

(1) Acceleration vector: g_A and g_B indicate the acceleration vectors of the sensor A and B, respectively.

(2) Angular velocity vector: ω_A and ω_B are the angular velocity vectors of the sensor A and B, respectively.

(3) Tilt angle: θ_A and θ_B mean the tilt angles of the sensor A (at the chest) and B (at the thigh), respectively.
(4) Standard Deviation: σ_{gA}, σ_{gB} and $\sigma_{\omega A}$, $\sigma_{\omega B}$ indicate the standard deviation of acceleration and angular velocity of the chest and thigh, respectively.
(5) ν_{gAR} and ν_{gBR}: the difference rate of relative acceleration of the chest and thigh, respectively.
(6) γ_{gA}, γ_{gB} and $\gamma_{\omega A}$, $\gamma_{\omega B}$: the gradient of acceleration and angular velocity of the chest and thigh, respectively.

3 Proposed Fuzzy Inference System

The above features will be selected for different rules of the fuzzy algorithm reciprocally to calibrate measured data as well as compare their feasibility. The proper features are firstly classified as fuzzy sets and became the input features of the fuzzy system [1, 14-16]. After fuzzification, we then induced the fuzzy rules which yield the activity recognition patterns in the future. Consequently, with defuzzification, we may carry out the output features that stand for static postures or dynamic motions, and create the patterns for activities. The procedure of the fuzzy algorithm is described by the following steps: Fuzzification, Rule Evaluation, and Defuzzization and discussed in the next subsection.

3.1 Fuzzification

In the fuzzification, the input parameters, tilt angles, gradient of acceleration and angular velocity and standard deviation of acceleration are fuzzified using predefined input membership functions as shown in Figures 4, 5, 6 and 7, respectively.

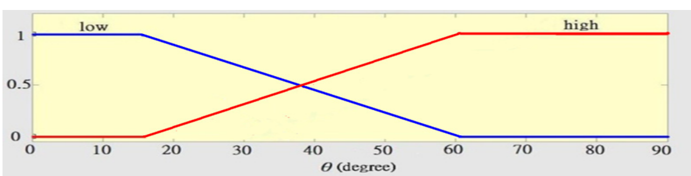


Fig. 4 Membership Function of θ_A and θ_B

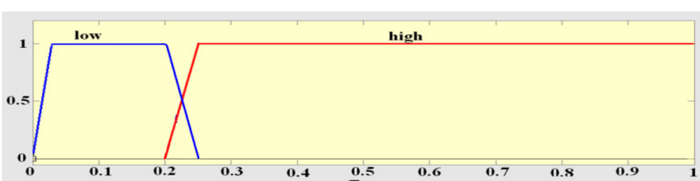


Fig. 5 Membership Function of σ_{gAx}

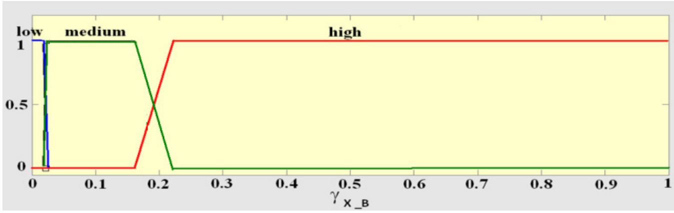


Fig. 6 Membership Function of γ_{gBx}

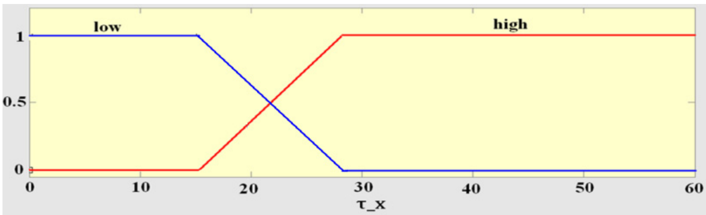


Fig. 7 Membership Function of $\gamma_{\omega Ax}$

3.2 *Rule Evaluation*

In this work, we use fuzzy rules for the activity recognition pattern. These rules are generated by the fuzzy logic with "if-then-else" syntax to recognize the input features and carry out the output features; e.g., if (θ_A = "low-angle", θ_B = "high-angle", γ_{gBx} = "low" and $\gamma_{\omega Ax}$ ="low" with neglect of σ_{gAx}) then output = "sit". The set of fuzzy rules is listed in Table 1. Herein, the membership function of output feature for activities such as "lie," "sit," "stand," "walk," and "run" can be quantified by using the triangle distribution as shown in Fig.8.

Table 1. Fuzzy rules for defuzzification

θ_A	θ_B	σ_{gAx}	γ_{gBx}	$\gamma_{\omega Ax}$	Output
L	L	-	L	L	Stand
L	H	-	L	L	Sit
H	H	-	L	L	Lie
-	-	L	M	H	Walk
-	-	H	H	H	Run

3.3 *Defuzzification*

It is necessary to convert the resultant fuzzy decision sets into a precise quantity. We use Centroid Of Area (COA) to perform the defuzzification. The corresponding membership of the output is defined in Fig.8.

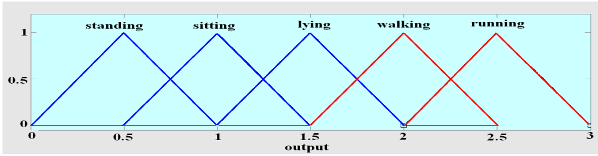


Fig. 8 Membership Function of output

4 Performance Evaluations

In this study, ten laboratory members wore the developed WSN sensors on their chests and thighs to deliver body motion data to test the procedure above. In the test, seven of them obeyed the specific order of activities to determine the membership functions and fuzzy rules and these recognition parameters were verified by the rest ones for the blind test.

The samples of acceleration and angular velocity components corresponding to various movements are detected by the sensor A and B. The input feature sets including {θ_A, θ_B, σ_{gBx}, γ_{gBx}, $\gamma_{\omega Ax}$} are extracted for the fuzzy rules which determine the output features to represent "lie," "sit," "stand," "walk," or "run". The testers are trained to follow the same postures and motions orderly for calibrating necessary fuzzy parameters and creating the recognition logics in Table 1. These parameters can detect the regular activities perfectly excluding irregularly changing status. By substituting this pattern in a blind test, the recognition results are shown in Table 2.

By comparing output data and real motions in the blind test, we found good recognition rate for static postures and lose accuracy for dynamic motions, averagely 99% versus 93%; but imperfection as changing the status which is not our scope though in this study. It is observed that most of stable and routine movements can be recognized successfully; for instance, the static postures and dynamic motions that obey the identical sample actions. If the testers are restricted completely by expected rules, their movements can be well recognized. These results approve the accuracy of the developed devices and the proposed procedure in this study.

Table 2 The successful rate (%) of activity recognition

Lie	Sit	Stand	Walk	Run	Average
Sample Test					
100	100	100	100	100	100
Blind Test					
98	100	100	96.31	90.48	96.96
static posture			dynamic motion		
99			93		

In addition, we are further interested whether the procedure is feasible to the simple rehabilitation of physical therapy or not. As shown in Fig.9, we utilized the proposed procedure by wearing the sensors marked by C and D at the wrist and the upper arm, respectively, to detect two fundamental rehab activities: (a) raise the wrist to the elbow position at 45^o, and (b) straighten up the arm over the head with 45^o step-by-step ($0^o \to 45^o \to 90^o \to 135^o \to 180^o$). Relative to other input features, as shown in Fig.10, tilt angles are observed significantly for both of restricted motions in repeat per five seconds. The feature set $\{\theta_C, \theta_D\}$ follows the fuzzy logic of previous $\{\theta_A, \theta_B\}$ to recognize angles of the activity on threshold $\{40^o \sim 45^o, 85^o \sim 90^o, 130^o \sim 135^o, 175^o \sim 180^o\}$. The result is shown in Table 3 that records qualified angles versus required counts of the rehab posture. We learn that the requested routine activity can be recognized by counting how many times the tester moves the elbow to the right position. Contrarily, if the tester does not obey the therapy, the motion cannot be recognized and qualified. For the irregular

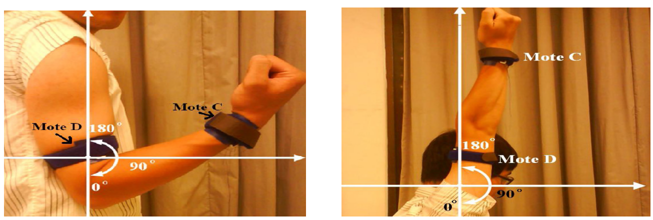


Fig. 9 Fundamental rehabilitation postures: raise the wrist to the elbow position (left), and straighten up the whole arm over the head (right)

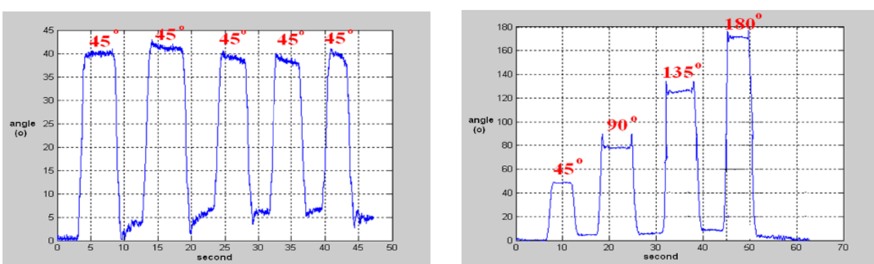


Fig. 10 Tilt angle by mote C for raising the wrist to the elbow position (left) and by mote D for straightening up the whole arm over the head (right)

Table 3 Recognition for rehab activities

Feature Set	Motion ID	Recognized Degree	Threshold Degree	Pass Count
θ_C	(a)	40, 42, 39, 40, 40	40~45	4/5
θ_D	(b)	45, 78, 130, 172	{40~45, 85~90, 130~135, 175~180}	2/4

movements such as non-routine “walk,” “run,” or “sway the elbow” those involve sophisticated personal habitual behaviors, more types of specific membership functions or machine learning algorithms are suggested for the available solution.

5 Conclusions

In this study, we proposed a recognition procedure to detect the essential body activity including static postures and dynamic motions by using the self-developed wearable WSN sensor. The required movement data are transported to backend via wireless sensor network when the user wears two sensors on the specific locations of body. The fuzzy algorithm is applied to calibrate WSN measurement data. In which, the membership functions are based on distribution of the input features such as the tilt angles, the standard deviation and difference rate of acceleration and angular velocity; then the output features of activities can be ruled by the fuzzy-based logic for recognition. With processes of fuzzification and defuzzification, the patterns are yielded and output movements can be obtained. The patterns are verified by a blind test that two typical feature sets are adopted in the study to perform good recognition rate for steady and regular movements. Thus the proposed procedure is approved by two rehabilitation activities to recognize qualified postures. It can be concluded that the recognition rate is available to be a threshold of the restricted postures in rehabilitation. This fundamental study prior to hospital practice contributes the facile and feasible process with the convenient device for the ubiquitous healthcare in physical therapy.

Acknowledgement. This work was supported by NSC grants 100-2221-E-130-004 and 100-2625-M-039 -001.

References

1. Medjahed, H., Istrate, D., Boudy, J., Dorizzi, B.: Human Activities of Daily Living Recognition Using Fuzzy Logic for Elderly Home Monitoring. In: Proc. of IEEE on Fuzzy System, pp. 2001–2006 (2009)
2. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Real-Time Tracking of the Human Body. IEEE Trans. Pattern Analysis and Machine Intelligence 19(7), 780–785 (1997)
3. Hwang, J.Y., Kang, J.M., Jang, Y.W., Kim, H.C.: Development of Novel Algorithm and Real-time Monitoring Ambulatory System Using Bluetooth Module for Fall Detection in the Elderly. In: Proc. of the IEEE on Engineering in Medicine and Biology Society, pp. 2204–2207 (2004)
4. Jones, V., Bults, R., Konstantas, D., Vierhout, P.A.M.: Healthcare PANs: Personal Area Networks for trauma care and home care. In: Proc. 4th Wireless Personal Multimedia Communications, pp. 1369–1374 (2001)
5. Zhao, Z., Cui, L.: EasiMed: A remote health care solution. In: Proc. 27th Conf. of the Engineering in Medicine and Biology Society, pp. 2145–2148 (2005)

6. Jovanov, E., Milenkovic, A., Otto, C., de Groen, P.C.: A Wireless Body Area Network of Intelligent Motion Sensors for Computer Assisted Physical Rehabilitation. J. of Neuro Engineering and Rehabilitation 2(6) (2005), doi:10.1186/1743-0003-2-6.
7. Lisetti, C., Nasoza, F., LeRougeb, C., Ozyera, O., Alvarez, K.: Developing Multimodal Intelligent Affective Interfaces for Tele-home Health Care. Int'l J. of Human-Computer Studies 59(1-2), 245–255 (2003)
8. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless Sensor Networks: a Survey. Computer Networks 38(4), 393–422 (2002)
9. Ghasemzadeh, H., Jafari, R.: Coordination Analysis of Human Movements with Body Sensor Networks: A Signal Processing Model to Evaluate Baseball Swings. IEEE Sensors Journal 11(3), 603–610 (2011)
10. Lind, L., Sundvall, E., Karlsson, D., Shahsavar, N., Åhlfeldt, H.: Requirements and Prototyping of a Home Health Care Application Based on Emerging JAVA Technology. Int'l J. of Medical Informatics 68(1-3), 129–139 (2002)
11. Ahmad, S., Eskicioglu, R., Graham, P.: Design and Implementation of a Sensor Network Based Location Determination Service for use in Home Networks. In: Proc. of IEEE on Mobile Adhoc and Sensor Systems, pp. 622–626 (2006)
12. Kan, Y.-C., Chen, C.K.: A Wearable Inertial Sensor Node for Body Motion Analysis. IEEE Sensors Journal, (in Press) doi:10.1109/JSEN.2011.2148708
13. TinyOS Community Forum, `http://www.tinyos.net`
14. Chiang, S.-Y., Kan, Y.-C., Tu, Y.-C., Lin, H.-C.: A Preliminary Activity Recognition of WSN Data on Ubiquitous Health Care for Physical Therapy. In: 2011 International Conference on Data Engineering and Internet Technology, pp. 517–520 (2011)
15. Jeong, D.-U., Do, K.-H., Chung, W.-Y.: Implementation of the Wireless Activity Monitoring System Using Accelerometer and Fuzzy Classifier. The International Journal of Information Systems for Logistics and Management, 617–716 (2008)
16. Alvarez-Alvarez, A., Alonso, J.M., Trivino, G., Llamazares, A., Ocaña, M.: Human Activity Recognition Applying Computational Intelligence Techniques for Fusing Information Related to WiFi Position and Body Posture. In: Proc. of IEEE on Fuzzy Systems, pp. 1–8 (2010)

Adaptive Reinforcement Learning for Dynamic Environment Based on Behavioral Habit

Akihiro Mimura, Shota Sumino, and Shohei Kato

Abstract. In our previous works, we proposed the adjustment method for learning rate of reinforcement learning: ALR-P. In this method, the learning rate can be adjusted adaptively considering learning progress using a simple and general value TD-error. And we confirmed that the adaptive learning can be realized with the proposed method through a maze problem as dynamic environment. In this paper, we propose the additional ability for this method to realize the learning agent taking behavioral habit into the consideration. The behavioral habit has been taken in human consideration for important decision making in real world. We believe that the learning agent also should have the behavioral habit and take action considering it. We applied ALR-P with the some behavioral habits (ALR-BH) to dynamic maze problem. The experimental results show that the adaptively adjustment of the learning rate is effective for dynamic environment and ALR-BH enabled the learning agent to behave appropriate actions based on the behavioral habit.

1 Introduction

Recently, there are a lot of researches about robotics [4][7] using the reinforcement learning [8] that the learning agent learns and acquires appropriate control rules autonomously through the trial and error. In the reinforcement learning, there are some meta parameters that affect the learning behavior. If these meta parameters of the learning such as learning rate and the discount rate are inappropriate, the learning agent cannot behave sufficient performance. In general, the optimal value does not exist in these meta

Akihiro Mimura · Shota Sumino · Shohei Kato
Dept. of Computer Science and Engineering, Graduate School of Engineering,
Nagoya Institute of Technology,
Gokiso-cho Showa-ku Nagoya 466-8555 Japan
e-mail: {mimura,sumino,shohey}@juno.ics.nitech.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 201–211.
springerlink.com

parameters, and the designer should properly adjust them according to the target task. In addition, it is difficult for the learning agent to adapt to the real world in which environmental changes occur, because a lot of the general reinforcement learning methods using fixed meta parameters premises on the static environment. Thus, in our previous works, we proposed ALR-P [3] in order to realize the robust learning for the real world problems (e.g., environmental change and noise), in which the learning rate is adaptively adjusted considering the learning progress. And we suggested that ALR-P enabled the leaning agent cover real world problem through the maze problems in which environmental changes occurred.

Meanwhile, we believe that the *characteristics* of the problem also should be taken into the consideration. The objective of an agent in the reinforcement learning is generally to maximize the sum of the rewards. However this objective is not entirely appropriate. The objective of agent in the reinforcement learning is generally to maximize sum rewards, however it can be not entirely appropriate. For instance, in motion acquisition of robot, the robot possibly breaks down by repeating tumbles in the trial and error for the learning. Under such the situation, the learning should be progressed carefully for avoiding tumble in order to prevent the most critical accident: break down and loss of control. That is to say, it is thought that the *risk management* meeting the characteristics of the problem (e.g., risk of failure) is important.

In this paper, we propose the *behavioral habit* as the additional ability of the ALR-P. We define the index which determines "which is emphasized the learned solution or the new one" as the behavioral habit in this paper. In ALR-P, the learning rate approaches 0 as the learning converges, and the agent becomes negative for the learning. Then, immediately after the environmental change, the agent tends to select the action that is expected to be suitable through the previous learning, and this is the behavioral habit emphasizing the "past success." On the other hand, "finding the new action" can be demanded in other problems, especially in a problem with dynamic environment, and then the positiveness for the learning should be maintained in a high level throughout the learning. In this paper, we propose the reinforcement learning agent that can cover dynamic environment by using ALR-BH that is adjustment method for the learning rate considering the behavioral habit.

2 Reinforcement Learning

The reinforcement learning agent takes an action and receives the reward for the action, and intend to maximize sum rewards throughout an episode. The agent learns control rule by repeating a sequence of taking an action based on action-value (Q-value) and receiving a reward. The reward indicates what is good in an immediate sense. Q-value is predicted value of the total amount of reward that the agent will receive over the future.

The agent dynamically updates the action-value function based on the received rewards. In this paper, we use Q(λ)-learning [9] as learning algorithm. In Q(λ)-learning, the action-value of the action a in the state s, $Q(s,a)$ is updated according to the following equation.

$$Q(s,a) \leftarrow Q(s,a) + \alpha\delta(s,a)e(s,a), \tag{1}$$

where α, $\delta(s,a)$, and $e(s,a)$ show the learning rate, the TD-error of action a in the state s, and the eligibility trace, respectively. TD-error is calculated by the following equation.

$$\delta(s,a) = r + \gamma \max_{a' \in A'} Q(s',a') - Q(s,a), \tag{2}$$

where r, γ, s', and A' show the reward, the discount rate, the following state after the action a, and set of actions that the agent can select in state s', respectively. Roughly, TD-error indicates the difference of the predicted value and the observed value in the state-action pair. The eligibility trace is the index that shows whether the updating is proper when the action-value function is updated. As shown the following equation, the eligibility trace of a state-action pair become 1 when the state-action pair is selected, and ones of other state-action pairs decrease.

$$e(s,a) \leftarrow \begin{cases} 1 & (s,a) \text{ is current state-action pair} \\ \gamma\lambda e(s,a) & \text{otherwise} \end{cases}, \tag{3}$$

where λ shows the discount rate of trace.

3 Learning Rate

As shown in the Equation (1), a learning rate is a meta parameter which determines updating width of action-value function. Generally, in the case of that the learning rate is high, the progress of learning is fast although the learning is not steady. On the other hand, the learning progresses slowly though the learning is steady in the case of that the learning rate is low. This means that the learning rate is a parameter that balances trade-off between speed and stability of the learning.

In our previous works, we proposed adjustment method named *adaptive learning rate considering learning progress: ALR-P* [3]. ALR-P adjusts learning rate according to each state (or state-action pair) for every step. The learning in the early stage or environment changes should be progressed speedy even if it can be roughly, because the learning agent have to acquire more proper action-value function as soon as possible in that cases. Meanwhile, the learning in the end stage should be progressed more precisely in order to converge action-value on optimal value in high accuracy. From the above things, the learning rate should be adjusted considering the learning progress; the learning rate should be high to value the learning speed in a state that the action-value cannot be correctly estimated, while the one in a state that the action-value can be correctly estimated should be low to value the stability.

And the learning rate can be considered as the index which determines positiveness for the learning. The high and low learning rates are positive and negative for the learning, respectively. It is thought that addition behavioral habit to the learning agent is possible by using this characteristic of the learning rate.

4 Learning Agent Covering Dynamic Environment

4.1 ALR-P

ALR-P is a method of adjusting learning rate for every step based on the degree of learning progress. The learning agent can adapt to dynamic environment by using ALR-P. In this method, we focus on TD-error which can be thought a difference of predicted and observed rewards. In a state such as early stages of learning and environmental change, TD-error is high because prediction of reward is not proper. On the other hand, in a state such as converging stages of learning, TD-error is low and approaches 0 because prediction of reward is proper. Thus, we define the expected value of absolute TD-error approximated by exponential moving average as the degree of learning progress, and use the degree of learning progress to adjust the learning rate. The degree of learning progress in state s and action a, $d(s,a)$ is defined by the following equation.

$$d(s,a) \leftarrow d(s,a) + \omega[|\delta(s,a)| - d(s,a)], \tag{4}$$

where ω shows updating width. The maximum $d(s,a)$ in each state-action pair is saved as $d_m(s,a)$. Using the calculated values from Equation (4), the learning rate α is adjusted by the following equation for every steps.

$$\alpha = \frac{d(s,a)}{d_m(s,a)}. \tag{5}$$

Because the learning rate is adjusted based on the degree of learning progress that each state-action pair has, so the learning rate can be adjusted according to the each state-action pair.

In ALR-P, only TD-error is used to calculate the degree of learning progress. Because the TD-error is a simple value generally calculated in existing reinforcement learning methods, it is easy to build ALR-P into almost all reinforcement learning methods such as Q-learning, Sarsa, Actor-Critic and so on.

4.2 Behavioral Habit

In above, we proposed ALR-P which can adjust the learning rate in **a step**, and the agent can adaptively learn the appropriately under dynamic environment. Meanwhile, the objective of agent in the reinforcement learning is to maximize the total reward it receives in **the long run** (i.e., throughout many steps), and the learning should be realized based on the behavioral habit. For example, in the robot control where mistakes such as tumbles cannot be allowed, the learning should be carefully progressed for less mistakes even if it takes a lot of times. This concept is not for only the reinforcement learning agent but also humans, the behavioral habit has been taken in human consideration for important decision making in real world. The behavioral habit is not taken into consideration in ALR-P, and the agent prioritizes to maximize total reward even if the above-mentioned environment where mistakes cannot be allowed and a risk is involved in order to maximize total reward. In this paper, we propose the ability for adding behavioral habit to the agent in ALR-P.

In ALR-P, $d_m(s, a)$ which is reference value for adjusting the learning rate increases monotonically, and the learning rate approaches 0. Approaching the learning rate to 0 means that the action-value function is hardly updated and the learning becomes negative. Thus, the learning rate is maintained with a certain amount of level by multiplying $d_m(s, a)$ by a *persistence rate* $0 < \kappa \leq 1$ as shown the following equation.

$$d_m(s, a) \leftarrow \kappa d_m(s, a). \tag{6}$$

This value $d_m(s, a)$ is updated for each step the state-action pair (s, a) is selected. In case of κ is large, $d_m(s, a)$ is not decreased much, and the learning rate tends to transition in relatively low level (i.e., the agent values learned solution and becomes negative for the learning). On the other hand, in case of small κ, the learning rate tends to transition in relatively high level (i.e., the agent values new solution and becomes positive for the learning). Therefore, the learning agent can have various behavioral habit by adjusting degree of the positiveness based on κ. The method ALR-P added the ability which determines the behavioral habit is named *Adaptive Learning Rate based on Behavioral Habit: ALR-BH*. In ALR-P, the value $d_m(s, a)$ is not updated by the equation (6). Therefore, ALR-P is considered as a special case of ALR-BH such that persistence rate $\kappa = 1$. Q(λ)-learning applying ALR-BH is shown in Table 1. Although there are some methods for adjustment the learning rate [5][2], the objective of these methods is not addition of the behavioral habit like ALR-BH.

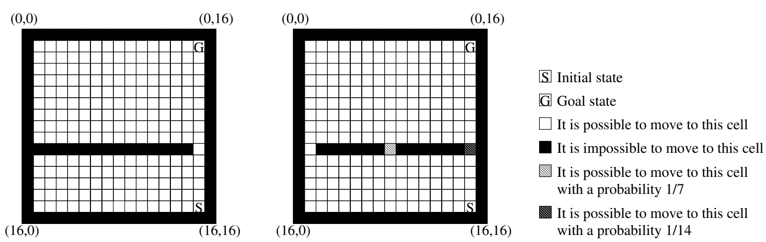
Table 1 Q(λ)-learning applying ALR-BH

Initialize $Q(s,a)$ arbitrarily for all s, a
Initialize $e(s,a) = 0$ and $d(s,a) = 0$ for all s, a
Repeat (for each episode):
  Initialize s, a
  Repeat (for each step of episode):
    Choose a' from s' using policy derived from Q (e.g., softmax-method)
    $\delta(s,a) \leftarrow r(s,a) + \gamma \max_{a'} Q(s',a') - Q(s,a)$
    $d(s,a) \leftarrow d(s,a) + \omega[|\delta(s,a)| - d(s,a)]$
    If $d(s,a) > d_m(s,a)$:
      $d_m(s,a) \leftarrow d(s,a)$
    $\alpha = d(s,a)/d_m(s,a)$
    $e(s,a) = 1$
    For all s, a:
      $Q(s,a) \leftarrow Q(s,a) + \alpha\delta e(s,a)$
      $e(s,a) \leftarrow \gamma\lambda e(s,a)$
    $d_m(s,a) \leftarrow \kappa d_m(s,a)$
    $s \leftarrow s'$
    until s is terminal

5 Adaptive Learning Experiment

5.1 Maze Problem with Environmental Changes

To verify the effect of behavior on the learning agent by addition of the behavioral habit, we apply ALR-BH to the maze problems (Fig. 1) in which the environmental changes occur. Each cell shows state, and the coordinates are defined as follows: left-uppermost is $(0,0)$, right-uppermost is $(0,16)$, left-lowermost is $(16,0)$, and right-lowermost is $(16,16)$. An initial state and a goal state are set to $(15,15)$ located in lower-right part of the maze and $(1,15)$ located in upper-right part, respectively. The environmental change in this experiment is defined as follows; until 100th episode (Fig. 1(a)), agent can reach goal state by passing through $(10,15)$, from 101st episode (Fig. 1(b)),



(a) Until 100th episode (b) From 101st episode

Fig. 1 Maze problem

$(10, 1)$ and $(10, 8)$ is opened up, and the states which are $(10, 8)$ and $(10, 15)$ change whether transition is possible or not with a fixed probability for every steps. The agent can move on to the state $(10, 8)$ with the probability 1/7 and the state (10,15) with the probability 1/14, respectively. Thus, after environmental changes, the agent can reach goal state by passing along one of the three paths. Each expected sum rewards of these three paths are large in order of the right side path which contains $(10, 15)$, the center path which contains $(10, 8)$, and the left side path which contains $(10, 1)$. From above things, it can be said that the right side path is high-risk-high-return and the left side path is low-risk-low-return.

Comparative methods are the conventional general methods that learning rate is each fixed to 0.3 as *negative behavioral habit* for the learning, 0.6 as *neutral behavioral habit* for the learning, and 0.9 as *positive behavioral habit* for the learning. Q(λ)-learning is used as learning algorithm, and softmax method is used as action selection rule. Each parameter is set as follows; the discount rate $\gamma = 0.95$, the discount rate of trace $\lambda = 0.95$, and the update width of the degree of learning progress $\omega = 0.1$. This experiment consists of 300 episodes, and one episode is that learning agent reaches in goal state or passes over 100 steps. The agent moves to top, bottom, right, or left cell, and receives reward that is -1 in a step.

5.2 Experimental Results

Success Rate of Task. The above-mentioned experiments were conducted 100 times for each method. The results shown in Table 2 indicates that success rate of the task and selection path at the success episode after environmental changes. From the result in case of using conventional methods that the learning rates are fixed, we confirm that the rates of the selection probabilities of each path are different. The learning agent with low learning rate selects the right (risky) path which contains a state $(10, 15)$ mainly (item of negative behavioral habit in Table 2) and the agent with high learning rate selects

Table 2 Success rate of the task and selection path at the success episodes after environmental changes

Setting	Learning rate	Success rate (%)	Selection rate of each path (%)		
			$(10, 1)$	$(10, 8)$	$(10, 15)$
Negative behavioral habit	$\alpha = 0.3$ (fixed)	61.10	2.19	44.51	53.30
	ALR-BH ($\kappa = 1.0$)	81.76	10.08	33.08	56.84
Neutral behavioral habit	$\alpha = 0.6$ (fixed)	66.22	34.81	46.03	19.16
	ALR-BH ($\kappa = 0.98$)	74.95	32.40	45.81	21.79
Positive behavioral habit	$\alpha = 0.9$ (fixed)	73.29	66.76	25.71	7.53
	ALR-BH ($\kappa = 0.90$)	74.67	72.82	20.53	6.65

the left (safety) path which contains a state $(10, 1)$ mainly (item of positive behavioral habit in Table 2). It is thought that in case of low learning rate, the agent becomes insensitive to environmental changes and tends to select the right side path along which the agent was able to pass before environmental changes even though the path is blocked off. On the other hand, the agent with high learning rate is sensitive to the changes and tends to select the left side path which is opened up newly. From above things, the agent whose the behavioral habit is positive for the learning has the tendency to avoid the state including uncertainty.

In the proposed method, each behavioral habit's persistence rate κ which determines positiveness was set such that the each selection rate of paths becomes mostly equivalent to one of the conventional methods. And the persistence rate κ was set as followings; $\kappa = 1.0$ as *negative behavioral habit* for the learning, $\kappa = 0.98$ as *neutral behavioral habit* for the learning, and $\kappa = 0.90$ as *positive behavioral habit* for the learning. From Table 2, we confirm that the selection rates of both methods are approximately equivalent, and the success rates of ALR-BH are higher than conventional methods in the all behavioral habits. This result suggests that the learning agent could adjust the learning rate appropriately and adapt to this problem by using ALR-BH.

Addition of the behavior habit to the agent seems to be possible by using fixed learning rate, however the success rate of the task becomes low as the risk increases. On the other hand, ALR-BH not only adjusts adaptively the learning rate, but also maintains high success rate. The results suggest that ALR-BH was able to take action appropriately according to corresponding the behavior habit.

Transition of Sum Rewards. The results shown in Fig. 2, Fig. 3, and Fig. 4 indicate that the transition of sum rewards in case of the negative, the neutral, and the positive behavioral habit for the learning, respectively. Before the environmental changes, the proposed method have learning converge more quickly than the conventional methods. The reason why this results is that the learning progressed roughly and speedily by adjusting the learning rate high in the early stage of the learning, and the learning rate was lowered gradually as the learning progressed by using ALR-BH. Also, after the environmental changes, we confirm that the proposed method could relearn quickly. Especially, in case of the negative behavioral habit for the learning, the difference of convergence speed between ALR-BH and the conventional method is a lot.

Transition of Learning Rate. Fig. 5 shows an example of transition of learning rate in the state-action pair that state is $(11, 15)$ and action is moving to up. From this figure, we confirm that after environmental changes learning rate was adjusted high again. Because the location of state $(11, 15)$ was near the state in which environmental changes occurred, it is thought that the state was profoundly affected by the environmental changes. ALR-

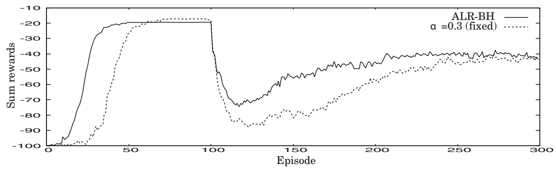


Fig. 2 Transition of sum rewards in case of the negative behavioral habit for the learning

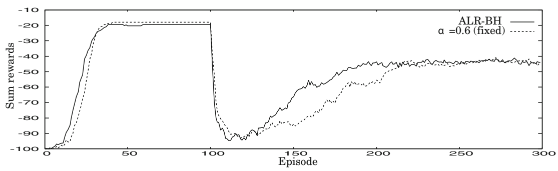


Fig. 3 Transition of sum rewards in case of the neutral behavioral habit for the learning

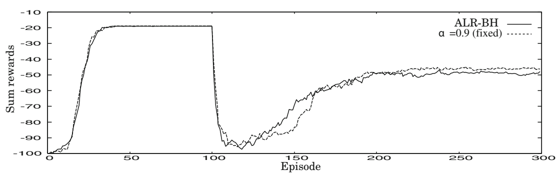


Fig. 4 Transition of sum rewards in case of the positive behavioral habit for the learning

BH could promote relearning by adjustment of the learning rate according to the variation of TD-error. Fig. 6 shows a typical example of transition of learning rate in state-action pair that state is initial state $(15, 8)$ and action is moving to right. From this figure, we confirm that the learning rate was not high even though it was after the environmental changes. Because the location of state $(15, 8)$ was far from the state in which environmental changes occurred, it is thought that the state is not seriously affected by the environmental changes. From these results, it was confirmed that ALR-BH could adjust learning rate to appropriate value on each state.

Fig. 7 shows a typical example of transition of learning rate in the episode in which the agent selected right side (the shortest) path. In the states except the state $(10, 15)$ in which environmental changes occurred, the learning rate is changing with low level. On the other hand, the learning rate is enlarged in the state in which environmental changes occurred, when the state was passed. In this state, the learning is not converged in the first place because of the environmental changes, so TD-error also varied widely and the learning rate was adjusted according to the changes. In this way, updating of action-value function which reflected having succeeded greatly is possible, then the

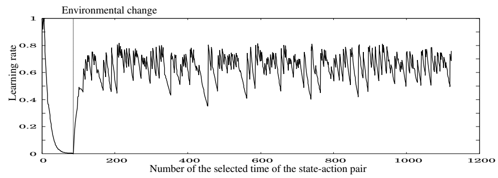


Fig. 5 A typical example: transition of learning rate in state-action pair that state is $(11, 15)$ and action is moving to up

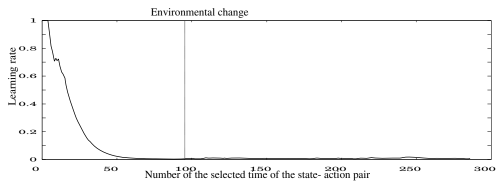


Fig. 6 A typical example: transition of learning rate in state-action pair that state is $(15, 8)$ and action is moving to right

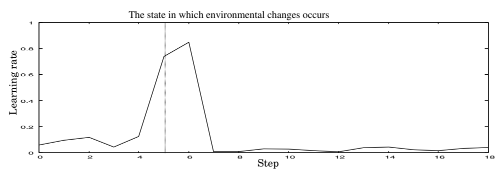


Fig. 7 A typical example: transition of learning rate in the episode in which the agent selected right side path

agent become prone to also select the path which contains $(10, 15)$ in the next episode. On the other hand, in conventional method, action-value function does not reflect the success because the learning rate is stationary regardless of success or failure. It is thought that this thing mentioned above is a factor of difference of the success rates (Table. 2) between proposed and conventional methods.

6 Conclusion and Future Works

In this paper, we proposed the behavioral habit as the additional ability of the ALR-P [3] which is proposed in our previous works. We applied ALR-P with the behavioral habit (ALR-BH) to dynamic maze problem having some paths whose expected sum rewards were different to goal. Through the experiment, we confirmed that the adaptively adjustment of the learning rate was effective for environmental changes and ALR-BH enabled the learning agent to take suitable actions based on the behavioral habit.

In this paper, we used the fixed behavioral habit (i.e., the persistence rate κ), however the behavioral habit should be changed based on the situation in order to realize more effective learning. Therefore, we will describe the recognition method for the characteristics and situation of a problem, and realize the dynamic behavioral habit change. And the behavioral habit can be considered as the characteristic of a learning agent. In recent years, it has been studied that the relation between the influence of the meta parameters in the reinforcement learning and the influence of the neurotransmitter in human behavior [1][6]. Thus, we believe that the reinforcement leaning considering the behavioral habit possibly enables to help the development of more humanly communication robot, and presentation of the novel opinion for the human learning progress through a perspective of the computer science.

Acknowledgements. This work was supported in part by the Ministry of Education, Science, Sports and Culture, Grant.in.Aid for Scientific Research under grant #20700199.

References

1. Doya, K.: What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural Networks 12(7), 961–974 (1999)
2. Geoge, A.P., Powell, W.B.: Adaptive stepsizes for recursive estimation with applications in approximate dynamic programming. Machine Learning 65(1), 167–198 (2006)
3. Mimura, A., Kato, S.: Adaptive reinforcement learning based on degree of learning progress. In: 17th International Symposium on Artificial Life and Robotics (2012)
4. Mimura, A., Nishibe, S., Kato, S.: Kinetic chained throwing humanoid robots by using reinforcement learning. In: 12th International Symposium on Advanced Intelligent Systems, pp. 188–191 (2011)
5. Noda, I.: Adaptation of stepsize parameter to minimize exponential moving average of square error by newton's method. In: 9th International Conference on Autonomous Agents and Multiagent Systems, pp. M-2–1 (2010)
6. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D.: Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442(7106), 1042–1045 (2006)
7. Riedmiller, M., Gabel, T., Hafner, R., Lange, S.: Reinforcement learning for robot soccer. Autonomous Robots 27(1), 55–73 (2004)
8. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press (1998)
9. Watkins, C.J.C.H., Dayan, P.: Q-leaning. Q-leaning. Machine Learning 8(3-4), 279–292 (1992)

An Exploration of Individual Level Tacit Knowledge Integration

Dao Cheng Hong, Hong Ling, and Cheng Hong Zhang

Abstract. With the advanced knowledge economy, much attention has been placed on knowledge integration because of human beings' cognitive limits. As a spread phenomenon, tacit knowledge integration has becoming a hot topic because of the embedness in social activities and interactions. Previous literatures have advocated the direct effect of social networks (SNs) on knowledge integration despite the fact that most social relations eventually lead to the integration action mediating individual intentions. Given this apparent gap, this paper investigates integrative process of tacit knowledge based on social networks and proposes a novel theory of tacit knowledge integration (TKI) drawn on SNs perspective, technology acceptance model (TAM) and theory construction methodology. The explorative study demonstrates several findings.

1 Introduction

With the advanced knowledge economy, few works can be done only by a single person in modern organization. Individuals developing and possessing different knowledge especial tacit knowledge do their best to communicate and cooperate with each other in working toward the same goal. Much more attention has been increasingly paid on knowledge which is power and one of most valuable assets in organization. Researcher further asserted that without the adequate management of organizational knowledge, the consequences for organizations could be devastating [1]. Therefore, much more research and practice of knowledge management has been conducted in an effort to capture, store, and disseminate knowledge effectively and efficiently for organizational performance.

Furthermore, knowledge especial tacit knowledge being recognized as the 90 percent of organization knowledge capacity which resides in human beings is critical to organization's competitiveness [2,3]. Researchers have begun to pay much

Dao Cheng Hong
Shanghai Key Laboratory of Intelligent Information Processing,
School of Computer Science, Fudan University, Shanghai 200433, China

Hong Ling · Cheng Hong Zhang
School of Management, Fudan University, Shanghai 200433, China

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 213–224.
springerlink.com

more attention on integration for fragmented pockets of tacit knowledge, such as expertise, specialized knowledge. Tacit knowledge integration leverages individual and organization's performance and attracts more attentions because of human beings' cognitive limits and knowledge dispersancy. Therefore, tacit knowledge integration of many different individuals in the process of producing goods and services is playing a critical and constructive role for organization.

However, with the approaching of information booming era, someone has proven more amenable to being inquired about tacit knowledge than others within the same social networks (SNs). In other words, individuals have many obstacles or resistance to acquirement of tacit knowledge from others of their social networks. These practical observations lead to main research questions: what factors of SNs influence the integration of tacit knowledge? How do factors impact on the integration of tacit knowledge?

Technology acceptance model (TAM) posits that perceived usefulness and perceived ease of use towards objects determine an individual's intention to actual behavior with intention serving as a mediator of action [4]. Thereby, this article focuses on individual activities and analyses the factors influencing on tacit knowledge integration (TKI) from TAM perspective. In other words this paper explores influences on the integration of tacit knowledge and develops TKI theory drawn on SN, TAM and theory construction methodology [5, 6]. The contributions of this research are focusing on a hot topic and proper integration, which provide a novel and integrated perspective both for academia and practice neglected by most researchers. The most important but not the last one is that under the theoretical framework most of TKI phenomenon can be explained properly.

2 Theoretical Concepts

In the booming society it's unnecessary and impossible for one person to master and apply all knowledge and technology properly. Researchers of knowledge collaboration have asserted that individuals mostly develop and rely on their social networks in deciding with whom to collaborate and how to collaborate with partners [7]. With the increasing number of evidence, tacit knowledge integration is conducted by people and embedded in their social activities and interactions. People are known to acquire knowledge from their own interpersonal networks which extend beyond the formal organizational structures for knowledge collaboration [8, 9]. Knowledge integration depending on how members know, synthesize and apply current their individually owned knowledge is ultimately about micro-social interactions among individuals [10, 11]. Researchers presumed that tacit knowledge integration is conducted by actors and determined by social connections or ties; obstacles to an effective integration include lack of familiarity among individuals, unfamiliar language, disparities in verbal skill, distinctive thought worlds, insufficient conflict, and status differences [10]. Withal, knowledge integration has been recognized as "in essence a process of the management of social networks" [12].

As a much more widespread phenomenon of modern society, knowledge integration is a hot topic and has been studied both in research and industry. Two main research streams of knowledge integration include organizational capability perspective [11, 13-17] and process perspective [18-22]. Anyway, tacit knowledge is what the actor knows, and is derived from experience and embodies beliefs and values. Mostly, tacit knowledge residing in human minds can only be shared through up-close demonstration, observation, or hands-on experience [23]. Integration of pocketed and dispersed tacit knowledge emphasizes the economic value of specialization and the effectiveness for both organization and individuals. Tacit knowledge which is the most important basis for the generation of new knowledge is action-oriented knowledge and even more the most valuable assets for both individual and organization. For the purpose of research, tacit knowledge integration is defined as an ongoing collective process of promoting individual tacit knowledge development by transferring and sharing tacit knowledge through communication and interaction for a successful project, task in organization [21, 24].

3 TKI Theory

Practically, individuals are known to acquire knowledge depending on their own personal networks which extend beyond the formal organizational structures. Research on knowledge collaboration presumed that tacit knowledge integration is embedded in social interactions., individuals often develop and rely on their social networks in deciding with whom to collaborate and how to collaborate with each other. Additionally, perceived usefulness and perceived ease of use towards knowledge accession determine an individual's intention to actual behavior [4].

In TKI theory, the dependent variable is tacit knowledge integration and the main constructs are expressive richness of ties, ties strength, individual heterogeneity and intention of integration. Each of main constructs is posited to have a positive effect on tacit knowledge integration mediating individual intention. In other words, as each of constructs increases, the tacit knowledge becomes more amenable to integration. It doesn't mean that a tacit knowledge integration process with low degree of social networks can't be conducted; rather, it means that it would be more amenable to being steered if these characteristics degrees are high. The dependent variable of tacit knowledge integration is continuous rather than discrete, and should be thought of as a matter of degree, not of kind. This is a critical distinction. The propositions of TKI theory would not be interpreted as on/off.

3.1 Knowledge Integration

Three dimensions of knowledge integration have been presented in existent literatures which are efficiency, scope and flexibility [14, 25, 26]. Researchers have discussed these contents as follow. Firstly, flexibility of knowledge integration is the extent to which organization can access additional knowledge and reconfigure existing knowledge. Secondly, scope of knowledge integration refers to the breadth of tacit knowledge that the project draws upon. Knowledge integration efficiency, finally, is the extent to which organization accesses and utilizes the tacit knowledge held by individual members. In order to utilize the specialized knowledge stored within every individual, it is vital that all members have a good understanding of each other's capability. As the statement of previous session, we take tacit knowledge integration as a continuous rather than discrete process, and a matter of degree, not of kind. In this research, we can measure tacit knowledge integration using efficiency, scope and flexibility.

3.2 Intention of Tacit Knowledge Integration

An individual's positive or negative feelings about performing the target influence behavioral intention and further lead to actual behavior [4]. Attitudes towards usage and intentions about performing the target behavior may be ill-formed or lacking in conviction or else may occur only after preliminary strivings to evolve. Thus, actual performance may be a direct or immediate consequence of such attitudes and intentions. Individuals may integrate tacit knowledge when they have clear intention of tacit knowledge integration within their social networks.

Hypotheses 1. The greater (lower) the intention of tacit knowledge integration, the higher (lower) level of tacit knowledge integration is to being developed.

3.3 Social Networks

Social network is a powerful theory to explain social interaction and knowledge integration especially, which asserts that the world is made up of relationships; individuals and organizations are encircled by social connections [27]. Social activities promote interaction between people/objects and form multi-dimensional relationships in daily lives [28]. Most previous researches emphasized the structure dimension of social networks and neglected ties dimension and nodes dimension with the essence of interaction and the mechanism of knowledge integration.

Social ties. Two broad extreme types of social ties have been distinguished in social networks theory [29-33]. One end is expressive tie involving expressions of interpersonal affect which are sources of social support and provide a sense of identity and personal belonging, and serve to transmit normative expectations, such as friendships. The other end is instrumental ties involving gathering information, advice, and resources necessary for accomplishment, such as work-related-advice ties. In general, expressive ties are normative and affection based, whereas instrumental ties are information and cognition based [34, 35]. The importance of social interaction and connections between individuals should be stressed in order to promote effectiveness of tacit knowledge integration, because integration of tacit knowledge requires high level of personal interaction through reciprocity exchange relationships.

Tacit knowledge is a highly personal knowledge and hard to formalize which is rooted in action, commitment, and involvement in a specific context [36]. Research in social networks suggests that individuals who have mutually expressive ties provide support and assistance to one another [7]. Social ties facilitate tacit knowledge integration through the expressive richness of ties. Expressive ties promote tacit knowledge integration by providing a sense of organizational identity and personal belonging which are perception and motivations to assist a contact. Therefore, much more expressive richness of ties facilitates identity and personal belonging with individual intention of tacit knowledge integration.

Hypotheses 2. Expressive richness of ties promotes intention of tacit knowledge integration.

Referring to social ties, the strength of ties is necessary and has been attracting more attention. Tie strength, a concept ranging from weak ties at one extreme to strong ties at the other, characterizes the closeness and interaction frequency of a relationship between two partners [37, 38]. At the dyadic level, theories have arisen around both extremes of the tie-strength concept, with research findings to both strong and weak ties. Researcher studied strong ties and demonstrated the importance of it in transferring tacit, complex knowledge across departmental boundaries [39]. Moreover, strong ties bind cliques of individuals and primarily convey affluent news. Strong ties develop the relationship-specific heuristics and specialized language which is conducive to conveying complex chunks of knowledge [7, 40]. People who interact with each other regularly are more likely to have similar perceptions than people who do not interact [41]. Withal, strong ties promote the creation and maintenance of social connections because of individual obligation and responsibility sense. Strong ties cultivate the individual obligation and responsibility perceptions leveraging trust and coordination of intention.

Hypotheses 3. Ties strength promotes intention of tacit knowledge integration.

Network Nods. Diversity of participants in workplace has been investigated in depth, such as social category diversity, value diversity, and informational diversity [42]. Value diversity is the fundamental dimension of nodes heterogeneity in social networks. As the basic diversity of individuals, value diversity provides various ideas and perceived knowledge which are useful for tacit knowledge integration. Value diversity exists when individuals differ with each other in terms of what they think of real task, goal, target, or mission should be. In other words, individuals' similarity in goals and values enhances the same voice and prevent creative mind and behavior. Furthermore, considerable evidence points to the detrimental effects of value diversity on coordination. Consequently, we draw out the final hypothesis:

Hypotheses 4. Value diversity of individuals promotes intention of tacit knowledge integration.

4 Quantitative Assessment of Research Model

4.1 Survey Sample

We conducted this study involving a random survey of individuals with works in the mainland of China. The survey continued over a period of three months and raw data were collected from volunteers. In our study all of the respondents are full-time professionals in a variety of industries. In order to decrease misunderstanding, we provided the definition and several explanations in the questionnaire (Chinese version). The context of team/project work also has been emphasized to match the definition of tacit knowledge integration. Table 1 lists the characteristics of the pilot investigation sample.

4.2 Survey Measures

Tacit knowledge integration. Within existent literatures [14, 25, 26], efficiency, scope and flexibility have been adopted to measure knowledge integration. In order to develop proper measures (TKI), we adopt prior research items and modify several ones to adapt this research aims under team-work condition focusing on individual tacit knowledge integration.

Intention of tacit knowledge integration. We measured intention of tacit knowledge integration [INT] with adapted classic-item scale developed by Davis [4]. We asked individuals to rate their intentions based on their own personal interactions with others during the latest project or task in organization.

Table 1 Description of the sample

N=157	Frequency	Percent	Valid Percent
Gender			
Male	106	67.5	67.5
Female	51	32.5	32.5
Age			
≤25	20	12.7	12.7
26-35	115	73.2	73.2
36-45	22	14.0	14.0
Tenure of current job			
≤1	11	7.0	7.0
1-3	30	19.1	19.1
3-5	36	22.9	22.9
5-7	27	17.2	17.2
≥7	53	33.8	33.8
Education level			
College	18	11.5	11.5
Bachelor	102	60.5	60.5
Master or PhD	37	23.6	23.6
Job position			
Staff	77	49.0	49.0
Junior manager	48	30.6	30.6
Middle manager	29	18.5	18.5
Senior manager	3	1.9	1.9
Total	**157**	**100.0**	**100.0**

Social ties. We measured expressive richness of ties using three-item scales (ET) in this study [43-46]. Respondents were inquired to recall and evaluate their social ties within knowledge networks during their latest project or task in workplace. The strength of social ties (ST) has been measured with frequency, intimacy, and reciprocity proposed by Granovetter [45].

Value diversity. After our preliminary analysis, in order to promote reliability, we developed a three-item scale for value similarity (VAL) based on the measurement of Jehn et al.[42].

In our research, all items were measured using a 7-Likert scale, ranging from 1 to 7 (strongly disagree to strongly agree). Next part, we conduct confirmatory factory analyses and analysis hypotheses by testing the proposed path model with structural equation modeling tool (LISREL 8.70).

4.3 Test Results

The empirical study requires that the reliability and validity of variables in study should be rigorously tested. All of the adopted multiple items in this study have established acceptable reliability and validity level from past research. However, in an effort to improve the appropriateness of instrument items, some questions were slightly altered in wording and style in Chinese. Factor analysis was conducted to establish convergent and discriminant validity of the constructs. Each multi-item construct was modeled as reflective of the latent variable rather than formative because we expected the items measuring each constructs to covary. Overall, factor analysis results indicate strong support for the construct validity of the structural variables. The value of Cronbach's alpha for all the extracted

Table 2 Reliability and Validity of Measurement Model

Variable	Item	Cronbach's Alpha	X^2/d.f	IFI	NFI	CFI	GFI	RMSEA
Expressive Tie	ET1 ET2 ET3	.957	1.99	0.98	0.97	0.98	0.88	0.08
Tie Strength	ST1 ST2 ST3	.892						
Value Diversity	VAL1 VAL2 VAL3	.939						
Individual Intention	INT1 INT 2 INT 3	.899						
Tacit Knowledge Integration	TKI1 TKI2 TKI3	.925						

Table 3 Correlation Matrix of Latent Variables

	ET	ST	VAL	INT	TKI
ET	1.00				
ST	0.65	1.00			
VAL	-0.69	-0.56	1.00		
INT	0.75	0.63	-0.78	1.00	
TKI	0.74	0.59	-0.69	0.87	1.00

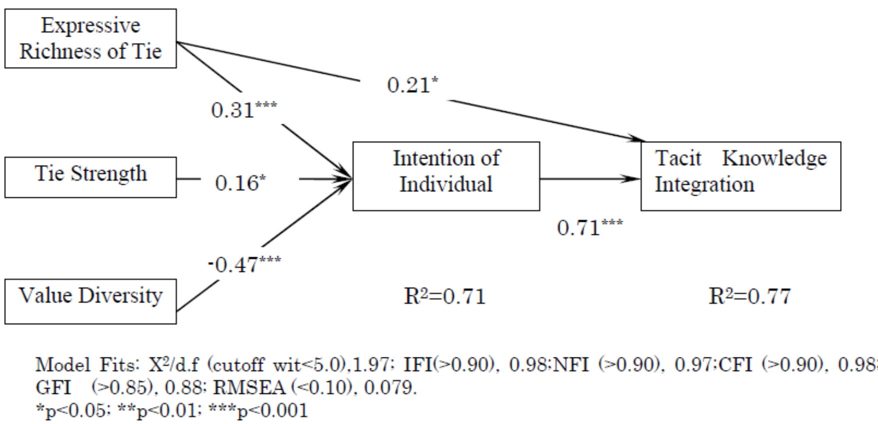


Fig. 1 Results

constructs and the measurement model fits also are presented in Table 2, and the correlation of latent variables is described as shown in Table 3.

Fig.1 is an adjusted and optimal graphical depiction of the LISREL analysis results which shows details of structure model testing of our study. The hypothesized path between intention and TKI is significant, and H1 has been supported. As expected, expressive richness of tie and tie strength have a significant impact on individual intention of performance (H2, H3). Contrary to previous prediction, the hypothesized path between value diversity and intention of tacit knowledge integration is negative prominence (H4). It means that the main effect of value diversity is hampering tacit knowledge integration through individual will. As the basic diversity of individuals, value diversity decreases satisfaction, intent to remain and commitment to organization. In addition, the results verify that the impact of tie expressive richness on tacit knowledge integration is partially mediating individual intent to performance. Therefore, hypotheses 1, 2 and 3 are supported, but hypotheses 4 are not supported in this empirical research.

5 Discussion

Our research discusses the effect of individual social ties and personal intention when conducting tacit knowledge integration to solve problems or create new knowledge in workplace. The main contribution of this study includes identifying ways to increase integrative capabilities of individuals who must draw on social ties and individual intention.

First of all, we propose a theoretical model of tacit knowledge integration. This framework is constructed based on a review of the social networks, TAM, and knowledge management literatures, along with the results of a previous qualitative study. Secondly, TAM is a new perspective for knowledge integration and attracts more and more attention in individual tacit knowledge integration domain. Thirdly, we pay more attentions on social ties and value diversity of nodes heterogeneity for exploring the mechanism of tacit knowledge integration. It is a complement to previous research that has emphasized the structure of social networks. We find that the main effect of value diversity is hampering tacit knowledge integration. Especially, a point of interest in this work is the finding that value diversity is harmful to tacit knowledge integration mediating individual intent. As the basic diversity of individuals, value diversity decreases satisfaction, intent to remain, and commitment to organization. That's to say, with high value similarity, individuals will have the same believes and criterions beneficial for specialization and communication with each other [24, 47]. Consequently, tacit knowledge integration can't be conducted smoothly and reach the objective level. Finally, the preliminary research also verifies that expressive richness of tie, tie strength and value similarity advance tacit knowledge integration mediating the individual intention of performance. The finding also illuminates us the importance of social networks and reminds us that we should cultivate and develop higher level of intent for tacit knowledge integration.

This research demonstrates the mechanism of tacit knowledge integration based on social networks from TAM perspective and provides some implications for both academia and practice. In the first place, this study is a novel perspective to explain and predict factors of social networks during knowledge integration procedures for researchers. This research apparently emphasizes the tie dimension and nodes dimension of social networks and intention for performance which are playing a critical role in knowledge integration. This aspect of the study will help researchers better understand what and how social networks continue to have profound effects on tacit knowledge integration. In the second place, this study is also a practical framework for managerial activities in which tacit knowledge integration is conducted. The results provided by this study will help managers with ability to distinguish and evaluate expressive richness of ties for development of strong intention and tacit knowledge integration. The preliminary results suggest that affection-based expressive ties should be encouraged and promoted through social activities during work time. The kernel value and train for staff in organization should be encouraged and cultivated carefully. Social interactions and activities among individuals should be promoted which facility performance intention and further tacit knowledge integration.

Definitely, the new theoretical framework that constructs socio-technical factors of social ties and intent of performance affecting knowledge integration is limited by certain properties. First, to theorize this work, we mainly draw on social networks and TAM theories. All framework constructs come from our interpretation with integrative perspective and more practice should test this framework to improve and perfect this TKI framework. Second, the theoretical framework is constructed with TAM perspective which is power for explanation of knowledge share, transfer and integration but not enough. Other factors effect on tacit knowledge integration may be explicitly examined, such as managerial factors. Certainly, this research just is conducted from a specific perspective and other new views will be adopted in the next research schedule.

6 Conclusions

Individuals are known to acquire knowledge from their own social networks which extend beyond the formal organizational structures, and will develop intention of performance in social interactions which has been recognized as the mechanism of knowledge integration. Tacit knowledge integration is defined as a developing collective process of promoting individual tacit knowledge development through social activities for a successful project, task or work. This study illuminates knowledge integration based on social networks from TAM perspective, and tentatively suggests that expressive richness of ties, strength of ties and similar value of individuals should be encouraged and promoted for high level of tacit knowledge integration. The empirical results provide a novel mechanism for researching tacit knowledge integration which is a hot topic and is playing a pivotal role in this turbulent business. Practically, this theoretical framework makes contributions by providing strong potential guidance for management, such as understanding and predicting tacit knowledge integration in practice.

Acknowledgments. These authors equally contributed to this study and Prof. Cheng Hong ZHANG is the corresponding author. This work was supported partly by NSFC (70871027, 70972047, 61074119, and 61170095) and Shanghai Committee of Science and Technology (10DZ2272800 and 11DZ2273100).

References

[1] Cameron, P.: Managing the wealth. CMA Management 74(9), 46–49 (2000)
[2] Alavi, M., Leidner, D.E.: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. MIS Quarterly 25(1), 107–136 (2001)
[3] Wang, J., Zhao, S., Yang, J.: Research on the Sharing Model and Mechanism of Tacit Knowledge. Science of Science and Management of S.&.T. 15(10), 65–67 (2004)
[4] Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13(3), 319–339 (1989)
[5] Eisenhardt, K.M.: Building theories from case study research. Acad. Management Rev. 14(4), 532–550 (1989)
[6] Wand, Y., Weber, R.: Research commentary: Information systems and conceptual modeling—A research agenda. Information Systems Research 13(4), 363–377 (2002)
[7] Reagans, R., McEvily, B.: Network Structure and Knowledge Transfer: The Effects of Cohesion and Range. Administrative Science Quarterly 48(2), 240–267 (2003)

[8] Cross, R., Cummings, J.N.: Tie and Network Correlates of Individual Perform-ance in Knowledge-Intensive Work. The Academy of Management Journal 47(6), 928–937 (2004)

[9] Cross, R., Sproull, L.: More than an answer: Information relationships for actionable knowledge. Organization Science 15(4), 446–462 (2004)

[10] Okhuysen, G.A., Eisenhardt, K.M.: Integrating Knowledge in Groups: How Formal Interventions Enable Flexibility. Organization Science 13(4), 370–386 (2002)

[11] Kogut, B., Zander, U.: Knowledge of the Firm, Combinative Capabilities and the Replication of Technology. Organization Science 3(3), 383–397 (1992)

[12] Huang, J.C., Newell, S.: Knowledge integration processes and dynamics within the context of cross-functional projects. International Journal of Project Management 21(3), 167–176 (2003)

[13] Grant, R.M.: Prospering in Dynamically-Competitive Environments: Organizational Capability as Knowledge Integration. Organization Science 7(4), 375–387 (1996)

[14] De, B.M., Van Den Bosch, F.A.J., Volberda, H.W.: Managing organizational knowledge integration in the emerging multimedia complex. Journal of Management Studies 36(3), 379–398 (1999)

[15] Grant, R.M.: The knowledge-based view of the firm: Implications for management practice. Long Range Planning 30(3), 450–454 (1996)

[16] Smith, K.G., Collins, C.J., Clark, K.D.: Existing knowledge, knowledge creation capability, and the rate of new product introduction in high-technology firms. Academy of Management Journal 48(2), 346–357 (2005)

[17] Tiwana, A.: The Influence of knowledge integration on project success. Georgia State University, Atlanta (2001)

[18] Farrell, J., Flood, P., Curtain, S., et al.: CEO leadership, top team trust and the combination and exchange of information. Irish Journal of Management 26(1), 22–40 (2005)

[19] Hamel, G., Prahalad, C.K.: Strategy as Stretch and Leverage. Harvard Business Review 71(2), 75–84 (1993)

[20] Henderson, R.M., Clark, K.B.: Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. Administrative Science Quarterly 35(1), 9–30 (1990)

[21] Hong, D.C., Ling, H., Zhang, C.H., Li, Y.M.: Theoretical Insight into the Mechanism of Knowledge Integration. In: Proceeding of 2nd IEEE International Conference on In-formation Management and Engineering, pp. 696–701. China (2010)

[22] Zahra, S.A., Ireland, R.D., Hitt, M.A.: International expansion by new venture firms: International diversity, mode of market entry, technological learning, and performance. Academy of Management Review 43(5), 925–950 (2000)

[23] Hamel, G.: Competition for competence and inter-partner learning within international strategic alliances. Strategic Management Journal 12(Summer), 83–103 (1991)

[24] Ling, H., Hong, D.C., Zhang, C.H.: Research on tacit knowledge integration: a synthesis of social ties and TMS. Knowledge Management Research & Practice 9(3), 256–262 (2011)

[25] Grant, R.M.: Toward a knowledge-based theory of the firm. Strategic Management Journal 17(Special Issue), 109–122 (1996)

[26] Subramanian, A.M., Soh, P.H.: Knowledge Integration and Effectiveness of Open Source Software Development Projects. IIMB Management Review 20(2), 139–148 (2008)

[27] Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press (1994)
[28] Butts, C.T.: Social network analysis: A methodological introduction. Asian Journal of Social Psychology 11(1), 13–41 (2008)
[29] Fombrun, C.J.: Strategies for network research in organizations. Acad. Management Rev. 7(2), 280–291 (1982)
[30] Ibarra, H.: Race, opportunity, and diversity of social circles in managerial networks. Acad. Management J. 38(3), 673–703 (1995)
[31] Lincoln, J.R., Miller, J.: Work and friendship ties in organizations: A comparative analysis of relational networks. Administrative Science Quarterly 24(2), 181–199 (1979)
[32] Podolny, J.M., Baron, J.N.: Resources and relationships: Social networks and mobility in the workplace. American Sociological Review 62(5), 673–693 (1997)
[33] Tichy, N.M., Tushman, M.L., Fombrun, C.: Social network analysis for organizations. Acad. Management Rev. 4(4), 507–519 (1974)
[34] Coleman, J.S.: Social capital in the creation of human capital. American Journal of Sociology 94(S), 95–120 (1988)
[35] Coleman, J.S.: Foundations of Social Theory. Belknap Harvard University Press, Cambridge (1990)
[36] Nonaka, I.: A dynamic theory of organizational knowledge creation. Organization Science 5(1), 14–37 (1994)
[37] Granovetter, M.: The strength of weak ties. American Journal of Sociology 78(6), 1360–1380 (1973)
[38] Marsden, P., Campbell, K.: Measuring tie strength. Soc. Forces 63(2), 482–501 (1984)
[39] Hansen, M.T.: The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. Administrative Science Quarterly 44(1), 82–111 (1999)
[40] Uzzi, B.: Embeddedness in the Making of Financial Capital: How Social Relations and Networks Benefit Firms Seeking Financing. American Sociological Review 64(4), 481–505 (1999)
[41] Erickson, B.H.: The relational basis of attitudes. In: Wellman, B., Berkowitz, S. (eds.) Social Structures: A Network Approach, pp. 99–121. Cambridge University Press, NY (1988)
[42] Jehn, K.A., Northcraft, G.B., Neale, M.A.: Why Differences Make a Difference: A Field Study of Diversity, Conflict, and Performance in workgroups. Administrative Science Quarterly 44(4), 741–763 (1999)
[43] Allbritton, M.: A model of contractual project-based work: Personal social network connectivity. In: ICT Use, and Self-Monitoring, Syracuse University (2006)
[44] Ang, S.H., Leong, S.M.: Out of the mouths of babes: Business ethics and youths in Asia. Journal of Business Ethics 28(2), 129–144 (2000)
[45] Lee, D.Y., Dawes, P.L.: Guanxi, trust, and long-term orientation in Chinese business markets. Journal of International Marketing 13(2), 28–56 (2005)
[46] Su, C., Sirgy, M.J., Littlefield, J.E.: Is guanxi orientation bad, ethically speaking? A study of Chinese enterprises. Journal of Business Ethics 44(4), 303–312 (2003)
[47] Nunner-Winkler, G.: Two moralities? A critical discussion of an ethic of care and responsibility versus an ethic of rights and justice. In: Kurtines, W.M., Gewirtz, J.L. (eds.) Morality, Moral Behavior and Moral Development, pp. 348–361. Wiley, New York (1984)

Analysis of Tourist Subjective Data in Smartphone Based Participatory Sensing System by Interactive Growing Hierarchical SOM

Takumi Ichimura, Shin Kamada, and Kosuke Kato

Abstract. A self organizing map (SOM) is trained using unsupervised learning to produce a two-dimensional discretized representation of input space of the training cases. Growing Hierarchical SOM is an architecture which grows both in a hierarchical way representing the structure of data distribution and in a horizontal way representing the size of each individual maps. The control method of the growing degree by pruning off the redundant branch of hierarchy in SOM has been proposed and the criteria were designed by the adjustment of parameter settings according to the quantization error and the size of map. Moreover, the interface tool for the proposed method called interactive GHSOM has been developed. A mobile phone participatory sensing system in tourist subjective data collection has been developed with Android smartphone. In this paper, we examined to classify the subjective data by interactive GHSOM and the results were compared with knowledge extracted by C4.5.

1 Introduction

The current information technology can collect various data sets because the recent tremendous technical advances in processing power, storage capacity and network connected cloud computing. The sample record in such data set includes not only numerical values but also language, evaluation, binary data such as pictures. The technical method to discover knowledge in such databases is known to be a field of data mining and developed in various research fields.

Takumi Ichimura · Shin Kamada
Faculty of Management and Information Systems, Prefectural University of Hiroshima,
1-1-71, UjinaHigashi, Minami-ku, Hiroshima 731-8558, Japan
e-mail: ichimura@ieee.org

Kosuke Kato
Faculty of Applied Information Science, Hiroshima Institute of Technology,
2-1-1 Miyake, Saeki-ku, Hiroshima 731-5193 Japan
e-mail: k.katoh.me@it-hiroshima.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 225–235.
springerlink.com

The data mining is seen as an increasingly important tool by modern business to transform unprecedented quantities into business intelligence giving an informational advantage. Some data mining tools are currently used in a wide range of profiling practices such as marketing, fraud detection and medical information. The traditional method in data mining tools includes Bayes' theorem and regression analysis. As data sets have growth in size and complexity, automatic data processing technique in the field of computer sciences is required. The data mining in the field of neural networks(NNs), clustering, genetic algorithms, decision trees and support vector machines come into existence in the research of innovative soft computing methodologies.

Data mining is the process of applying these methods to data with the intention of uncovering hidden patterns. An unavoidable fact of data mining is that the subsets of data being analyzed may not be representative of the whole domain, and therefore may not contain examples of certain critical relationships and behaviors that exist across other parts of the domain. Moreover, the probability of inclusion of missing data and/or contradictory data becomes high because the means or instrumentality for storing information is to store the raw data in the storage by the automated collecting process.

Self organizing map (SOM)[1] is a type of artificial neural network that is trained using unsupervised learning to produce a low dimensional, discretized representation of the input space of the training samples, called a map. SOM is known to be an effective clustering method because it can learn regardless of data size and can intuitively show clustering results visually using maps. However, the clustering result by SOM has ambiguity because the boundary of clusters is not clear. In order to improve the clustering capability, Rauber et al. proposed a basic algorithm of the growing hierarchical self organizing map (GHSOM)[2]. The algorithm has been chosen for its capability to develop a hierarchical structure of clustering and for the intuitive outputs which help the interpretation of the clusters. However, GHSOM divides a data set into sub clusters immoderately if the distribution of samples is complex. There is a trade off for human designers between the investigation of the shape of a partial detailed cluster and the entire distribution of samples. In order to grasp an overview of tree structure of GHSOM, we propose an interactive GHSOM [3] to restrain the growing of hierarchy in GHSOM by reforming the map in each layer interactively. The criteria for interactive GHSOM are designed by the adjustment of parameter settings according to the quantization error and the size of map.

Mobile phone based Participatory Sensing (MPPS) systems involve a community of users sending personal information and participating in autonomous sensing through their mobile phones. Sensed data can be obtained from sensing devices present on mobiles such as audio, video and motion sensors, the latter available in high-end mobile phones. Sensed data can also be obtained from external sensing devices that can communicate wirelessly to the phone. For example, the technology for determining the geographic location of cell phones and other hand-held devices is becoming increasingly available. The tourist subjective data collection system with Android smartphone has been developed [4]. The application can collect subjective data such as pictures with GPS, geographic location name, the evaluation and comments in real sightseeing spots where a tourist visits and more than 500 subjective data are stored in the database. In this paper, we examined to classify them to investigate the effectiveness of interactive GHSOM and the results of interactive GHSOM are compared to the knowledge extracted by C4.5 [5].

The remainder of this paper is organized as follows. In section 2, the original algorithm of GHSOM is explained briefly. Section 3 describes the algorithm of interactive GHSOM and its interface tool. Experimental classification results of MPPS data by interactive GHSOM and C4.5 were investigated in Section 4. In Section 5, we give some discussions to conclude this paper.

2 Growing Hierarchical SOM

This section describes a basic algorithm of GHSOM [2]. The algorithm has been chosen for its capability to develop a hierarchical structure of clustering and for the intuitive outputs which help the interpretation of the clusters. These capabilities allow different classification results from rough sketch to very detailed grain of knowledge. This technique is a development of SOM, a popular unsupervised neural network model for the analysis of high dimensional input data [1]. Fig.1 shows the overview of hierarchy structure in GHSOM. Fig.2 shows the algorithm of GHSOM.

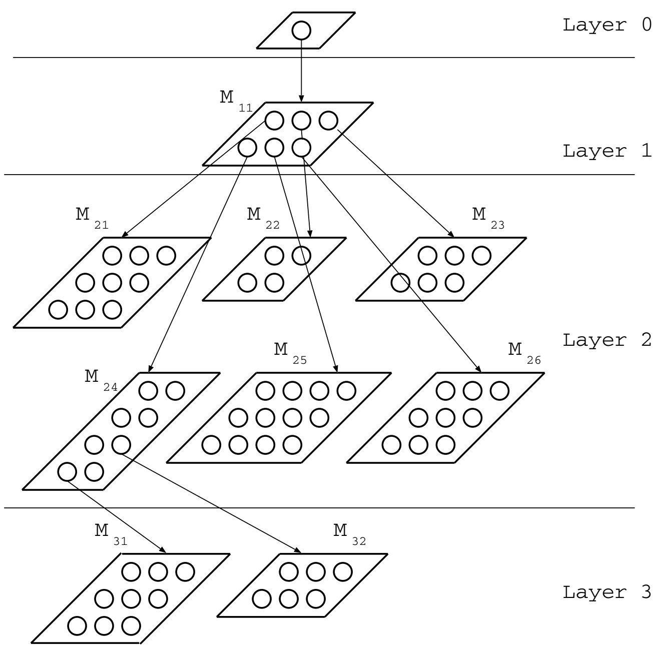


Fig. 1 A Hierarchy Structure in GHSOM

Step 1) All parameters are initialized.

Step 2) The mean quantization error of a unit i is calculated in the layer 0 in Fig.1.

$$mqe_0 = \frac{1}{n_I} \sum_{\mathbf{x}_i \in I} \| \mathbf{m}_0 - \mathbf{x}_i \|, n_I = |I|, \tag{1}$$

where $\mathbf{m}_0$ is the average of input samples, n_I is the number of input data set I, and $\mathbf{x}_i$ is an input vector for a sample.

Step 3) Let $\mathbf{M}_{\ell,v} (v = 1, 2, \cdots)$ be a map in the layer $\ell (= 1, 2, \cdots)$. The initial size of $\mathbf{M}_{\ell,v}$ is $u_0 (= 2 \times 2)$. The SOM clustering algorithm is employed in each map. The samples are divided into subcategories I and let k be an each winner unit in a subcategory.

Step 4) Calculate the mean quantization error for each winner unit k in $\mathbf{M}_{\ell,v}$ by using Eq.(2) and Eq.(3).

$$\begin{aligned} mqe_k = \frac{1}{n_C} \sum_{\mathbf{x}_j \in \mathbf{C_k}} \| \mathbf{m}_k - \mathbf{x}_j \|, \\ n_C = |\mathbf{C}_k|, \mathbf{C}_k \neq \phi, \end{aligned} \tag{2}$$

$$qe_k = \sum_{\mathbf{x}_j \in \mathbf{C_k}} \| \mathbf{m}_k - \mathbf{x}_j \|, n_C = |\mathbf{C}_k|, \mathbf{C}_k \neq \phi, \tag{3}$$

where mqe_k is the mean quantization error and qe_k is the quantization error. The $\mathbf{m}_k$ is the reference vector of winner unit k, $\mathbf{C}_k$ is the samples which are allocated to the unit k, and $\mathbf{x}_j$ belongs to $\mathbf{C}_k$.

Step 5) Let e be a unit representing the largest error between input and reference vectors among winner units in $\mathbf{M}_{\ell,v}$. Compare the reference vector of the unit e and its neighbor units, let d be the unit with largest difference in the neighbor units.

Step 6) Calculate the mean quantization error $mqe_{\mathbf{M}_{\ell,v}}$ for the subset $\mathbf{u}$ of a winner unit k in $\mathbf{M}_{\ell,v}$ by using Eq.(4)

$$mqe_{\mathbf{M}_{\ell,v}} = \frac{1}{n_u} \sum_{k \in \mathbf{u}} qe_k, n_u = | \mathbf{u} | \tag{4}$$

Step 7) If the mean quantization error is satisfied with Eq.(5), units are inserted in row/column as shown in Fig.3.

$$mqe_{\mathbf{M}_{\ell,v}} \geq \tau_1 qe_w, \tag{5}$$

where w is the unit in the map of the layer $\ell - 1$.

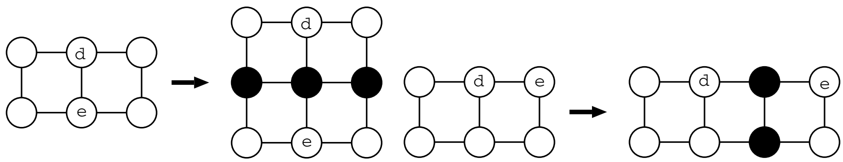
Step 8) The initial weights are given as the average of d and e.

Step 9) If the mean quantization error is not satisfied with Eq.(5), the insertion process of unit stops and returns to the map in the above layer and SOM is employed in the clustering in the map.

Step 10) After Step 9), if the quantization error is larger(Eq.(6)), then add a new layer to the map.

$$qe_k \geq \tau_2 qe_0 \tag{6}$$

Fig. 2 The algorithm of GHSOM



(a) generation pattern (1) (b) generation pattern (2)

Fig. 3 Unit Generation Process in GHSOM

3 Interactive GHSOM and Its Android Smartphone Tool

3.1 Control of Growing Hierarchies

The process of unit insertion and layer stratification in GHSOM works according to the value of τ_1 and τ_2. Because the threshold of their parameters gives a criterion to hierarchies in GHSOM, GHSOM cannot change its structure to data samples adaptively while training maps. Therefore, only a few samples are occurred in a terminal map of hierarchies. In such a case, GHSOM has a complex tree structure and many nodes(maps). As for these classification results, the acquired knowledge from the structure is lesser in scope or effect in the data mining. When we grasp the rough answer from the specification in the data set, the optimal set of parameters must be given to a traditional GSHOM. It is very difficult to find the optimal values through empirical studies.

We propose the reconstruction method of hierarchy of GHSOM even if the deeper GHSOM is performed. A stopping criterion for stratification is defined. Moreover, if the quantization error is large and the condition of hierarchies is not satisfied, the requirement for redistribution of error is defined.

Case 1) If Eq.(7) and the hierarchies are satisfied, stop the process of hierarchies and insert new units in the map again.

$$n_k \leq \alpha n_I, \tag{7}$$

wheren_k, n_I mean the number of input samples for the winner unit k and of the all input samples I, respectively. The α is a constant.

Case2) If Eq.(6) is not satisfied, the addition of layer is not executed, the quantization error of a unit is larger than the quantization error in an overall map. If Eq.(6) is satisfied, then a new unit is inserted.

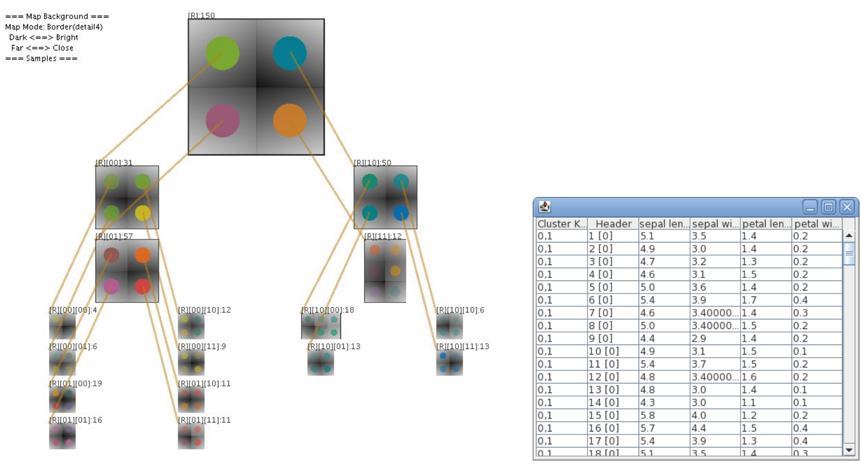
$$qe_k \geq \beta \tau_1 \sum_{y \in \mathbf{S}_k} qe_y, \tag{8}$$

where $\mathbf{S}_k$ is the set of winner units k.

3.2 Android Smartphone Based Interface of GHSOM

We developed the Android smartphone based interface of interactive GHSOM to acquire the knowledge intuitively. Android consists of a kernel based on the Linux with middleware, libraries and APIs. Smartphone and Tablet are equipped the eminent human computer interface with easy operations such as tap, flick, drag, and pinch. This tool was developed by Java language and was porting it to Android. Fig. 4(a) shows the clustering results by GHSOM. When we touch a unit in a map, the other window as shown in Fig. 4(b) is displayed and the samples in allocated in the corresponding unit is listed in the table. The color of unit shows the pattern of sample represented in Munsell color system [7], which consists of three independent dimensions: hue, value, and chroma. The similar color of units represents an intuitive understanding of similar pattern of samples. If the number of units in a map are increased, only a few samples could be classified into a new generated unit. Once the unit connected to the map is selected, the method re-calculates to find an optimal set of weights in the local tree structure search and then a better structure is depicted.

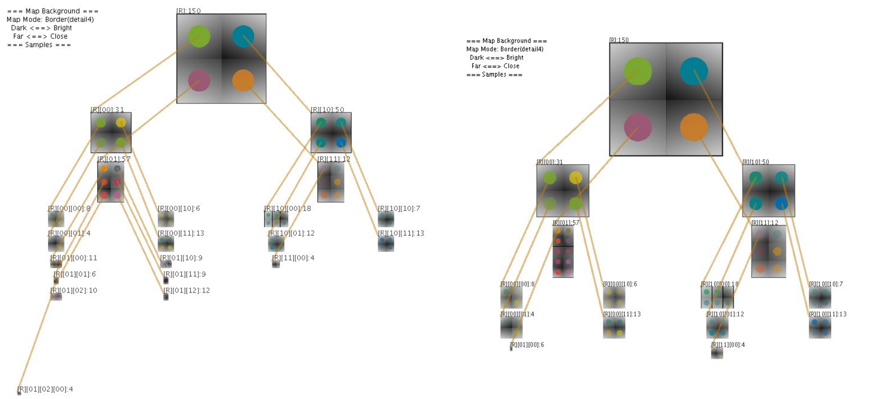
Figs. 5(a) and 5(b) show the clustering results of Iris dataset [6] by using interactive GHSOM. The left side as shown in Fig.5(a) is first representation of calculation result by GHSOM. The right side as shown in Fig.5(b) was the calculation result



Cluster K...	Header	sepal len...	sepal wi...	petal len...	petal wi...
0,1	1 [0]	5.1	3.5	1.4	0.2
0,1	2 [0]	4.9	3.0	1.4	0.2
0,1	3 [0]	4.7	3.2	1.3	0.2
0,1	4 [0]	4.6	3.1	1.5	0.2
0,1	5 [0]	5.0	3.6	1.4	0.2
0,1	6 [0]	5.4	3.9	1.7	0.4
0,1	7 [0]	4.6	3.40000...	1.4	0.3
0,1	8 [0]	5.0	3.40000...	1.5	0.2
0,1	9 [0]	4.4	2.9	1.4	0.2
0,1	10 [0]	4.9	3.1	1.5	0.1
0,1	11 [0]	5.4	3.7	1.5	0.2
0,1	12 [0]	4.8	3.40000...	1.6	0.2
0,1	13 [0]	4.8	3.0	1.4	0.1
0,1	14 [0]	4.3	3.0	1.1	0.1
0,1	15 [0]	5.8	4.0	1.2	0.2
0,1	16 [0]	5.7	4.4	1.5	0.4
0,1	17 [0]	5.4	3.9	1.3	0.4
0,1	18 [0]	5.1	3.5	1.4	0.3

(a) Simulation Result (b) A Detailed Description

Fig. 4 Simulation Result for Iris data set



(a) Classification samples with unclear data (b) Re-classification by interactive GHSOM

Fig. 5 Simulation Result by Interactive GHSOM

by interactive GHSOM, when a user pushes the unit: $[R][01][02][00]$. The notation represents the location of unit in the connection from the top level.

4 Experimental Results

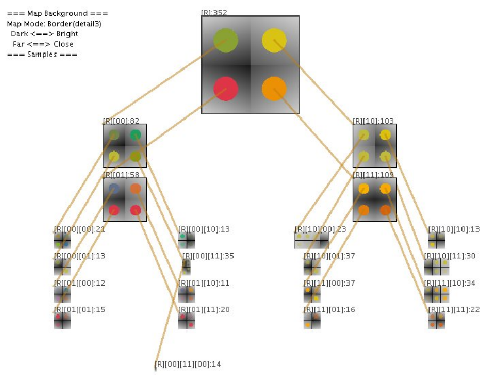
4.1 Tourist Subjective Data in MPPS

Participation of mobile phone users in sensorial data collection both from the individual and from the surrounding environment presents a wide range of opportunities for truly pervasive applications. For example, the technology for determining the geographic location of cell phones and other hand-held devices is becoming increasingly available. Our developed Android smartphone application [8] can collect the tourist subjective data in the research field of MPPS. The collected subjective data consist of jpeg files with GPS, geographic location name, the evaluation of $\{0,1,2,3,4\}$ and comments written in natural language at sightseeing spots to which a user really visits. The application must obtain GPS data before taking a picture so that the pictures provide evidence to prove that the tourist visited there. More than 500 subjective data are stored in our developed database through MPPS. Because most of user comments are short messages within 256 letters such as Twitter, the supplement information should be retrieved from web such as tourism association websites [9, 10, 11, 12, 13, 14, 15] and tourism blogs [16, 17]. The words in the websites are extracted and numerical value evaluated by TF-IDF (term frequency inverse document frequency) method [18] are added to the subjective data. The TF-IDF calculates a weight often used in information retrieval and text mining. In this paper, the partial information except comments and images are analyzed

Table 1 Sample records in our developed MPPS

Lat.	Lon.	Alt.	Name	Eva.	Date	Time
34.363369	132.470307	32.30	Oyster Street	2	2011-04-29	14-36-42
34.484011	132.269203	258.8	Fishing Lake	3	2011-04-30	12-06-17
34.484362	132.269326	272.6	Fishing Lake	4	2011-04-30	12-38-04
34.473791	132.240430	356.2	Rodge	1	2011-04-30	13-09-42
34.367706	132.175777	357.5	Futae Yaki(Cake)	4	2011-04-30	14-24-05
34.388838	132.103882	575.7	Spa Rakan	4	2011-04-30	14-57-39
34.393745	132.436148	41.4	Game spot	3	2011-05-06	20-26-58
34.410682	133.197108	174.8	Onomichi	4	2011-07-16	15-23-42
34.387643	132.430239	50.7	Tomato noodle	4	2011-07-20	20-56-38
34.393464	132.459653	52.3	High quality Japanese Restaurant	4	2011-07-22	18-49-54

(a) First classification of MMPS



(b) Re-classification

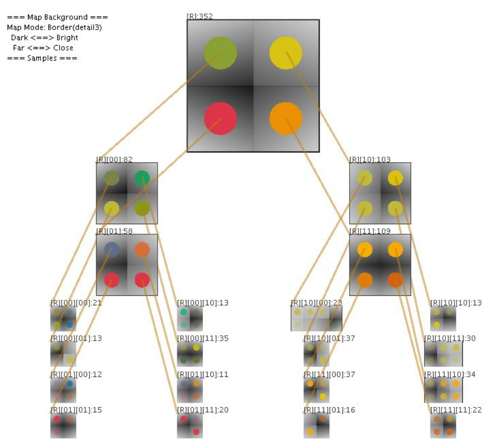


Fig. 6 Simulation Result by Interactive GHSOM

```
| | | lon <= 132.408 :
| | | | lat <= 34.3096 :
| | | | | lat > 34.3016 : [R][00][11][00][12] (5.0/2.0)
| | | | | lat <= 34.3016 :
| | | | | | evaluation <= 0 : [R][00][10][11] (2.0)
| | | | | | evaluation > 0 : [R][00][11][00][10] (4.0/1.0)
| | | | lat > 34.3096 :
| | | | | lat <= 34.3367 :
| | | | | | lon <= 132.308 : [R][00][11][00][03] (3.0)
| | | | | | lon > 132.308 : [R][00][11][00][13] (2.0)
| | | | | lat > 34.3367 :
| | | | | | evaluation <= 0 : [R][00][10][01] (2.0)
| | | | | | evaluation > 0 :
| | | | | | | lat <= 34.4228 : [R][00][11][01] (2.0)
| | | | | | | lat > 34.4228 : [R][00][11][00] (2.0)
| | | lon > 132.408 :
| | | | evaluation <= 0 : [R][00][10][00] (7.0)
| | | | evaluation > 0 :
| | | | | lat > 34.3776 : [R][00][11][10] (7.0)
| | | | | lat <= 34.3776 :
| | | | | | lon <= 132.452 : [R][00][11][12] (3.0)
| | | | | | lon > 132.452 : [R][00][11][11] (7.0)
```

(a) Tree structure of C4.5 (before)

```
| | lon <= 132.485 | | | evaluation <= 0 :
| | | | lon > 132.424 : [R][00][10][00] (7.0)
| | | | lon <= 132.424 :
| | | | | lat <= 34.33 : [R][00][10][11] (2.0)
| | | | | lat > 34.33 : [R][00][10][01] (2.0)
| | | evaluation > 0 :
| | | | lat <= 34.3128 :
| | | | | lon <= 132.356 : [R][00][11][10] (14.0)
| | | | | lon > 132.356 : [R][00][11][00] (2.0)
| | | | lat > 34.3128 :
| | | | | lat <= 34.3798 : [R][00][11][11] (10.0)
| | | | | lat > 34.3798 : [R][00][11][01] (9.0)
```

(b) Tree structure of C4.5 (after)

Fig. 7 Knowledge part changed by interactive GHSOM

by interactive GHSOM, because the extracting method should be discussed on the technologies in web intelligence and we will describe in near feature [19]. Table 1 shows some samples of sightseeing spots collected by our developed MPPS.

4.2 Data Representation in Interactive GHSOM

Fig. 6 shows the classification results for tourist subjective data by interactive GHSOM. Fig. 1(a) was a result of classification. When the node on the map, [R][00][11][00] was clicked, GHSOM calculated to classify samples in the node and the connected upper map again and depicted Fig. 1(a).

Fig. 7 shows a part of tree structure of knowledge extracted by C4.5. The prior structure as shown in Fig. 7(a) became simple knowledge as shown in Fig. 7(b) by GHSOM implementation. The result represents that the GHSOM tool can give the intuitive graphical comprehension for real world data such as tourist subjective data in MPPS.

5 Conclusive Discussion

We proposed an interactive GHSOM to restrain the growing of hierarchy in GHSOM by reforming the map in each layer interactively. The interactive GHSOM and its graphical tool on Android smartphone can give intuitive comprehension of multi dimensional data. In order to verify the effectiveness of our proposed method, we developed Android smartphone based Participatory Sensing system for Hiroshima tourism information which can collect the tourist subjective data in sightseeing spots. As computational results, the tree structure by GHSOM almost equals the knowledge extracted by C4.5. The differences between classification results as shown in Fig.7 are represented knowledge in If-Then rule. The GHSOM analysis tool on the Android smartphone is expected to be solved the problems in real world and within real time.

Located in Hiroshima, which is remote from big cities, it is well-known widely as a sightseeing spot of World Heritage Site and a lot of guests seem to come all the way from Kyoto, Osaka, and Tokyo. Certainly, Hiroshima has two spots of UNESCO World Heritage Site. However, the recent utilization of local resources for tourism makes no difference to our desire to travel, because the means of transport has been developed from downtown to peripheral suburbs. Not only tourism association but the local citizens should give the innovative and attractive information in sightseeing to visitors. Our developed clustering system and MPPS for tourism information can discover valuable information in the unknown user subjective data. However, the analysis of the main subjective data such as pictures does not implemented. We will develop the image analysis method in tourism information in future.

References

1. Kohonen, T.: Self-Organizing Maps. Springer Series in Information Sciences, vol. 30. Springer, Heidelberg (1995)
2. Rauber, A., Merkl, D., Dittenbach, M.: The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. IEEE Transactions on Neural Networks 13, 1331–1341 (2002)
3. Ichimura, T., Yamaguchi, T.: A Proposal of Interactive Growing Hierarchical SOM. In: Proc. of 2011 IEEE SMC 2011, pp. 3149–3154 (2011)
4. Kamada, S., Ichimura, T.: Android Smartphone Based Collection System of Sightseeing Spot Information in Hiroshima. In: 2011 IEEE SMC Hiroshima Chapter Young Researchers WorkShop, pp. 21–24 (2011) (in Japanese)
5. Covoer, T., Hart, P.: Nearest neighbor pattern classification. Trans. on IEEE Information Theory 13(1), 21–27 (1967)
6. Fisher, Iris Dataset, UCI Machine Learning Repository (1936), http://archive.ics.uci.edu/ml/datasets/Iris
7. Kuehni, R.G.: The early development of the Munsell system. Color Research and Application 27(1), 20–27 (2002)
8. ITProducts, Hiroshima Tourist Map, https://market.android.com/details?id=jp.itproducts.KankouMap (retrieved November 15, 2011)
9. Tourism Promotion Office of the Hiroshima Prefectural Government, Hiroshima Tourist Navigator, http://www.kankou.pref.hiroshima.jp/foreign/english/index.html (retrieved November 15, 2011)
10. The Tourism Association of Kure City, http://www.urban.ne.jp/home/kurecci/ (retrieved November 15, 2011) (in Japanese)
11. The Tourism Association of Hatsukaichi City, Hatsukaichi Travell Navigator, http://www.hatsu-navi.jp/ (retrieved November 15, 2011)
12. The Tourism Association of Onomichi City, Onomichi Tourist Association: ONONAVI, http://www.ononavi.com/ (retrieved November 15, 2011)

13. The Tourism Association of Fukuyama City, http://www.fukuyama-kanko.com/event/ (retrieved November 15, 2011) (in Japanese)
14. The Tourism Association of Miyoshi City, http://www.kankou-miyoshi.jp/ (retrieved November 15, 2011) (in Japanese)
15. Dept. of the Tourism, Akitakata City Office, http://www.akitakata.jp/kankou/ (retrieved November 15, 2011) (in Japanese)
16. Nachan's playground in Hiroshima -Personal Blog-, http://www.hiroshima-asobiba.net/ (retrieved November 15, 2011) (in Japanese)
17. Hiroshima Outdoor Map, User Participatory Sensing Blog, http://hiroshima.o-map.com/ (retrieved November 15, 2011) (in Japanese)
18. Wu, H.C., Luk, R.W.P., Wong, K.F., Kwok, K.L.: Interpreting TF-IDF term weights as making relevance decisions. ACM Transactions on Information Systems 26(3), 137 (2008)
19. Kamada, S., Ichimura, T.: Analysis of Tourist Subjective Text Data in Smartphone based Participatory Sensing System by Interactive Growing Hierarchical SOM. In: Proc. of Chugoku & Shikoku Research Workshop, Japan Society for Fuzzy Theory and Intelligent Informatics (2011) (to appear) (in Japanese)

Framework on the Innovative Education System for Entrepreneurship

Mei-Chen Lo

Abstract. Education is the comprehensive activity in which we come to know ourselves and the world around us. The education system change is a huge issue in any country as of their multi-level education concerns. Since the technology lead our way of behavior change even to our thinking of life are gradually move in adoption new tool assistance for education. Therefore, a great deal of activity in what might be broadly termed the field of enterprise and entrepreneurship education in schools and colleges are important. This study presents the concepts for innovative education system (IES) and enterprising characteristics for fitting the requirement of high-tech enterprise activities. The framework is built to link each causal-factor for structure study and tasks involvement which gives the ideas of forming an IES to be worked and usefulness.

Keywords: Innovation management, Innovative education system (IES), Entrepreneurship.

1 Introduction

Education is the comprehensive activity in which we come to know ourselves and the world around us. It is the activity in which a society transmits its traditions, its values, its hopes and its fears from one generation to the next. Throughout the world there is currently a great deal of activity in what might be broadly termed the field of enterprise and entrepreneurship education in schools and colleges. In the Taiwan, a substantial number of different programs are developed with support of a variety of private and public organizations. Although, there are many activities at the local, regional and national levels aimed at different target groups and which involve, in their delivery, many types of institutions. Moreover, they demonstrate deferent levels of innovation and degrees of national penetration.

Mei-Chen Lo
Department of Business Management, National United University
No.1, Lienda Rd., Miaoli 36003, Taiwan
e-mail: meichen_lo@nuu.edu.tw

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 237–244.
springerlink.com

Many of the programs for enterprise or entrepreneurship education throughout the world take place within, or as an adjunct of, the innovative education system (IES). In this respect a number of important issues arise in relation to:

- acceptance of this form of education by schools, education advisers and indeed the formal educational system as a whole;
- the degree to which the programs can link with the mainstream school curriculum and form part of it;
- and the degree to which the education system ought to be involved in the development of particular personal competencies or skills hitherto rather more associated with training than education.

These issues are important in understanding the attitudes of teachers, education authorities and pupils towards the entrepreneurship and enterprise education curriculum. The issue of enterprise and entrepreneurship education is controversial in a number of respects.

Entrepreneurship education is focused upon the development of personal attributes. It does not necessarily embrace the context of the 'for profit' small business or indeed the entrepreneur although arguably it is substantially linked with the development of an enterprise culture within which the entrepreneur will flourish [8].

2 Entrepreneurship Education in the IES

The lecture has been praised as a cost-effective method for providing new information to a large audience. For example, some reports have suggested that students who are satisfied with the quality of lectures and teaching perform better on tests of knowledge and measures of learning. On the other hand, other studies have criticized the lecture as being passive and ineffective. This is consistent with educational theory, which suggests that learning must be an active and an innovative process where new concepts are incorporated into an individual's existing knowledge. One potential strategy to improve the lecture is to increase interactivity—for example, by using a computerized audience response system, which enables the teacher to pose specific questions and immediately assess the student's level of knowledge.

2.1 Entrepreneurial Culture and Education System

Entrepreneurship course features considered most important were development of a business plan project and entrepreneurs as speakers and role models. Cases ranked next in importance followed by lectures and assigned readings. Besides, an interactive lecturing has been shown to enhance students' attention, test scores, productivity, and interaction.

2.1.1 Interactive Lecturing

Lectures as a method of teaching and transmitting information have come under increasing criticism [1], [11],[12]. One of the major reasons for this critique is the observation that lectures are less effective than other methods when instructional goals involve the application of information or facts, the development of thinking skills, or the modification of attitudes [6], [14], [15]. In addition, students are frequently seen as passive recipients of information, and as a result, not engaged in the learning process.

Interactive lecturing involves an increased interchange between teachers, students and the lecture content. The use of interactive lectures can promote active learning, heighten attention and motivation, give feedback to the teacher and the student, and increase satisfaction for both. The value of interactive lecturing rests on the premise that active participation and involvement is a prerequisite for learning beyond the recall of facts, and that students must be attentive and motivated in order for learning to occur. For the reason of coping with the change in innovative technology development, some new technology devises to be adopted in class activities are encouraged; for example, the audience response system is a wireless technology in which each course attendee can respond to questions posed by the speaker on an individual keypad that registers an anonymous response. Each response is collected and summarized into aggregate statistics that can then be displayed to the entire group virtually instantaneously. (Interactive technology such as audience response system has been available for more than 30 years.)

2.1.2 Interactive Techniques

Some commonly used interactive techniques, such like [20]: (1) Breaking the class into smaller groups; (2) Questioning the audience; (3) Straightforward questions (e.g. Brainstorming, Rhetorical questions); (4) Surveying the class; (5) Quizzes and short answers; (6) Using audience responses; (7) Use of empirical cases; (8) Use of written materials; (9) Organizing debates, reaction panels and guests; (10) Using simulations and role plays; (11) Using films and videotapes; (12) Audiovisual aids; (13) Using effective presentation skills.

However, while many teachers accept the notion that other teaching methods might be better than lectures for encouraging students to be more actively involved in learning, and for promoting the application of knowledge, few have the time, resources or opportunity to use the small group methods that promote such involvement and application [19]. Also, when done effectively, the lecture can transmit new information in an efficient way, explain or clarify difficult notions, organize concepts and thinking, challenge beliefs, model problem solving, and foster enthusiasm and a motivation for learning [7], [4], [5], [18].

Much has been written about entrepreneurship and a growing literature is emerging in entrepreneurship education and graduate entrepreneurship [10], [3]. Dainow [3] conducted a survey (1985-1994) of the entrepreneurship education literature for a ten-year-period, up to 1984. His goal was “to assess the current state of the art, based on published articles, with a view to identifying the strengths and weaknesses that can guide future efforts”.

2.2 Education and Training Concerns for Small Business and Entrepreneurship

Some study [3] and [9] presents a marked increase in the empirical research, particularly in the area of educational process and structure. Results indicate considerable consensus that entrepreneurship can be taught and that teaching methods, can be enhanced through active participation. In addition, there is a clear indication of considerable growth in entrepreneurship courses and programs, although there is very little consistency in approach. There is also preliminary evidence that entrepreneurial attributes can be positively influenced by educational programs and that much entrepreneurship. Nonetheless, there is also strong evidence of resistance by small business owner/managers to education and training. Overcoming this barrier will require recognition of the need to develop programs tailored to the specific needs of target markets.

Training as an entrepreneur, the work content may involve complex tasks which require certain degree understanding of the fields of their domain knowledge.

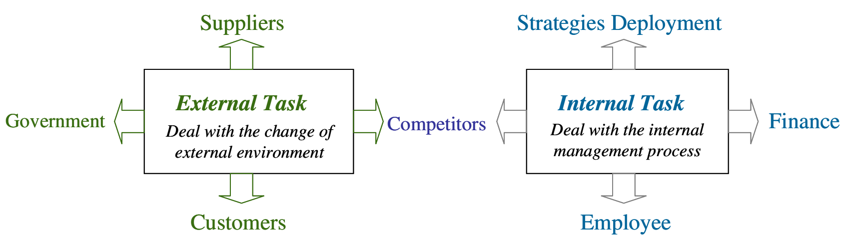


Fig. 1 Capabilities concerns for Start up a business – Entrepreneur

It may need to realize how the daily task is able to handle as self-employment. And the flow can be divided as two major aspects as of external and internal's, which the highlighted details can be recognize as Fig.1.

3 Characteristics of Education on Innovation Management for Entrepreneurship

For some time there has been widespread recognition that entrepreneurship is the engine that drives the economy of most nations. This has lead to an increasing interest in the development of education programs to encourage and enhance entrepreneurship and a recognition that much research needs to be carried out into what makes an entrepreneur and how these characteristics may best be imparted.

Overall, the current state of data in Taiwan has been insufficiently consistent and comprehensive to provide a sound platform upon which specific questions can be considered. For example:

- What should be the nature of future curricula development in enterprise and entrepreneurship education at universities, based upon current experiences?

– How many enterprises may like to involve the education task inside their business circle?
– How does existing course provision contribute to the development of entrepreneurial learning outcomes, and what might constitute good practice?
– What is the nature of the engagement by student types and by faculties/centers?
– What is the overall scale and scope of provision and engagement and how is this changing?

The emphasis is placed on presenting an overview with selected highlights of variations for enterprise education and non–accredited entrepreneurship support program. This includes: current and planned course provision over time; student profiles and targets; primary learning outcomes; non-accredited provision and student engagement; primary funding sources; and the development of a range of institutional characteristics conducive to supporting student enterprise and graduate entrepreneurship.

More so the focus for this paper is to build upon the existing understanding of frameworks of provision and engagement in support for student enterprise and graduate entrepreneurship. In this context there are few relevant studies that have comprehensively framewroked entrepreneurship education and support in Higher Education.

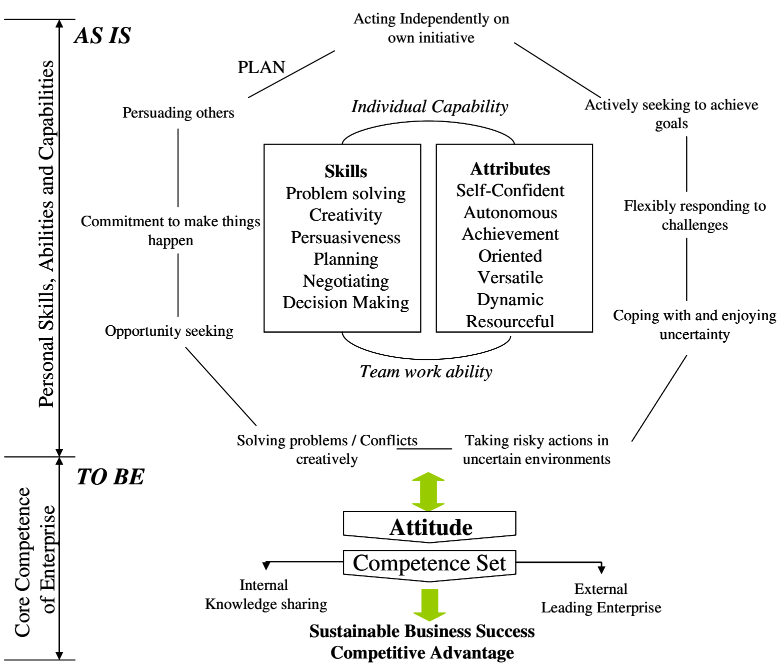


Fig. 2 ASIS-TOBE for Enterprises

The characteristics some of which can be more strictly defined in behavior terms and others as associated capabilities, skills and personal attributes [21], [16], [17], and [13].

Among the *AS-IS* are the personal skills (problem solving, creativity, persuasiveness, planning, negotiating and decision making), abilities (individual ability and team work ability), attributes (self-confident, autonomous, achievement oriented, versatile, dynamic, resourceful) and capabilities (planning, leading, communication, motivating people, analysis, coordination, flexibility, commitment and achievability) which initiate as acting independently on own initiative ambition to move thing changeable and accept the challenges. The personal attributes in enterprising are set out in Fig.2. Among the *TO-BE* are termed as of form the core competence of enterprise; which include attitude (demonstrating from intelligence transferability and pressure affordability) and competence set (in showing the capability accumulation from internal knowledge sharing system and external leading position). Then, the individual competence (e.g. self-awareness, self-confidence; creativity perseverance; persuasiveness; resourcefulness; negotiating skills; and motivation and commitment to achieve) expect to affect others to form competence set within enterprise training and activities. The concept is initiating from individual success and then turn to the accumulating competence to extend and to formulate group success as well (win-win concept). Therefore, the sustainable business success model can be able to reflect through the enterprise operation and to lead a niche position for its competitive advantage.

4 Discussions

A considerable number and variety of topics, approaches and directions are reflected in the conceptual papers reviewed for this article.

Of particular note is the need to distinguish among individual student, entrepreneurship, enterprise and small business management education and to differentiate each of these from traditional approaches to management education. This theme permeates the literature but is reflected most especially in articles dealing with suggested teaching strategies and proposed curricula. More specifically, the components of the ideal structure (as Fig.2) include the following: a focus on attributes, skills and capabilities as well as tasks, an element of concrete experience derived from active participation through projects and the like, and content directed to stage of venture development and emphasizing functional integration.

As Fig. 3, it demonstrates the way of thinking about the choice of individual, school resource allocation, enterprise opportunities (or requirements). The framework leads multi-layers design concept for giving clear look inside if the education system will be driven by the change of pressure to renovation or student's individual needs orient the change of our education system. From this study, the *SERVICE* concept is embedded in the education system and to derivate a novel idea on world academic resource integration. As for the IES with the resource integration through new technology devise system are suggested for consideration:

– Virtual devise in the hardware system in Education system (include library, school, classroom)
– Free access in learning
– Tutors assignment (for individual student-industrial and academic tutors)
– Knowledge pool (Q&A or blackboard – forums for researchers and students)
– Researcher's Heaven (networking research and open research program)

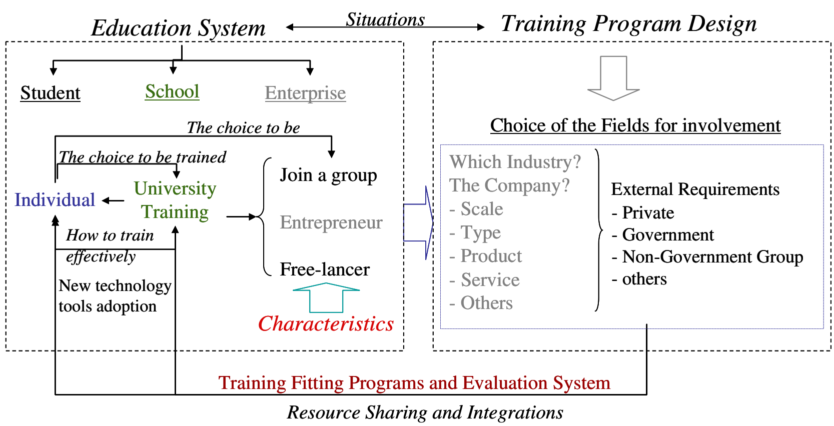


Fig. 3 Framework of Clarification on IES

5 Conclusion

It is explicit to the model of enterprise education that be developed here that a clear distinction can be made between such education and small business and entrepreneurship education and training. Much of enterprise education as it is being widely introduced into primary, secondary, further and higher education in the Taiwan, is not aimed directly at stimulating entrepreneurship whether defined as independent small business ownership or the development of opportunity-seeking leaders of high profile companies. Its major objective is to help develop enterprising people and in particular, to inculcate an attitude of self-reliance [2] through the 'process' of learning.

References

1. Bligh, D.: What' s the Use of Lectures?, England, Penguin (1972)
2. Cotton, J.: The Enterprise Education Experience. Education and Training 33(4), 6–13 (1991)
3. Dainow, R.: Training and education of entrepreneurs: the current state of the literature. Journal of Small Business and Entrepreneurship 3(4), 10–23 (1986)
4. Foley, R., Smilansky, J.: Teaching Techniques. McGraw Hill, New York (1980)

5. Frederick, P.: The lively lecture 8 Variations. College Teaching 34, 43–50 (1986)
6. Frederick, P.: Student involvement: active learning in classes. In: Weimer, M.G. (ed.) New Directions for Teaching and Learning, Teaching Large Classes Well, vol. 32, pp. 45–56. Jossey-Bass, San Francisco (1987)
7. Gage, N., Berliner, D.: Educational Psychology. Houghton-Mifflin, Dallas (1991)
8. Gibb, A.A.: Designing Effective Programmes for Encouraging the Small Business Start-Up Process. Journal of Education and Industrial Training 4, 24–32 (1987)
9. Gibb, A.A.: Enterprise Culture and Education: Understanding Enterprise Education and Its Links with Small Business, Entrepreneurship and Wider Educational Goals. International Small Business Journal 11(3), 11–34 (1993)
10. Gorman, G., Hanlon, D., King, W.: Some research perspectives on entrepreneurship education, enterprise education, and education for small business management: a ten-year literature review. International Small Business Journal, pp. 56–77 (April/June 1997)
11. Kimmel, P.: Abandoning the lecture: curriculum reform in the introduction to clinical medicine. The Pharos 55, 36–38 (1992)
12. Kroenke, K.: The lecture method: where it wavers. American Journal of Medicine 77, 393–396 (1984)
13. McGahan, A.M., Porter, M.E.: How Much Does Industry Matter, Really? Strategic Management Journal 18(Summer Special Issue), 15–30 (1997)
14. McKeachie, W.: Teaching Tips. MA, D.C. Heath and Co., Lexington (1994)
15. Newble, D., Cannon, R.: A Handbook for Medical Teachers. Kluwer Academic, Boston (1994)
16. Porter, M.E.: From Competitive Advantage to Corporate Strategy. Harvard Business Review, 43–59 (May/June 1987)
17. Porter, M.E.: Towards a Dynamic Theory of Strategy. Strategic Management Journal 12(Winter Special Issue), 95–117 (1991)
18. Saroyan, A., Snell, L.: Variations in lecturing styles. Higher Education 33, 85–104 (1997)
19. Schwartz, P.: Active small group learning with a large group in a lecture theatre: a practical example. Medical Teacher 11, 81–86 (1989)
20. Steinert, Y., Snell, L.S.: Interactive lecturing: strategies for increasing participation in large group presentations. Medical Teacher 21(1), 37–42 (1999)
21. Yu, P.L.: A class of solutions for group decision problems. Management Science 19(8), 936–946 (1973)

Bio-Inspired Image Analysis Using E-Ubias Cognitive Systems

Lidia Ogiela and Marek R. Ogiela

Abstract. This publication presents the idea of cognitive systems of data analysis for 3D reconstructions of image data. Systems that conduct cognitive analyses are designed so that semantic analyses can be used to interpret data. These types of systems are built for various types of data, but the primary subject of this publication is the presentation of cognitive systems for analysing many-dimensional medical images of large vessels, namely coronary arteries. An analysis conducted in UBIAS - Understanding Based Image Analysis Systems - will be used to illustrate the method of enhancing this class of systems with processes whereby these systems learn new solutions about which the systems have no knowledge. The proposed solution demonstrates the method of building a new class of systems: E-UBIAS - Extended Understanding Based Image Analysis Systems.

Keywords: semantic analysis, cognitive data analysis systems, UBIAS systems, E-UBIAS systems, multi-dimensional images, 3D reconstructions of images.

1 Introduction

Systems for the cognitive analysis of data are used for semantically analysing data of various types, from image data to numerical data [1]. One example of the first type of data consists in medical images which were used to analyse lesions found within the central nervous system, namely the spinal cord, lesions of long bones of lower and upper extremities, of foot, hand and wrist bones [4, 5, 6, 7, 8, 9, 10]. The flow in coronary vessels has also been analysed [21].

The semantic analysis and interpretation of data is made possible by defining a set of semantic features characteristic for the analysed data sets [11]. These types of analyses and interpretations are conducted using cognitive resonance, whose essence lies in indicating the consistencies found between the set of characteristic features of the analysed data and the set of expectations obtained on the basis of a base of expert knowledge available in the system. Cognitive resonance enables

Lidia Ogiela · Marek R. Ogiela
AGH University of Science and Technology
Al. Mickiewicza 30, PL-30-059 Krakow, Poland
e-mail: {logiela,mogiela}@agh.edu.pl

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 245–252.
springerlink.com

information systems to undertake semantic analyses of images using defined semantic actions. This type of analysis, applied in processes of reasoning about the structure of lesions, is carried out for medical images, as this type of data (image data) is characterised by a large quantity of semantic information [3, 12, 13, 14]. This information may concern:

- lesion occurrence;
- number of lesions present:
  - homogenous;
  - heterogeneous;
  - located in the same human organ;
  - located in various human organs;
- the size of every observed lesion;
- the length, width and thickness of the lesion;
- repetition frequency;
- lesion location etc.

Determining the semantics of the analysed lesions allows further processes to be carried out. These include reasoning about the current condition and projecting the future one (e.g. the treatment initiated, the atrophy of the lesion or its growth).

Analysis processes founded on the aspect of cognitive interpretation and reasoning are characteristic for cognitive systems, among which different data interpretation and analysis classes are distinguished [14, 15, 16]. The variety of these classes depends on the area to which the individual types of systems are dedicated (Fig. 1).

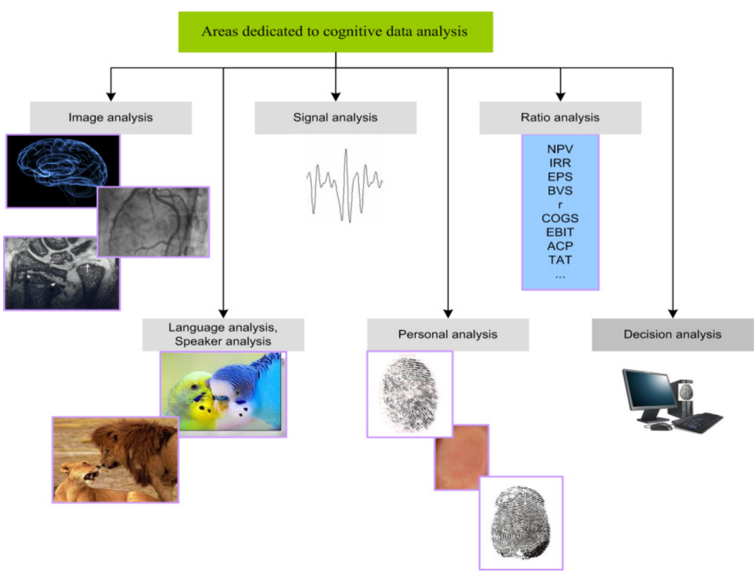


Fig. 1 Areas of cognitive analysis application

A detailed description of classes of cognitive data analysis systems can be found in the following publications [14, 17, 18, 19].

2 Analysis Processes in Cognitive Systems

Processes imitating human cognitive processes are considered to be the basic analysis processes executed in cognitive systems [1, 3, 14, 19]. Their characteristic feature is a certain non-determinism observed in people, which at the same time demonstrates their individual diversity and uniqueness.

These processes form the foundation of the design of cognitive information systems which, enhanced with semantic analysis aspects, lead to semantic reasoning and interpreting processes.

Cognitive analysis carried out in cognitive systems is based on the cognitive resonance which indicates the presence of links and consistencies between the set of expectations about the analysed data and the set of features characteristic for it. This phenomenon (Fig. 2) means that the presence of a consistency of the set of expectations with the set of characteristic features becomes the basis for carrying out a semantic data analysis, i.e. processes of data understanding.

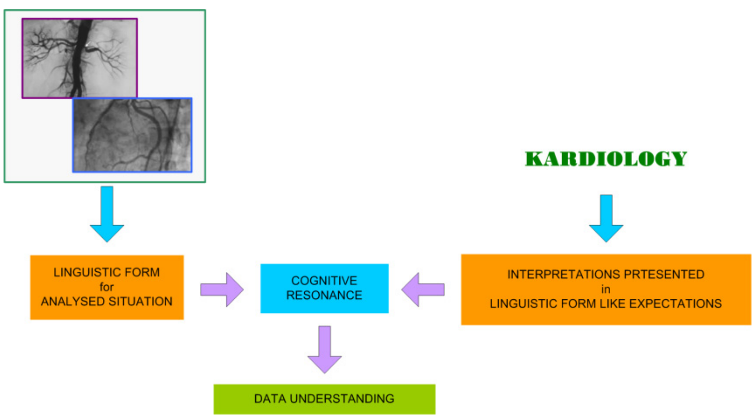


Fig. 2 The cognitive resonance and data understanding schema

The process of data understanding, illustrated in Fig. 2, and based on cognitive resonance, applies in particular to image analysis processes described in this publication. The main subject of this scientific dissertation is the analysis of medical 3D images portraying lesions in large vessels, namely coronary arteries. Images of this type are acquired, inter alia, in heart disease diagnostics used to assess the lesions of the cardiac muscle and also of coronary vessels. It is precisely the coronary vessel analysis which will be characterised and described in the section below.

3 3D Reconstructions of Image Data

3D image data is analysed in cognitive data analysis systems on the basis of defined linguistic formalisms. In the case of analysing lesions found in coronary vessels, context-free graph grammars have been used to semantically analyse the lesions observed. Such work was aimed at proposing grammar formalisms powerful enough for analysing 3D images of coronary arteries and vessels [21].

Images of arteries, just like other multi-dimensional images showing lesions in human organs, can be subjected to a semantic analysis due to the diversity of lesions occurring. The most frequent lesions of large vessels include:

- persistent Botall's arterial duct;
- pulmonary artery stenosis (a congenital defect);
- trilogy, tetralogy and pentalogy of Fallot;
- coronary artery sclerosis;
- and
- congenital as well as acquired heart defects contributing to changes (or lesions) in the structure of coronary arteries.

The above groups of lesions can be semantically analysed with regard to:

- the healthy structure of coronary arteries and vessels;
- enlarged cardiac cavities;
- the dilation of large vessels;
- valve enlargement or stenosis;
- valvular incompetence;
- combined mitral and aortal or mitral/aortal defects;
- multi-valvular defects;
- coronary artery and vessel sclerosis;
- the occurrence of aneurysms;
- ischemic heart disease; and
- the occurrence of tumours.

In the data analysis process, the identification of the lesion occurring, its location, size and the frequency of its occurrence makes it possible to determine the significance and the impact of the analysed lesion on the subsequent diagnostic and treatment process.

Data obtained in the analysis process includes:

- information on the patient's condition;
- recommended (applied) treatment process;
- the subsequent indispensable diagnostic process;
- the impact of the analysed lesion on other organs;
- determining the relationship between the lesion occurring and the possibility of other disease units appearing in the near future.

The analysis of coronary arteries carried out in cognitive systems based on images acquired as part of cardiologic diagnostic processes offers an opportunity to assess the risk of a heart lesion or disease occurring. 256-slice or 320-slice computed tomography represents the modern method of acquiring images in cardiologic diagnostics.

CT images obtained as part of diagnostic processes are semantically analysed by UBIAS - Understanding Based Image Analysis Systems – in which it is possible to identify the analysed lesion, its parameters and significance.

Figure 3 shows an example semantic analysis of a lesion found in a coronary artery.

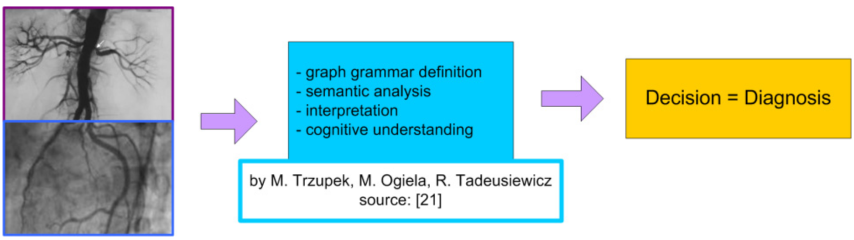


Fig. 3 An example UBMSS system – a very good cash-flow structure

Processes of the semantic analysis of coronary arteries have already been widely described in scientific literature and the reader can find an analysis of this subject in publication including the following [14, 21].

A significant novelty of this topic of the semantic analysis executed in UBIAS systems is the ability to enhance the operation of cognitive data analysis systems by adding stages at which they learn new solutions. This can occur if the UBIAS system is analysing an image about which it has no knowledge. In this case, the analysis process will fail.

The ability of the system to learn yields better results in the analysis processes conducted and at the same time enables the processes to be executed correctly for data for which the system has no defined pattern. As a superior class of semantic image analysis systems, E-UBIAS – Extended Understanding Based Image Analysis Systems – execute analysis and reasoning processes with the added stage at which the system learns new solutions (Fig. 4).

The solution proposed offers opportunities for conducting the semantic analysis of (new) acquired image data previously not defined in the system database and for that about which the system used to have no information (e.g. it was unable to compare the analysed characteristics of these images with expectations generated based on expert knowledge). The stage of introducing images that are new (for the system and its knowledge base) or of lesions not analysed previously offers opportunities to train the system in how to take decisions in a new situation. The training process can be based on the knowledge collected in the system or on knowledge acquired outside it. For such cases, it is necessary to extend the expert knowledge base which will be used to semantically analyse the new images (or sets of them).

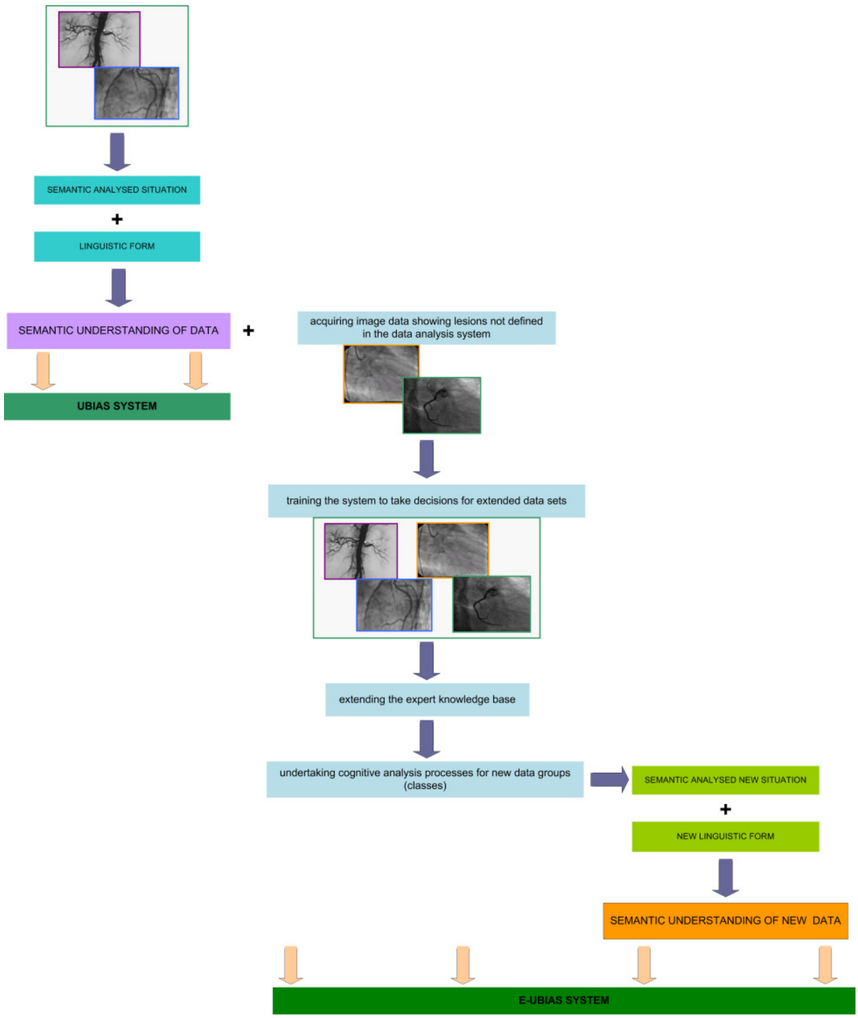


Fig. 4 E-UBIAS systems for analysing CT images of coronary arteries

4 Conclusion

The E-UBIAS systems proposed for the semantic analysis of data allow semantic analysis processes to be carried out for significantly extended groups of data. The semantic analysis conducted initially was described with regard to data sets for which the system running the analysis could provide the answer. So that semantic analysis related to a group, in a sense closed, of cases (defined in the system). This kind of an analysis was described for diverse classes of image data: medical images showing various lesions of human organs. The example cited in this publication was of a system analysing large vessels, i.e. coronary arteries.

The novelty of the proposed approach to methods for the cognitive analysis of data is the aspect of enhancing UBIAS systems with the systems' ability to learn how to take decisions in new situations, in cases for which the system does not know the answers. The ability of cognitive data analysis systems to learn how to take decisions and conduct the analysis for new solutions adds to the abilities known so far and offers an opportunity for the further development of semantic analysis systems. This development should go in the direction of using the processes of training cognitive information systems conducting analyses in various areas.

Acknowledgments. This work has been supported by the National Science Centre, Republic of Poland, under project number N N516 478940.

References

1. Cohen, H., Lefebvre, C. (eds.): Handbook of Categorization in Cognitive Science. Elsevier, The Netherlands (2005)
2. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. A Wiley-Interscience Publication, John Wiley & Sons, Inc. (2001)
3. Meystel, A.M., Albus, J.S.: Intelligent Systems – Architecture, Design, and Control. A Wiley-Interscience Publication John, Wiley & Sons, Inc., Canada (2002)
4. Ogiela, L.: Cognitive Understanding Based Image Analysis Systems (UBIAS) of the Diagnostic Type. In: IEEE International Workshop on Imaging Systems and Techniques – IST 2007, Krakow, Poland, May 4-5 (2007)
5. Ogiela, L.: UBMSS as an Example of the Application of Cognitive Analysis in Data Analysis. In: CISIM 2007, IEEE Proceedings 6th International Conference CISIM 2007 – Computer Information Systems and Industrial Management Applications, Ełk, Poland, June 28-30, pp. 77–80 (2007)
6. Ogiela, L.: Cognitive Systems for Medical Pattern Understanding and Diagnosis. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part I. LNCS (LNAI), vol. 5177, pp. 394–400. Springer, Heidelberg (2008)
7. Ogiela, L.: Modelling of Cognitive Processes for Computer Image Interpretation. In: Al-Dabass, D., Nagar, A., Tawfik, H., Abraham, A., Zobel, R. (eds.) EMS 2008 European Modelling Symposium, Second UKSIM European Symposium on Computer Modeling and Simulation, Liverpool, United Kingdom, September 8-10, pp. 209–213 (2008)
8. Ogiela, L.: Syntactic Approach to Cognitive Interpretation of Medical Patterns. In: Xiong, C.-H., Liu, H., Huang, Y., Xiong, Y.L. (eds.) ICIRA 2008. LNCS (LNAI), vol. 5314, pp. 456–462. Springer, Heidelberg (2008)
9. Ogiela, L.: Cognitive Computational Intelligence in Medical Pattern Semantic Understanding. In: Guo, M., Zhao, L., Wang, L. (eds.) Fourth International Conference on Natural Computation, ICNC 2008, Jinan, Shandong, China, October 18-20, vol. 6, pp. 245–247 (2008)
10. Ogiela, L.: Innovation Approach to Cognitive Medical Image Interpretation. In: 5th International Conference on Innovations in Information Technology, Innovation 2008, Al Ain, United Arab Emirates, December 16-18, pp. 722–726 (2008)

11. Ogiela, L.: UBIAS Systems for Cognitive Interpretation and Analysis of Medical Images. Opto-Electronics Review 17(2), 166–179 (2009)
12. Ogiela, L.: Computational Intelligence in Cognitive Healthcare Information Systems. In: Bichindaritz, I., Vaidya, S., Jain, A., Jain, L.C. (eds.) Computational Intelligence in Healthcare 4. SCI, vol. 309, pp. 347–369. Springer, Heidelberg (2010)
13. Ogiela, L.: Cognitive Informatics in Automatic Pattern Understanding and Cognitive Information Systems. In: Wang, Y., Zhang, D., Kinsner, W. (eds.) Advances in Cognitive Informatics and Cognitive Computing. SCI, vol. 323, pp. 209–226. Springer, Heidelberg (2010)
14. Ogiela, L., Ogiela, M.R.: Cognitive Techniques in Visual Data Interpretation. Studies in Computational Intelligence, vol. 228. Springer, Heidelberg (2009)
15. Ogiela, L., Ogiela, M.R.: Cognitive Approach to Bio-Inspired Medical Image Understanding. In: Nagar, A.K., Thamburaj, R., Li, K., Tang, Z., Li, R. (eds.) Proceedings 2010 IEEE Fifth International Conference Bio-Inspired Computing: Theories and Applications, Liverpool, UK, Changsha, China, September 8-10, September 23-26, pp. 1010–1013 (2010)
16. Ogiela, L., Ogiela, M.R.: Cognitive Systems in Medical Data Interpretation and Image Understanding. In: Cho, Y., Kawata, S., Ko, F. (eds.) Proceedings NCM 2011 - the 7th International Conference on Networked Computing and Advanced Information Management, Hilton Hotel Gyeongju, Gyeongju, Korea, June 21-23, pp. 276–279 (2011)
17. Ogiela, L., Ogiela, M.R.: Semantic Analysis Processes in UBIAS Systems for Cognitive Data Analysis. In: You, I., Barolli, L., Tang, F., Xhafa, F. (eds.) IMIS 2011 - 2011 Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, June 30- July 2, pp. 182–185. Korean Bible University, Seoul (2011)
18. Ogiela, L., Ogiela, M.R.: Semantic Analysis Processes in Advanced Pattern Understanding Systems. In: Kim, T.-h., Adeli, H., Robles, R.J., Balitanas, M. (eds.) AST 2011. Communications in Computer and Information Science, vol. 195, pp. 26–30. Springer, Heidelberg (2011)
19. Ogiela, L., Ogiela, M.R.: Advances in Cognitive Information Systems. In: COSMOS 17. Springer, Heidelberg (2012)
20. Ogiela, M.R., Tadeusiewicz, R., Ogiela, L.: Image languages in intelligent radiological palm diagnostics. Pattern Recognition 39, 2157–2165 (2006)
21. Trzupek, M., Ogiela, M.R., Tadeusiewicz, R.: Intelligent Image Content Semantic Description for Cardiac 3D Visualisations. Engineering Applications for Artificial Intelligence 24, 1410–1414 (2011)
22. Zhong, N., Raś, Z.W., Tsumoto, S., Suzuki, E. (eds.): ISMIS 2003. LNCS (LNAI), vol. 2871. Springer, Heidelberg (2003)

Biorheological Delivery Engines in the Context of Facial Harmony

Tarik Ozkul, Haluk M. Ozkul, and John Ronczka

Abstract. What may be suggested is the plausibility that a nexus exists between Biorheology interfaces and facial harmony delivery engines ('human—machine Interoperability'). This nexus could lead to the development of practical procedures to assist patients with facial disconformities such as Down's syndrome; to those with the needs for facial plastic surgery through to patients with skin cancer. Research tends indicates that many facial parameters are likely to contribute to targeting interventions that enables the achievement of the outcome of restoring a harmonious face. Use of these parameters within the Biorheology is important since achieving a harmonious face is the primary goal of most dental and aesthetic surgical procedures. This paper suggests Informatics medicine is the cornerstone to the development of Biorheology based facial harmony 'Command—communication—control' Interfaces. A possibility is the development of an 'Informatics Medicine' Toolkit that could be used via Biorheology interfaces (e.g. Near-Field Communication (NFC) or Artificial Intelligent (AI) plasmas, foams, gels or glues) based on facial harmony KEPO's, using non-surgical interventions.

1 Introduction

A scan of a sample of available literature suggests that Orthodontists and aesthetic surgeons align to the prognosis that overall facial harmony of the patient directly

Tarik Ozkul
Computer Science and Engineering Department, American University of Sharjah, United Arab Emirates
e-mail: tozkul@aus.edu

Haluk M. Ozkul
Head, ENT Clinic, Haseki Hospital, Istanbul, Turkey
e-mail: mhalukozkul@superonline.com

John Ronczka
Adjunct to National Centre for Ports and Shipping (NCPS),
Maritime Logistics Management, Australian Maritime College, University of Tasmania,
Launceston, Tasmania, Australia
e-mail: ronczkaj@ozemail.com.au

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 253–262.
springerlink.com

aligns to the desired outcomes set for intervention or countermeasures (e.g. non-surgical; surgical operations). One outcome may be to meet the expectations of the patient by personalising the intervention and countermeasures.

Tweed [1] has mentioned the philosophy of orthodontic treatment as "the improvement of harmony in dental treatment." Merrifield [2] has studied the facial balance and aesthetics by studying profiles of patients in 1960's. Whereas, Kingsley [3] has presented a view of facial harmony improvements from that of an orthodontists. In summary, the importance of facial harmony is very well recognised by aesthetic surgeons, orthodontists, and the surgical procedures used. As such, there has been an evolution to patient 'Key expected performance and outcomes' (KEPO), such as sculpting a more harmonious face [4-6] that the patient is happy to live with.

This then leads to the use of patient focused facial parameters that may be 'key performance indicators' (KPI). The outcome sought could be to provide satisfactory results that improve the facial harmony of the patient but additionally meets the patients KEPO. Orthodontists recognised the concept of facial harmony but not in the context of other interventions and countermeasures as they related to Biorheological KEPO's. This study concentrates on the Facial harmony KEPO's and KPI's aligned to Biorheological based interventions and countermeasures. Within the Informatics medicine context, what are provided are Biorheological Facial harmony KEPO's with implicit intervention, countermeasures and relationships that might be distilled by a causality nomenclature. The causality nomenclature puts forward are plausible Biorheology based facial harmony entangled single–to–multiple chain 'causality logic gates' of 'knowledge—information—learning domains' (KILD's) Biorheology Semantic—Semiotic interfaces.

2 Key Terms

In the journey within the context of Biorheological Facial harmony KEPO's the initially step, is to have an understanding of the key words:

- **Medicine:** "the scientific study or practice of diagnosing, treating, and preventing disease or disorders of the body or mind of a person or animal" [7]; and
- **Informatics (medical context):** "the collection, classification, storage, retrieval, and dissemination of recorded knowledge" [8].
- **Biorheology:** apply to the rheology of living systems or materials or biological matter in terms of plasticity and non-Newtonian fluids [9]
- **Facial harmony:** is most commonly referred to be "beauty" [10].
- **Interfaces:** Integration—adaptation between entangled single–to–multiple chain (ESTMC)—KILD's that may have the capability for multiplexing between entities and events (e.g. Near-Field Communication (NFC) or Artificial Intelligent (AI) plasmas, foams, gels or glues).

Initially a nexus is suggested between practice and knowledge within 'Informatics medicine' and Biorheological based interfaces for external body use. Secondly a nexus exists between Biorheology based facial harmonies parameters. This could be a practical attempt to assist patients with Down's syndrome; to those with the needs for facial plastic surgery through to patients with skin cancer [11].

3 The Concept of Facial Harmony

Assessment of facial harmony is often necessary for medical and dental operations. The ultimate goal of any aesthetic or dental operation is to achieve the best possible harmonious face for the patient. There are countless parameters describing facial features. The aim of most surgical operations is to bring the facial parameters to normal, acceptable range while achieving the best possible harmony for the patient.

An assessment of facial harmony tends to be commonly referred to as "beauty." Some argue that concept of beauty varies within various ethnicities; and other researchers suggest that facial proportions are time dependent [10]. Indeed some studies indicated that, the perception of "good looking" do seem to vary regionally [12]. Such an interpretation was confirmed by another study conducted by Riedel [13], who concluded that, although experts do not seem to agree on what is "best," they all agree on what is "bad."

Even though what is "best looking" is disputable as suggested by the Riedel study; there is no disagreement about what should be accepted as "normal." Artists and scientists are studying facial proportions, which make a "normal face," since 15th century. Leonardo da Vinci studied these proportions intensively and established the ground rules for anatomical proportions which is very similar to what we use today [14].

Medical professionals devised dozens of parameters to describe crucial facial angles and proportions of "normal" faces. These angles and proportions are expected to be within certain range for the parameters to be accepted as "normal." These parameters have been studied extensively, forms part of a number the medical curriculum, and might be suggested to be for a Male Female subjects [15-17]:

- **Nasofrontal angle (BAG):** should be between 115-130 degrees. Within this range, the more obtuse results are preferred for females and more acute preferred for males.
- **Nasofacial angle (BAF):** ideally should be between 30-40 degrees.
- **Nasolabial angle (TDE):** should be between 95-105 degrees for male gender.
- **Nasomental angle (ABE):** ideally should be between 120-132 degrees.
- **Nasal projection (NT/AT):** is a ratio and ideally should be around 0.55-0.60.
- **Rule of the third (ROTT):** Face should be divided, into three equal portions vertically marked by tip of the chin, tip of the nose, eyebrows and top of the forehead. Ideally, this ratio should be equal to 1.

- **Rule of the fifth (ROTF):** Face should be divided into five equal vertical portions marked by the tip of the ears, outer edge of the eyes, and outer edge of the nose.
- **Base of the nose (BOTN):** Base of the nose should be confined, to an equilateral triangle. Ideally, this ratio should be 1 for equilateral triangle.
- **Mentocervical angle (HFG):** is defined as the angle defined by glabella–pogonion line intersecting with menton—cervical point line.

4 The Concept of Biorheological KEPO's

Current research effort has focused on the nexus of Biorheological KEPO's as ESTMC—KILD's using 'Biorheology causality logic gates' (B–COR gates) [18]. These ESTMC's might exist in multiple operating dimensions as *Möbius strips* hybrid 'B–COR gates'. The B–COR tendency may be to act as a biological 'knowledge—information—learning delivery engines' (KILDEE's). Conceivably, the fusion of human—machine *Markov chain* Biorheology interfaces (Figure 1) might be the next 'Biorheological KEPO's based facial harmony evolution [19].

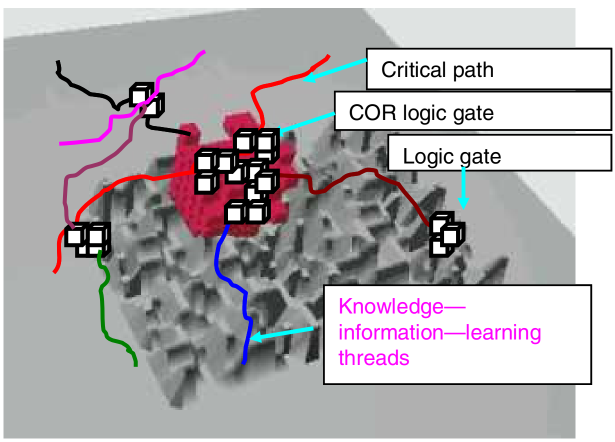


Fig. 1 Entangled–single–multiple chains—*Biorheology interface* [18] [19]

The B–COR KILD's paradigm undergoes ESTMC's convergence—merging:

- C^5M ('Command—communication—control' with *'Management—causalities—consequences'*) [18] [20];
- SIAN (Synergy, integration (Interoperability), assimilation and narratives) [18] [20].

- BRI (Biorheological informatics: mathematical study of biological interfaces in the life sciences (e.g. biology, genome) [21]. Semantic—Semiotic neutral nets may be used for Biorheology interface hybrid access ('devices—entities—events—processes') to overcome impairments? Within the dynamics of deformation and flow of matter (Rheology), biological logic gates (B–gates) bio-delivery engines might be the reality for 'human—machine Interoperability' communication–conversation. Like in human communications misunderstanding (random radicals) are possible within the supporting spectrums and bandwidths as they exist over multiple hybrid access interfaces.

5 The Interface Problems

Problems that need to be addressed may be summaries as (Table 1):

- **Facial harmony** parameters to describe crucial facial angles and proportions of "normal" faces.
- **Human impairments** overcome/minimised.
- **Enhancements** via B–COR KILD—ESTMC delivery engines.
- **Interoperability** communication–companion Human—machines using Informatics medicine.
- **Biorheology interfaces** Human—machine (AI) plasmas–foams–gels–glues.
- **Low rejection** rates of host Biomaterial in facial harmony intervention.
- **Packets:** The journey of a Human—machine packets within 'Biorheological KEPO's based facial harmony' interventions and countermeasures (Fig 2) of KILD's C^5M, SIAN and BRI based ESTMC—KILD's?

Packets within Human—machine 'Biorheological KEPO's' based facial harmony' might have messages or fragments of messages that are individually routed between human—machine nodes. With no previously established communication path, interface packets are routed to their destination through the most expedient route [10] due to the way packet switching versus Message switching occurs with:

- **Human impairments** overcome/minimised: facial harmony due to cancer.
- Long message has more chance to **incur error** than short packet and are **not suitable** for interactive application due to waiting delay on other messages.
- Delay for one message is generally longer than the **total delay** of packets the message is divided into [22].

Semantic Interfaces:

- **<u>effective</u>** means of communication which has proven itself in practice?
- **pre-established** human-defined ontology rely upon?
- **communication** can only proceed by the interpretations of behaviours, common behaviours?

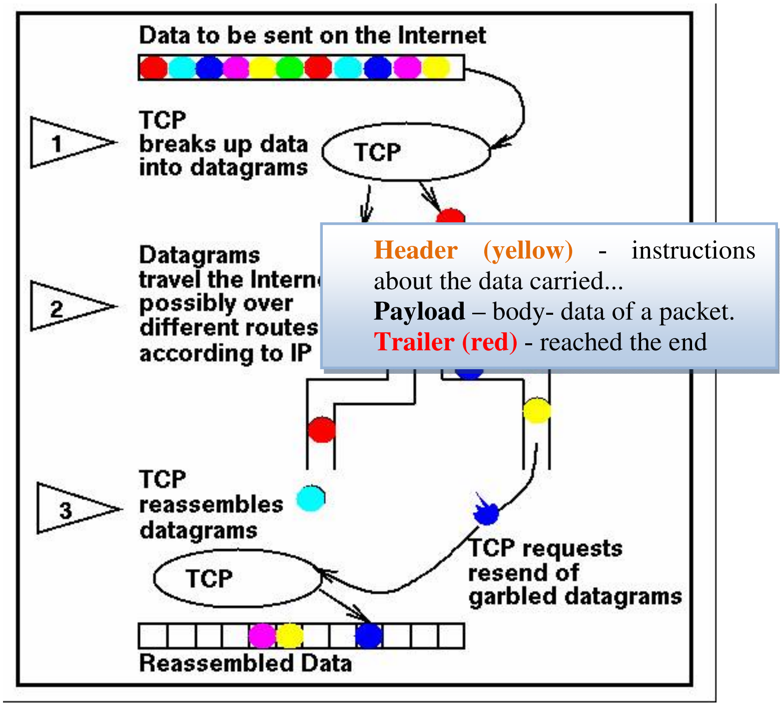


Fig. 2 Packet switching versus Message switching [22]

Semiotic Interfaces:

- **Metaphor** such as a new idea is created from the fusion of the two original ideas, or our understanding of the first idea.
- **Mental models that** are cognitive based for human-computer interaction.
- **User System usage comprises *D1***: concepts the user knows and uses; ***D2***: concepts used only occasionally and not initially known; ***D3***: the user's model of the system (i.e. the set of concepts the user *thinks* exists in the system); ***D4***: the actual system. [23] [24].

Table 1 ‘Interface problems - Biorheology

Interface problems	Biorheological		
	Plasticity	Non-Newtonian fluids	Biological
Facial harmony parameters	Y	Y	Y
Human impairments	Y	Y	Y
Enhancements	Y	Y	Y
Interoperability	Y	Y	Y
Semantic and Semiotic	Y	Y	Y
Low rejection rates	Y	Y	Y
Packets:	Y	Y	Y

6 Domains

Semantic based mapping process uses Themes' (Table 1; Fig. 1) to assist in their discovery of the Domains. The ESTMC's might exist in multiple operating KILD dimensions, yet reinforces a nexus to use 'Medical Informatics Domains' (Table 2) to demonstrate a paradigm shift to 'Informatics Medicine'.

Table 2 'Interface problems'– 'Informatics Medicine' – Domains

Interface problems	Medical Informatics Activity	Semantic Domain (Entity)
Facial harmony parameters	Medical Sociology	Medicine
Human impairments	Biostatistics; Computational medicine	Informatics
Enhancements	Relationship between the Patient and Doctor–medical practitioner	Medicine
Interoperability	Health information; Pathology informatics; Nursing informatics	Informatics
Semantic and Semiotic	Bioinformatics (computational genomics, computational genetics);	Informatics
Low rejection rates	Medical and Health Administration	Informatics
Packets:	Medical imaging informatics	Informatics

7 Causality

In causality evidence based nexus development, what are highlighted are evidence based relationships and processes that exist between the entities and events. Within the controls environment of a Biorheological human–machine interface (Table 3) infers is that causality based evidence drives the nexus [18].

Table 3 'Interface Problems– 'Informatics Medicine' – Causality

Interface problems	Symptoms	Immediate cause	Remote Cause
Facial harmony parameters		YES	
Human impairments	YES		
Enhancements		YES	
Interoperability			YES
Semantic and Semiotic			YES
Low rejection rates			YES
Packets:			YES

Causality: Symptoms: effects—factors specific to an event or occurrence. Immediate cause explains why the event or occurrence has occurred. Remote cause may explain why the event or occurrence has occurred [25] [26]. **Relationship** means between the Patient and Doctor–medical practitioner.

8 Conclusions

Entanglements of entities and events within 'Informatics Medicine' may involve ESTMC—KILD's based parameters for Facial Harmony, such as:

- **Complexities** in the supporting knowledge—information—learning domains;
- **Multiple** operating dimensions, personalised yet accessible 24/7 to advice;
- '**Biorheology** causality logic gates' (B–CORG);
- **Biological–*machine*** KILDEE's;
- **Unique** trigger events, information—knowledge that has Biological—*machine* temporal meshing; and
- **Fusion** of human—machine via Biorheology interfaces evolution.

In terms of Semantics to Semiotic based interfaces, this suggests a hybrid approach of ESTMC—KILD's with:

- **Effective** means of communication which has proven itself in practice;
- **Pre-established** human—defined ontology that is adaptable and reliable;
- **Communication** can only proceed by the interpretations of behaviours, common behaviours;
- **Metaphor** such as a new idea is created from the fusion of the two original ideas, or our understanding of the first idea;
- **Mental models** that are cognitive based for human-computer interaction; and
- **User System** focused on patient KEPO comprises **concepts** using both **actual** and a**ugmented** reality based interfaces that adapt to the user's needs [27].

The causality nomenclature put forward plausible Biorheology based facial harmony using a number of entangled single–to–multiple chain 'causality logic gates'. Future work will focus on developing a functioning 'Informatics Medicine' Biorheology based facial harmony interfaces focused on patient KEPO's, using non-surgical interventions that can be personalized and administered at home.

References

[1] Berneburg, M., Dietz, K., Niederle, C., Goz, G.: Changes in esthetic standards since 1940. American Journal of Orthodontics and Dentofacial Orthopedics 137(4), 450.e1-450.e9 (2010)

[2] Tweed, C.H.: Indications for the extraction of teeth in orthodontic procedure. American Journal of Orthodontics 30, 405–428 (1944)

[3] Merrifield, L.L.: The profile line as an aid in critically evaluating facial esthetics. American Journal of Orthodontics 52(11), 804–822 (1966)

[4] Kingsley, N.A.: Treatise on Oral Deformities as a Branch of Mechanical Surgery, pp. 465–509. Applegate, New York (1888)

[5] Klontz, H.A.: Facial balance and harmony: an attainable objective for the patient with a high mandibularplane angle. American Journal of Orthodontics and Dentofacial Orthopedics 114(2), 176–188 (1998)

[6] Brons, R.: Facial harmony: standards for orthognathic surgery and orthodontics. Quinteessence (1998)

[7] NICM, "Definition of complementary medicine", National Institute of Complementary Medicine (NICM) (January 2009), http://www.nicm.edu.au/content/view/31/35/

[8] Merriam-Webster, "Informatics", Cited a Reference, Merriam-Webster's Medical Dictionary, Merriam-Webster, Inc., 1 (2007), http://dictionary.reference.com/browse/informatics

[9] ISB, Biorheology., International Society of Biorheology, A symposium on Rheological Problems in Biorheology, Lund Sweden (2005), http://www.coe.ou.edu/isb/history.html, http://www.rheology.org

[10] Martin, J.G.: Racial ethnocentricisim and judgment of beauty. Journal of Psychology 63, 59–63 (1964)

[11] Boyle, M.B., Gikandi, W.W., Kassim, Z.G., Salih, S.: Colloidal and Surface Phenomena Aspects of Sunscreen Lotion. Department of Chemical Engineering, University at Buffalo, SUNY, CE457/527 Project Report (April 8, 2002), http://www.eng.buffalo.edu/Courses/ce457_527/ce457_pro/g10_doc.html

[12] New Mexico State University, *Faces,* Department Psychology Website. Las Cruses, NM (April 2011), http://www.cs.nmsu.edu/ipivkina/Spring05cs579/StudentPres/Faces.pdfS

[13] Riedel, R.A.: Esthetics and its relation to orthodontic therapy. Angle Orthodontist 20, 168–178 (1950)

[14] Naini, F.B., Moss, J.P., Gill, D.S.: The enigma of facial beauty: esthetics, proportions, deformity, and controversy. American Journal of Orthodontics and Dentofacial Orthopedics 130, 277–282 (2006)

[15] Powell, N., Humpreys, B.: Proportions of the Aesthetic Face. Thieme Startton, New York (1984)

[16] Papel, I.D.: Facial Plastic and Reconstructive Surgery. In: A Problem-Solving Manual, Thieme, New York (2002)

[17] Papel, I.D.: Facial Analysis and Nasal Aesthetics. In: Aesthetic Plastic Surgery. Springer, New York (2002)

[18] Ronczka, J.: Wisdom Open-system Semantic Identification (WOSSI) Mapping of Causality Logic Gates. In: ITA 2009, pp. 1–10. Glyndŵr University North East Wales, UK (2009), http://www.ita09.org

[19] FEDEE "Relocating and integrating business operations", Federation of European Employers (EDEE), The Strand, London, p. 1 (2002), http://www.fedee.com/natlaw.html

[20] Brand, S.: GBN Book Club Reviews–War and anti–war, p. 1 (1993), http://www.gbn.org/BookClub/War.html (accessed March 2000)

[21] Fu-berlin: Introduction to Dahlem Research School (DRS), p. 1. Freie Universität, Berlin (2010), http://www.fu-berlin.de/sites/en/promovieren/drs/index.html

[22] ESCI, "Network Protocols" (2002), http://escience.anu.edu.au/lecture/ivr/networkProtocols/printNotes.en.html

[23] Croft, "Aluminus", Host webpage (2003), http://alumnus.caltech.edu/~croft/research/agent/role/

[24] Avery, J., Yearwood, J.: Supporting Evolving Ontologies by Capturing the Semantics of Change, p. 3350. University of Ballarat, Victoria (2004), http://www.ballarat.edu.au

[25] Hobbs, M.L.: The Value of Urban Design, Ministry for the Environment NZ, p. 1 (2005), http://www.mfe.govt.nz/publications/urban/value-urban-design-summary-jun05/html/page12.html

[26] Bellinger, G., Castro, G., Mills, A.: Data. Information, Knowledge, and Wisdom. In: The International Society for the Systems Sciences (ISSS), p. 1 (2004), http://www.systems-thinking.org/dikw/dikw.html

[27] Ronczka, J.P.: Coalescence Theory application to rheology of entangled, single and multiple chains to suggest Biorheology logic gates. In: 5th Pacific Rim Conference on Rheology (PRCR-5), Sapporo Japan, pp. 1–2 (2010), http://prcr2010.com

Celebrities as Marketing EnhancerCase Analysis of the Alternative Food Movement and "Eco-Chic" Lifestyle Advocacy

Shinichiro Terasaki and Shin'ya Nagasawa

Abstract. Globalisation, industrialisationand neo-liberalism have caused social inequality and environmental degradation, while increasing numbers of celebrities, particularly in entertainment industry make use of this trend to maximise their influence over the public through environmental advocacies such as "alternative" food geographies and "eco-chic" lifestyle. The purpose of this paper is to examine how far celebrities' environmental advocacies work as a branding method of celebrities themselves.Following conceptual background discussions such as defining celebrity, types of celebrity and what "eco-chic" lifestyle means, critical assessment of the above issues takes place with a good array of empirical examples.The results indicates that celebrity advocacy is fundamentally a sort of star-driven "outside strategy" of social protest that popularises issues, enlarging the visibilities of celebrities themselves. This research has presented a new marketing trend that involves political ecology as a means of self-branding of celebrities who could be readily regarded as"products" in capitalism.

Keywords: branding, celebrity politics, political ecology, alternative food movement, environmental degradation, social inequality, marketing trend.

1 Introduction

Many public debates currently take place through television, internet, front page news, and "screens" (DeLuca and Peeples, 2002: 133). In such discussions, the presence of "the rich and famous" (Bryant, 2007), generally referred to as celebrities, continues to expand, involving many types of marketing channels. Contemporary mass media are full of images and information about celebrities. Subsequently, the public are exposed to high profiles, eccentric qualities, and attractive images of the famous (Giles, 2000; McCracken, 1989).

We cannot underestimate powerful images of celebrities. DeLuca and Peeples (ibid.) note that "emotions over rationality, speed over reflection, distraction over deliberation, slogans over arguments, the glance over the gaze, appearance over

Shinichiro Terasaki · Shin'ya Nagasawa
Graduate School of Commerce, Waseda University

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 263–272.
springerlink.com

truth, the present over the past." Celebrities derive from many fields, ranging from entertainment to sports, cooking, business and politics (Choi *et al.*, 2005: 85). Since the mass media encourage celebrities across nation boundaries, thereby building awareness and repute all over the world, celebrities are believed to conquer cultural barriers in worldwide marketing communications (Erdogan, 1999; Kaikati, 1987).

Deploying their "well-knowness," celebrities have begun to participate in various public discussions. The conventional scholarly research on the role of celebrities has focused on the accessibility to the news media, and the power to attract public attention to such issues (Meyer and Gamson, 1995; West and Orman, 2003). In contrast, contemporary research should also encompass the development of the internet, and the growing fragmentation of the mass media, both of which make the interaction between celebrities and the public intensive. Also, celebrities are capable of communicating with socio-economic elites, partly due to the blurred lines between celebrities and politicians (e.g. Ronald Reagan, Jesse Ventura, and Arnold Schwarzenegger etc.).

The objective of this paper is to examine how far celebrities' environmental advocacies such as "alternative" food geographies and "eco-chic" lifestyle work as a self-branding method. Following conceptual background discussion such as defining celebrity, types of celebrity and what "alternative" food geography means, critical assessment of the above issues takes place with a good array of empirical examples.

2 What Makes Celebrity?

Nowadays, celebrities play a central role in driving messages home via entertainment channels (Thrall *et al.*, 2008: 377). The word, celebrity can be a complex and sometimes controversial term. In some countries, celebrities do not necessarily satisfy both "rich and famous." For instance, in Japanese popular TV programme, *SekaiBaribariValue* (see wiki, 2008), infamous, but rich people were often referred as celebrities, and interviewed as samples of celebrities. Owing to the ambiguous use of the word, some of the public may be confused about the definition. To avoid such controversy, I analysed it from the angle of academic perspective.

Daniel Boorstin (Boorstin, 1961: 57; see also Lahr, 1978; Monaco, 1978; Schickel, 1985) describes celebrity as a person "known for his [or her] well-knownness." This "well-knowness" is attributed to their "highly dynamic, likeable and attractive" (Atkin and Block, 1983) personalities. Rein *et al.* (1997: 15; cited in Turner, 2004: 32) add that the commercial value of celebrities is as follows; "a celebrity is a person whose name has attention-getting, interested-riveting, and profit-generating value." This implies the correlation between "well-knowness" and commercial value of celebrities.

Some scholars portray celebrities in somewhat negative terms. Rojek (2001: 37–8) writes that celebrities emerged as part of a method to control the new urban masses by showing highly visible images. Thus, celebrity is intentionally manufactured, and celebrity figures are not necessarily the "deserving" (Meyer, 1995: 183). In addition, the social influence of the famous does not derive from formal, institutional power (Alberoni, 1972). In other words, celebrity is a "powerless elite" who reap the privileged social status not from institutionally-based social power, but from public attention (ibid.).

Although celebrities can be defined as "powerless elites" especially from the perspective of conventional view, some of the "rich and famous" have access to global political leaders. As we reviewed, many celebrities chose to become politicians as their secondary careers, blurring the lines between show business and politics. For Bono-the front man of U2-, celebrity is a form of "currency" (Singer, 2002) that enables not only communication with publicity and media, but also with world political leaders.

CEOs at trans-national corporations (TNCs) are also accessible to global politics. For instance, many global CEOs annually attend the World Economic Forum in addition to the political leaders and executives of international NGOs such as Oxfam, WWF and Friends of the Earth. CEOs can be less visible compared to celebrities in the entertainment industry, whereas they are powerful enough to influence the world political economy (e.g. credit crunch).

In this essay, I chose a good array of celebrities from the entertainment industry. Firstly, "alternative" food movements are discussed followed by "eco-chic" lifestyle of celebrities.

3 Alternative Food Movements

3.1 What Does Alternative Food Geography Mean?

Goodman and Goodman (2007: 2) describes "alternative" food networks (AFNs) as follows; "new and rapidly mainstreaming spaces in the food economy defined by—among other things—the explosion of organic, fair trade, and local, quality, and premium specialty foods. These are distinct from standardised industrial foods such as MacDonald's hamburgers and those of major retailers (ibid: 5), both of which aim for operational efficiency. Owing to their promotion, "alternative" movements have become prevalent in the North, and nowadays increasingly so in the South as well (Bryant *et al.*, 2008: 2).

Through this process, relatively wealthy consumers and often poor producers, particularly in the South are presumed to have "united," via bonds forged on account of fairness, ecological sustainability and product quality (ibid.: 3). Affluent consumers are sometimes referred to as 'reflexive' consumers who profess concern about the environmental and social impacts of their procurement (see DuPuis, 2000; Guthman, 2003). Though they tend to be well-educated, this does not imply that they have doctorates in nutrition or ethics or agronomy (Freidberg, 2004: 520).

AFN foods are delivered via new channels such as charity shops, food cooperatives, farmers' markets, community-supported agriculture or box schemes (Goodman and Goodman, 2007: 5). Nonetheless, the rapid growth of AFN market encouraged large-scale retailers to form supply chain with AFN producers, imposing strict cost-price disciplines (ibid.).

3.2 Quality Food Movements and Celebrity Chefs

Advocacy of quality foods such as organic commodities, those easily "distinguished" by taste and origin has characterised the mainstream social movement with respect to food and agriculture in the post-war period (Goodman and Goodman, 2007: 3; Banerjee *et al.*, 1995: 23). "Organic" emphasises the health aspects of environmentalism or the pros of "natural" products and ingredients (Banerjee *et al.*, 1995: 23). The modern food politics reflects the highly induvidualised purchasing decisions of "picky" consumers (Guthman, 2008: 1175).

This trend is generally considered as active opposition to standardised food products (Guthman, 2003: 46). Certainly, many still buy fast foods without any consideration for their health or communities involved in the production process. In contrast, people, especially the affluent and educated, are wary of the safety and nutritional value of those mass-produced "placeless and faceless" foods as a consequence of periodic food anxieties (e.g. BSE, foot and mouth disease, and avian influenza) (Goodman and Goodman, 2007: 6).

Growing concern for quality foods has been associated with increasing consumer knowledge about bulk-produced foods. However, Freidberg (2004: 520) notes that most information that renders the meaning and value of organic and fair trade labels derives from the popular media, not from formal education. Apart from this discussion, whether the information has a positive influence or not (e.g. the latest fad diet), celebrities can be considered as one of the major sources of information in "alternative" food movements.

In the contemporary UK, the quality food movement is flourishing, particularly at the level of vicarious satisfaction (Freidberg, 2004: 519). Celebrity chefs such as Jamie Oliver, Gordon Ramsey, and Hugh Fearnley-Whittingstall boost this movement largely through cooking shows (e.g. Food Fight). In this essay, I mainly focus on Jamie Oliver, a chef and food politics activist in the UK.

Oliver is not only a chef, but also a business person who exploits the social and aesthetic value of cooking as a tool to expand his socio-economic power in the UK society. His self-promotion is conducted via diverse media such as TV programmes, blogs, podcasts, and books, all of which are integrated in his official website (see JamieOliver.com, 2009). Through the website, consumers can directly purchase many services and products apropos to his "alternative" food activism. Since the number of website visitors generally depends on the celebrity's popularity, it is indispensable for him to maximise the number of his fans.

In this sense, his food activism to ban the junk in schools (see Channel4.com, 2009) contributes to reap his potential fans in younger generation, and the enhancement of his influence in food politics. According to McKinley *et al.* (2005: 549), British children often aspire to be like actors, music and sports stars because they are perceived to 'have lovely wee skinny bodies'.

Normally, well-shaped body is believed to be associated with healthy homemade cooking. "Ministry of Food", Oliver's new TV series that aim to let the public cook again, can help Oliver "pitch arguments for the rethinking of food culture" (Freidberg, 2004: 519). Although many viewers still look for "ready meal" at supermarkets (ibid.), Oliver will benefit from the TV programme in a sense that it induces “effective monetary demand for quality products” (Goodman and Goodman, 2007: 6). As Goodman and Goodman (ibid.: 12) imply, quality foods have “farming in nature’s image.” Thus, Oliver can also get hold of environmentally-friendly image throughout his activities.

Overall, the synergy of the above activities assists Oliver in acquiring socio-economic power. Nevertheless, there is a major potential risk of damaging his reputation; disgraceful attitudes to the customers as exhibited by fellow chef, Gordon Ramsey. In April, 2009, many mass media revealed that customers at four of Gordon Ramsey’s restaurants were likely to be served ready meals prepared outside of his restaurants (see Mailonline, 2009; Telegraph.co.uk, 2009). The food delivered to the restaurants have been reheated in each restaurant, and sold for up to six times compared to the initial costs, generating a “healthy” profit to buy his new Ferrari.

Akin to politicians, keeping a “healthy” image is vital for celebrity chefs to gain sustainable competitive advantages over rivals. Even though celebrity chefs can produce business models to socio-economic success, actual business operation is done by many “faceless” employees; it is reputation which enables them to realise their political, economic and environmental interests.

4 “Eco-Chic” Lifestyle Movements

4.1 The Emergence of “Going Green”

In current U.S political and social life, celebrity participation in social movements has become an increasingly common feature (Meyer, 1995: 200). Americans tend to acquire such information through casual news outlets (e.g. *Entertainment Tonight* or the *Daily Show*), rather than conventional news media (Baum, 2003; Kernell and Baum, 1999; Prior, 2005).

Irrespective of the range of media forms, “Going Green” is a constant celebrity topic in today’s media (Corkran*et al.*, 2008: 850). This emerging “eco-chic” lifestyle of celebrities has its roots in the growing awareness of environmental problems; typically global warming. The famous employ their popularity with the public to indicate their thoughts on such problems, and ways to minimise the effects on the environment (ibid.: 854). Brockington (2008: 560) names such “eco-chic” celebrities especially those who have won fame through the environmental-friendly promotions as celebrity conservationists.

In this chapter, we will critically explore this emerging “eco-chic” lifestyle movement by using three main examples; Leonardo DiCaprio, Live Earth, and the relationship between celebrities and global civil society.

4.2 Examples of “Eco-Chic” Lifestyle Movements, and Analysis

Due to the prevalence of internet technology, modern citizens easily search for celebrity charity news and information. “Look to the Stars,” for instance, provides and promotes such information, covering over 1770 celebrities from diverse industries (see Look to the Stars, 2009). Such a supportive environment contributes to enlarging celebrities' visibility to their audience.

Now, we look at the actual charitable activities of celebrities. Brad Pitt, for example, narrates an ongoing PBS series called *e² design* that is about sustainable architecture (PBS, 2008), whilst Keanu Reeves and Alanis Morrisette narrate a global warming documentary named the *Great Warming* (see thegreatwarming.com, no date given).

Perhaps, one of the most renowned advocates of anti-global warming is Leonardo DiCaprio. To date, the Leonardo DiCaprio foundation produced three movies related to environmental problems; *Global Warming* (2003), *Water Planet* (2005), and *The 11th Hour*. His global warming documentary, *the 11th Hour*, is especially well-known, winning the Jules Verne Audience Choice Award in 2007. In the movie, he featured a variety of authorities such as David Suzuki, Joseph Tainter, and Stephen Hawking (see Warner Brothers, 2007), whose support enhances the credibility of his advocacy.

The key feature of his "eco-chic" lifestyle is the exploitation of his ability to narrate environmental issues as an actor. As a result, the messages become somewhat emotional rather than rational. Alison Anderson (1991: 473) describes the method of such environmental advocacies as 'show-business type' approach. Since laypeople lack technical knowledge about environmental problems, they cannot critically evaluate the contents of such documentaries. In the long-term, the opinions of celebrities almost equate with public opinions about environmental problems. Eventually, this may affect environmental policy of the governments.

Some politicians make use of this "eco-chic" lifestyle movement to maximise their influence on the public. For instance, Al Gore, 45th vice president of the US, and over one hundred fellow celebrities attracted public attention to climate change via *Live Earth*, a 24-hour concert in 2007. According to the official website (Live Earth, no date given), approximately 2 billion people worldwide participated in the concerts through TV, radio, and internet.

There are largely two differences between individual charitable activities and collaborative "eco" advocacies of celebrities. Even though such concerts are less personalised and dramatised, the famous who participate in "eco-chic" concerts can avoid direct criticism. Second, each participant may reap a positive halo effect from the other participants. In such a big concert, some audiences can be forced to listen to the "eco-chic" messages from some specific artists disliked by them. However, their reluctance can be buffered by the rest of the artists. This expected halo effect will contribute to each artist's self-promotion to his/her socio-economic interests.

Overall, many celebrities nowadays strive to advance their political, economic and environmental interests through "eco-chic" advocacies. Nonetheless, such activism seems to assume neo-liberal taste. According to Thrall *et al.* (2008: 367-78), the celebrities listed on the *Forbes 100* are roughly twice more likely to engage in a wider range of issues and groups regarding celebrity advocacy than the rest of less visible celebrities.

Thrall *et al.* (ibid.) explain part of the reason that the most famous should be the most attractive to interest groups-it makes sense that they have more chance to get involved. Except for some established NGOs such as NRDC, Greenpeace, and Conservation International, all of which need not rely on the famous, smaller groups such as the Whaleman Foundation and Yele Haiti have greater difficulties making the news devoid of celebrity support. Accordingly, Yele Haiti, for instance, has received support from eight celebrities, including Susan Sarandon who explained her own political activism as follows; “If my privacy is going to be invaded and I’m going to be treated as a commodity, I might as well take advantage of it” (quoted in Brownstein 1990: 11). This suggests that social activism can be one method of career development for the “rich and famous”.

In addition, successful celebrities have more personal disposable wealth than the less famous. Kaufman and Wolf (2007: 156) alert that “philanthropy can be dehumanized and can become just a competition to see who gets on the media’s top philanthropist lists.” In the growing celebritisation of development (Goodman, no date given: 7),only the rich can benefit from donations from the angle of their personal socio-political interests. ArjunAppadurai (2000: 2; cited in ibid.: 2) asks a pregnant question: “Can the media ever be turned to the interests of the poor?.”

5 Conclusion

Celebrity advocacy is fundamentally a sort of star-driven “outside strategy” of social protest that popularises issues (Thrall *et al*., 2008: 363). Vivanco (2004: 10) also points out their simplification of social context. This simplified social context is translated into the casual news outlets most people depend on to get the information of celebrity advocacy. Through a series of information cycles, celebrities succeed in maximising their visibility widely considered as a key to career gain. Notwithstanding, this increasing visibility complicates celebrities’ relationship with the entertainment industry and other diverse audiences such as the mass media, NGOs, and a broader public (Meyer, 1995: 201). Some celebrities (e.g. Gordon Ramsey) suffered career damaging criticism from mass audiences because of the increased visibility. Thus, whether they can occupy visibly privileged socio-economic status largely relies on the individual strategy of social movements they engage in.

As conventional wisdom implies that the least famous should matter least (Thrall *et al*., 2008: 371), celebrity advocacy will become increasingly competitive as long as those involved feel sufficient merits from such activities.

References

1. Anderson, A.: Source Strategies and the Communication of Environmental Affairs. Media Culture & Society 13, 459–476 (1991)
2. Alberoni, F.: The Powerless “Elite”: Theory and Sociological Research on the Phenomenon of the Stars. In: McQuail, D. (ed.) Sociology of Mass Communications, pp. 75–98. Penguin, Harmondsworth (1972)
3. Appadurai, A.: Grassroots globalization and the research imagination. Public Culture 12(1), 1–19 (2000)

4. Atkin, C., Block, M.: Effectiveness of celebrity endorsers. Journal of Advertising Research 23, 57–61 (1983)
5. Banerjee, S., Gulas, S.S., Lyer, E.: Shades of Green: A multidimensional Analysis of Environmental Advertising. Journal of Advertising 24(2), 21–31 (1995)
6. Baum, M.: Soft News Goes to War. Princeton University Press, Princeton (2003)
7. Boorstin, D.: The Image: A Guide to Pseudo-Events in America. Harper and Row, New York (1961)
8. Brockington, D.: Powerful environmentalism: conservation, celebrity and capitalism. Media, Culture & Society 30(4), 551–568 (2008)
9. Brownstein, R.: The Power and the Glitter: The Hollywood-Washington Connection. Pantheon, New York (1990)
10. Bryant, R.: For richer, for poorer: Towards a political ecology of the rich and famous. Paper presented at King's College London Departmental Seminar Series (November 2007)
11. Bryant, R., Goodman, M., Redclift, M.: Spaces of intention as exclusionary practice: Exploring ethical limits to "alternative" sustainable development. Environment, Politics and Development Working Paper Series 4, 1–22 (2008)
12. Channel4.com. Jamie's School Dinners [internet] (2009), http://www.channel4.com/life/microsites/J/jamies_school_dinners/campaign/index.html (accessed December 12, 2011)
13. Choi, S.M., Lee, W., Kim, H.: Lessons from the rich and famous: A cross-cultural comparison of celebrity endorsement in advertising. Journal of Advertising 34(2), 85–98 (2005)
14. Corkran, G., Guthrie, J., Heavner, B.: Leadership and Global Warming. In: Oglesby, R.A., Adams, M.G. (eds.) Business Research Yearbook: Global Business Perspectives, vol. XV. IABD, Beltsville (2008)
15. Deluca, K.M., Peeples, J.: From Public Sphere to Public Screen: democracy, activism, and the "violence" of Seattle. Critical Studies in Media Communication 19(2), 125–151 (2002)
16. DuPuis, E.M.: Not in my body: rBGH and the rise of organic milk. Agriculture and Human Values 17, 285–295 (2000)
17. Erdogan, B.Z.: Celebrity Endorsement: A Literature Review. Journal of Marketing Management 15(4), 291–314 (1999)
18. Freidberg, S.: The ethical complex of corporate food power. Environment and Planning D: Society and Space 22, 513–531 (2004)
19. Giles, D.C.: Illusions of Immortality: A Psychology of Fame and Celebrity. MacMillan, London (2000)
20. Goodman, D., Goodman, M.: Alternative Food Networks. In: Kitchin, R., Thrift, N. (eds.) International Encyclopaedia of Human Geography. Elsevier, Oxford (2007)
21. Goodman, M.: The Mirror of consumption: Celebritisation. developmental consumption and the shifting cultural politics of fair trade
22. Guthman, J.: Fast food/organic food: reflexive tastes and the making of "yuppie chow". Social and Cultural Geography 4, 45–58 (2003)
23. Guthman, J.: Thepolyanyian way?: Voluntary food labels and neoliberal governance. Antipode 39, 456–478 (2007)
24. Guthman, J.: Neoliberalism and the making of food politics in California. Geoforum 39, 1171–1183 (2008)
25. JamieOliver.com (2009), `http://www.jamieoliver.com` (accessed December 12, 2011)

26. Kaikati, J.G.: Celebrity Advertising: A Review and Synthesis. International Journal of Advertising 6(2), 93–106 (1987)
27. Kaufman, A., Wolf, Y.: Celebrity CEOs and Privacy Issues. In: Larsen, K.R., Voronovich, Z.A. (eds.) Convenient or Invasive: The Information Age. Ethica Publishing, Boulder (2007)
28. Kernell, S., Baum, M.A.: Has Cable Ended the Golden Age of Television? American Political Science Review 93(1), 99–114 (1999)
29. Lahr, J.: Notes on Fame. Harper's, 77–80 (January 1978)
30. Lang, T.: Going public: food campaigns during the 1980s and early 1990s. In: Smith, D.F. (ed.) Nutrition in Britain: Science, Scientists and Politics in the Twentieth Century, pp. 238–260. Routledge, New York (1997)
31. Lester, L.: Lost in the wilderness? Celebrity, protest and the news. Journalism Studies 7(6), 908–921 (2006)
32. Live Earth. Who We Are [Internet], `http://liveearth.org/en/liveearth` (accessed December 12, 2012)
33. Look to the Stars. Charitable Celebrities [Internet] (2009), `http://www.looktothestars.org/celebrity` (accessed December 12, 2012)
34. MailOnline. That's a lot of ready meals Gordon! Ramsey buys a new Ferrari [Internet] (2009), `http://www.dailymail.co.uk/tvshowbiz/article-1172653/Thats-lot-ready-meals-Gordon-Ramsay-buys-new-Ferrari.html` (accessed December 12, 2012)
35. McCracken, G.: Who Is the Celebrity Endorser? Cultural Foundations of the Endorsement Process. Journal of Consumer Research 16(3), 310–321 (1989)
36. Mckinley, M.C., Lowis, C., Robson, P.J., Wallace, J.M.W., Morrissey, M., Moran, A., Livingstone, M.B.E.: It's good to talk: children's views on food and nutrition. European Journal of Clinical Nutrition 59, 542–551 (2005)
37. Meyer, D.S., Gamson, J.: The Challenge of Cultural Elites: Celebrities and Social Movements. Sociological Inquiry 62(2), 181–206 (1995)
38. Miller, D.: The Dialectics of Shopping. University of Chicago Press, Chicago (2001)
39. Monaco, J.: Celebrity. Delta, New York (1978)
40. PBS. e^2 design: series overview [Internet] (2008), `http://www.pbs.org/e2/design.html` (accessed December 12, 2012)
41. Prior, M.: News vs. Entertainment: How Increasing Media Choice Widens the Gaps in Political Knowledge and Turnout. American Journal of Political Science 49(3), 577–592 (2005)
42. Rojek, C.: Celebrity. Reaktion Books, London (2001)
43. Schickel, R.: Imitate Strangers: The Culture of Celebrity. Fromm International, New York (1985)
44. Singer, S.: Warrior one. Vogue (October 2002)
45. Telegraph.co.uk. Gordon Ramsay accused of using ready meals in his restaurants [Internet] (2009), `http://www.telegraph.co.uk/news/newstopics/celebritynews/5169631/Gordon-Ramsay-accused-of-using-ready-meals-in-his-restaurants.html` (accessed December 12, 2012)
46. Thegreatwarming.com. Americans must act to address Climate Change [Internet], `http://www.thegreatwarming.com/calltoaction/calltoaction.html` (accessed December 12, 2012)

47. Thrall, T., Lollio-Fakhreddine, J., Berent, J., Donnelly, L., Herrin, W., Paquette, Z., Wenglinski, R., Wyatt, A.: Star Power: Celebrity Advocacy and the Evolution of the Public Sphere. The International Journal of Press/Politics 13(4), 362–385 (2008)
48. Turner, G.: Understanding Celebrity. SAGE, London (2004)
49. Vivanco, L.A.: The Work of Environmentalism in an Age of Televisual Adventures. Cultural Dynamics 16(1), 5–27 (2004)
50. Warner Brothers: Frontiers of the New Century: The 11th Hour. New Perspectives Quarterly 24(4), 38–57 (2007)
51. West, D.M., Orman, J.: Celebrity Politics. Prentice Hall, Upper Saddle River (2003)
52. Wiki, Baribari Value [Internet] (2008), http://en.wikipedia.org/wiki/Baribari_Value (accessed December 12, 2012)

Directional Bias between Communities of a Production Network in Japan

Takashi Iino and Hiroshi Iyetomi

Abstract. We investigate a production network constructed by about 800 thousand firms in Japan with focus on its transaction flow between communities. Communities detected by maximizing modularity often contain nodes with common properties such as characterized by regions or industry sectors. Communities may thus upstream-downstream relationship according to their characteristics. Such directional bias of the connections between communities is evaluated through a polarization matrix of the network direction. We also devise a visualization method for directed network based on physical model. We attempt to draw a map of Japanese transaction flow in viewpoint of community structure.

1 Introduction

Many kinds of real complex network attract scientific interest. They are neither regular nor random networks and in fact are composed of many clusters of nodes in networks. Such clusters called community often contain the elements that have common behavior or features in networks.

To detect community structure, Newman proposed a quality function called modularity [1, 2, 3]. The idea of the modularity is based on quantifying statistically unforeseen arrangements of edges. Finding the division at the highest modularity value determines an optimum community structure in a network.

In the present paper we investigate community structure of a production network in Japan based on the same data in [4, 5, 6, 7]. The nodes and links in the production network correspond to firms and their mutual transaction relations. The production network consists of about 800 thousand firms and four million transaction relations. It is no exaggeration to say that the network is so large that it virtually covers the whole production activities in the nation. The analysis of the network may thus be able to give a new insight into the Japanese economy. The production network is analyzed in viewpoint of community [5] and more detailed subcommunity strucutures [7]. Some of the communities or subcommunities are characterized by regions or industry sectors.

Takashi Iino · Hiroshi Iyetomi
Faculty of Science, Niigata University, Ikarashi, Niigata 950-2181, Japan
e-mail: iino@ad.sc.niigata-u.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 273–280.
springerlink.com

Here we shed light on the direction of transaction. The data of the production network contain the information of the direction. We evaluate the directional bias of the connections between the communities, which cannot detect only separating into communities. We also devise visualization for directed network based on spring-electorical model [6] by adding a “magnetic” effect. We attempt to draw a map of Japanese transaction flow in viewpoint of community structure.

We first detect the community structure of the directed production network. For easy comparison with previous studies, we employ the community structure in [7], which is detected without information of direction. In other words, we separate the production network by the community structure in [7]. We then introduce a polarization matrix to evaluate the bias of the direction and devise visualization for directed networks.

2 Method

2.1 Modularity

Let us suppose that a network V is divided into L nonnull subsets $\{V_1, V_2, \cdots, V_L\}$ which do not overlap each other. The modularity Q is then defined by

$$Q = \sum_{i=1}^{L} \left(e_{ii} - a_i^2\right), \tag{1}$$

where e_{ii} denotes the link density within subset V_i, and a_i represents the fraction of links that connect into subset V_i. Therefore, $a_i a_j$ corresponds to the expectation value of the link density between subset V_i and subset V_j, and a_i^2 gives the expectation value of e_{ii} for the uniform random null model. The term e_{ii} is canceled perfectly by a_i^2 when the network is encircled by a single set or separated randomly into subsets. When the link density of the division surpasses the expectation value of link density under the null model, the modularity becomes large value.

To find the division with the optimized modularity value, we employ the bisection method worked out by ourselves in [5], which is a top-down method taking advantage of simulated annealing. The maximized modularity value by the bisection method surpasses that by the greedy method [2, 8, 9] especially for large-scale networks. Although the bisection method needs more computational time than the greedy method, it can detect communities of networks as large as the submillion production network in realistic time.

2.2 Polarization of the Network Direction

To quantify the bias of the direction between communities, we define the polarization of the network direction as

$$P_{ij} = \frac{A_{ij} - A_{ji}}{A_{ij} + A_{ji}}, \tag{2}$$

where A_{ij} is an element of adjacency matrix by communities. The element A_{ij} is defined by the sum of the elements of adjacency matrix by nodes as

$$A_{ij} = \sum_{l \in i} \sum_{m \in j} A_{lm}, \tag{3}$$

where l and m represent nodes, and i and j represent communities. By definition, the polarization matrix is antisymmetric ($P_{ij} = -P_{ji}$).

2.3 Visualization for Directed Networks

Various algorithms have been developed to visualize networks [10], and each algorithm shows up different structural feature. We adopt a spring-electrical model; pairs of nodes with direct links are physically connected with springs and any pairs of nodes repel each other through a repulsive Coulomb force [6, 11]. The interaction force between nodes for the model is explicitly written as

$$F(r_{lm}) = -k_{lm} r_{lm} + \frac{q_l q_m}{r_{lm}^2}, \tag{4}$$

where k_{lm} is a spring constant between nodes l and m. If the nodes are directly connected, $k_{lm} = k$. Otherwise, $k_{lm} = 0$. The symbol q_l denotes a Coulomb charge for node l. We anticipate that the ground state in the model leads to a easily viewable visualization reflecting topological structure in the network. The spring-electrical model represents communities as clusters in the visualized images. The nodes in a same community have dense inner-community links, thus they are pulled together by dence springs against convergence of Coulomb charge.

In this paper, we use the model to visualize the network of communities. The spring constant between communities i and j, represented as k_{ij}, is defined as proportional to the number of links between these communities. The Coulomb charge of community i, represented as q_i, is defined as proportional to the number of nodes in community i.

We then attempt to deal with the direction in the visualization of networks. The interaction force (4) does not contain information of the direction. We suppose an "uniform field" $\boldsymbol{H}$ for directed links and define an additional "magnetic" force as

$$\boldsymbol{F}_{\text{mag}} = q_{\text{mag}} \left(k_l^{\text{in}} - k_l^{\text{out}} \right) \boldsymbol{H}, \tag{5}$$

where q_{mag} denotes magnetic charge, and k_l^{in} (k_l^{out}) represents the in-degree (out-degree) of node l. To visualize network of communities, we define the magnetic force on community i as

$$\boldsymbol{F}_{\text{mag}} = q_{\text{mag}} \sum_{l \in i} \left(k_l^{\text{in}} - k_l^{\text{out}} \right) \boldsymbol{H}. \tag{6}$$

The force works to sort the nodes or the communities into upper or lower stream.

To obtain an optimized configuration in the model, we employ a molecular dynamics (MD) method [12, 13]. The MD simulation works well to reproduce an ordered structure with the lowest energies such as crystals of materials through slow cooling starting from any initial configurations. We expect it is also successful in visualizing the network.

3 The Production Network

The data used in the present study was compiled by TOKYO SHOKO RESEARCH, LTD., which has been gathering information on firms through investigation of financial statements, corporate documents and hearing-based survey. The production network constructed here is an unweighted network, because information on the amounts of transactions is not available. The element A_{ij} of adjacency matrix thus takes 1 or 0, depending on whether there is a transaction relation between firms i and j or not. We simply ignore the weights in this paper.

We define the direction of links in the production network along order flow. If firm A orders to firm B, the direction of link is defined from A to B. In other words, if firm A is a customer of B, the direction of link is defined as A $\rightarrow$ B.

We paid attention to the largest connected component of the network. It contains $773,670$ nodes and $3,295,325$ links, which covers more than 99% of the whole data. The degree distribution $P(k)$ follows a power law, i.e., $P(k) \propto k^{-\gamma}$ with $\gamma \simeq 2.32$. Thus the production network possesses the scale-free property.

4 Community Structure in the Production Network

We extracted 118 communities in the production network with the maximized modularity value 0.654 [7]. The distribution of community size has gap. There are 16 communities that have over $10,000$ nodes and About 90 communities consist of less than 100 nodes. The size of the several remaining communities ranges from 100 to $10,000$.

Attributes of firms in top 10 large communities in the production network are shown in Table 1. We looked up the prefectures and industry sectors to which firms in each community belong. The industry sectors conformed to Japan Standard Industrial Classification are shown in Table 2, where only terms used in this paper are listed. To simplify notation, the names of industry sectors are abbreviated by three capitals.

The communities in the production network are well characterized by regions and industry sectors. The largest community consists of the firms of manufacture and wholesale trade of machinery. The second, fourth, fifth, sixth and ninth largest communities contain many constructions. Furthermore, the fourth largest community is strongly characterized as a geographical region; the ratio of Kyushu firms to all firms in the fourth largest community is more than 90 %. The ninth largest community is also strongly characterized as a geographical region; the ratio of firms in Tohoku region to all firms in the ninth largest community is more than 90 %. The third largest community consists of the firms of manufacture, wholesale trade and retail trade that deal with food.

Table 1 Attributes of firms in major communities in the production network. Top three major prefectures and industry sectors in each community are described. Decimals shown in parentheses represent the ratio of firms of corresponding attribute to community size. The color legend is corresponding to Fig. 2.

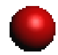
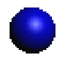
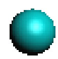
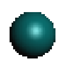
rank	size	prefecture (fraction)	industry (fraction)	color legend
1	88,840	Tokyo (0.189) Aichi (0.120) Osaka (0.110)	M-GM (0.144) W-ME (0.124) M-FM (0.105)	
2	84,280	Niigata (0.117) Tokyo (0.094) Aichi (0.086)	C-GE (0.400) C-SP (0.228) C-EI (0.075)	
3	78,529	Tokyo (0.110) Hokkaido (0.088) Aichi (0.055)	W-FB (0.262) M-FO (0.172) R-FB (0.137)	
4	48,903	Fukuoka (0.298) Kagoshima (0.137) Kumamoto (0.134)	C-GE (0.359) C-SP (0.146) C-EI (0.124)	
5	47,085	Aichi (0.135) Kanagawa (0.121) Tokyo (0.113)	C-GE (0.362) C-SP (0.205) W-BM (0.094)	
6	45,622	Tokyo (0.152) Kanagawa (0.112) Osaka (0.098)	C-EI (0.473) C-GE (0.101) W-ME (0.072)	
7	44,736	Tokyo (0.167) Aichi (0.094) Osaka (0.083)	T-RF (0.163) R-MB (0.123) R-FH (0.120)	
8	39,591	Tokyo (0.180) Osaka (0.118) Aichi (0.080)	R-DA (0.251) W-TA (0.201) M-AP (0.132)	
9	37,524	Fukushima (0.197) Aomori (0.167) Miyagi (0.154)	C-GE (0.357) C-SP (0.159) C-EI (0.089)	
10	37,488	Tokyo (0.320) Osaka (0.096) Aichi (0.068)	M-PR (0.175) W-MI (0.127) M-PP (0.077)	

Table 2 Abbreviation of industry sectors conformed to Japan Standard Industrial Classification defined by Statistics Bureau in Ministry of lnternal Affairs and Communications

abbreviation	Major groups	Divisions
C-EI	equipment installation work	construction
C-GE	construction work, general, including public and private construction work	
C-SP	construction work by specialist contractor, except equipment installation work	
M-AP	manufacture of apparel and other finished productsmade from fabrics and similar materials	manufacturing
M-FM	manufacture of fabricated metal products	
M-FO	manufacture of food	
M-GM	manufacture of general machinery	
M-PP	manufacture of pulp, paper and paper products	
M-PR	printing and allied industries	
R-DA	retail trade (dry goods, apparel and apparel accessories)	retail trade
R-FB	retail trade (food and beverages)	
R-FH	retail trade (furniture, household utensil and household appliance)	
R-MB	retail trade (motor vehicles and bicycles)	
T-RF	road freight transport	transport
W-BM	wholesale trade (building materials, minerals and metals, etc.)	wholesale trade
W-FB	wholesale trade (food and beverages)	
W-ME	wholesale trade (machinery and equipment)	
W-MI	miscellaneous wholesale trade	
W-TA	wholesale trade (textile and apparel)	

5 Result: Direction of Transactions between Communities

To elucidate directional bias between communities, we examined the polarization (2) between all pair of the top 10 major communities as shown in Fig. 1. To understand easily, we visualized the network of communities as shown in Fig. 2 by the method in Sec. 2.3. Each sphere represents a community, and the volume of a sphere is proportional to the number of nodes in corresponding community. The radius of arrows between spheres is proportional to the number of links between corresponding pair of communities. In the visualization, the communities of upper stream are configured top, and the lower stream communities are configured bottom by the magnetic uniform field $\boldsymbol{H}$.

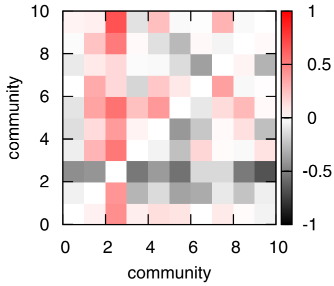


Fig. 1 The polarization matrix of the communities in the production network

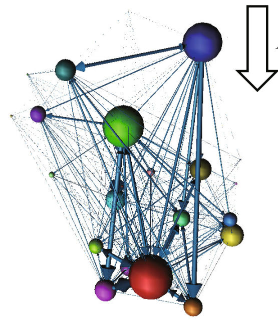


Fig. 2 The visualized image of the communities shown in Table 1

The third largest community is configured upper stream in the major communities. In other words, the firms that deal with food tend to be order side for other major communities. Especially, the 10th largest community, it consists of many firms of printing and paper products, has acceptance of order from the third largest community ($P_{3,10} = 0.675$).

The second largest community is configured upper stream for other major communities but the third largest community. As shown in Fig. 1, the second largest community tends to order to the fourth and ninth largest communities. In other words, the construction firms of central region in Japan tend to order to the local construction firms ($P_{2,4} = 0.289, P_{2,9} = 0.237$).

The community structure in this paper is separated without direction information of network. However, some of communities have biased transaction direction. The order flow between industrial clusters in Japan is reflected to the bias of transaction direction between communities.

6 Conclusion

The production network of Japanese firms is separated as some communities that are characterized by regions or industry sectors. The rough industrial structure in Japan is reflected in the communities of the production network. We evaluated the directional bias of orders between communities by the polarization matrix. Furthermore, for intuitive understanding, we visualized the network of communities by spring-electrical model with magnetic effect.

Despite the community detection is performed without direction information, we find that some of directions of orders between the communities are biased. The community of food firms is configured in upper stream for other top 10 major communities. The community of construction firms in central Japan tends to order to local construction firms. The polarization matrix and the visualization by spring-electrical model with magnetic effect are useful to get overview of the flow in networks.

Acknowledge. We thank Hideaki Aoyama, Yoshi Fujiwara, Yuichi Ikeda and Wataru Souma for useful discussions on the subjects treated in this article. We also thank the Research Institute of Economy, Trade and Industry (RIETI) for having provided us with the data on transaction relationship between firms in Japan. The present study was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Grants-in-Aid for Scientific Research (B), Grant No. 22300080 (2010-12).

References

1. Newman, M.E.J., Girvan, M.: Phys. Rev. E 69, 026113 (2004)
2. Newman, M.E.J.: Phys. Rev. E 69, 066133 (2004)
3. Newman, M.E.J.: PNAS 103, 8577 (2006)
4. Fujiwara, Y., Aoyama, H.: Eur. Phys. J. B 77, 565 (2010)
5. Iino, T., Kamehama, K., Iyetomi, H., Ikeda, Y., Ohnishi, T., Takayasu, H., Takayasu, M.: J. Phys.: Conf. Ser. 221, 012012 (2010)
6. Kamehama, K., Iino, T., Iyetomi, H., Ikeda, Y., Ohnishi, T., Takayasu, H., Takayasu, M.: J. Phys.: Conf. Ser. 221, 012013 (2010)
7. Watada, J., Phillips-Wren, G., Jain, L.C., Howlett, R.J., et al. (eds.): Intelligent Decision Technologies. SIST, vol. 10, pp. 537–545. Springer, Heidelberg (2011)
8. Clauset, A., Newman, M.E.J., Moore, C.: Phys. Rev. E 70, 066111 (2004)
9. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: J. Stat. Mech., p. 10008 (2008)
10. Di Battista, G., Eades, P., Tamassia, R., Tollis, I.G.: Graph Drawing: Algorithms for the Visualization of Graphs. Prentice Hall, Englewood Cliffs (1998)
11. Hu, Y.: The Math. J. 10, 37 (2006)
12. Allen, M.P., Tildesley, D.J.: Computer Simulation of Liquids. Oxford University Press, Oxford (1987)
13. Frenkel, D., Smit, B.: Understanding Molecular Simululation: From Algorithms to Applications, 2nd edn. Academic Press, San Diego (2002)

Enterprise Insider Detection as an Integer Programming Problem

Sergii V. Kavun, Ivan V. Sorbat, and Vyacheslav V. Kalashnikov

Abstract. The authors propose a new approach to the mathematical interpretation to overcome the problems associate with the identification (detect) of the unauthorized acquisition (intentional and accidental) of information by insiders within an enterprise of any industry. The approach is based on the use of mathematical tools based on integer programming. Secondly, the practical implementation of this approach has suggested that this approach has possible uses for applied software or as part of analytical tool to detect high risk behavior. This new approach can be using in complex problems with other tools or applications for detection of insiders within different enterprises. The authors propose a new approach that can be used as a tool within complex enterprises to increase the accuracy of detecting unauthorized activities of insiders.

1 Introduction

The progress of civilization and the constantly increasing role of various technologies in everyday human activities, has leaded to the need for permanent development of information security systems for enterprises (e.g., such as the commercial risk level, various commercial secrets, the income level, a strategic development plan, etc.). The importance of safety protection subsystem for such systems is undeniable. One of the approaches [1] to ensure the safety and protection of subsystems is to solve the task of identification (detection) of unauthorized access by insiders within the enterprise.

S.V. Kavun · I.V. Sorbat
KhNEU, Kharkiv, Ukraine
e-mail: kavserg@gmail.com

V.V. Kalashnikov
ITESM, Campus Monterrey, Monterrey, Mexico
and
CEMI RAS, Nakhimovsky pr. 47, Moscow, Russia
also, SumGU, Rimsky-Korsakov st. 2, Sumy, Ukraine
e-mail: kalash@itesm.mx

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 281–289.
springerlink.com

In the process of solving this task at the enterprises, a large variety of problems has been solved [2], among which is the mathematical interpretation of such a problem. As a rule, when trying to identify (detect) unauthorized access by insiders, enterprise rely on various sociological methods on the base of visual observations. The analysis carried out by the authors [3], has shown the prospective usefulness of the mathematical interpretation of the problem in question.

The well-known experts and scholars in this area are: Vereen V.P. [1], Klebanova T.S., Kizim M.O. [4], Gurov M.P., Kudryavtsev Y.A. [5], Oleynikov E.A. [6], Kurkin N.V. [7], Shkarlet S.N. [8]. Their works have demonstrated a systematic approach to address threats to information and economic security, but most of these studies relate to external threats. Since the unauthorized information access within an enterprise by insider activity brings financial losses, there is a need to address the urgent task of preventing or identifying an insider or a group of insiders (the insider trading activity).

2 Statistical Information in Area of Insider Detection

The authors have conducted a study based upon the open sources of information leakage, on data from the world of statistics, which were published in the media, the Web forums, reporting analyst firms, thematic blogs, and other public resources. Table 1 shows the statistics (on base the aggregation by authors from many open sources) of the number of incidents of intentional and accidental leaks of information for a certain period of time in the world (on base the open sources[9]).

Fig. 1 shows the dynamics of information leakages for a given period of time, built on the basis of the calculated data (Table 1). Also were calculated the trends depending on the basis of polynomial's curves and values of reliability (R^2), which a researcher may use for the further predictions.

In addition, a statistical analysis may assist in providing an indication of the magnitude of the problem of inside intrusions, its detection and non detection for the modern enterprises and how important it is to overcome the problem. Results of the research, has the potential to indicate in which regions of world the enterprises must increase their efforts for enhancing the tools for insider detection task. However, all the well-known studies by experts and scientists do not propose a formalization of the task in a mathematical form, which does not allow the use of mathematical methods for the modern enterprises.

Table 1 Data on the types of information leaks over a certain period of time

Type of Leak	2010		2009		2008		2007	
	quantity	*%*	*quantity*	*%*	*quantity*	*%*	*quantity*	*%*
Deliberate	402	48,0	375	51,0	241	45,5	154	29
Accidental	390	46,4	320	43,5	223	42,1	376	71
Not established	47	5,6	40	5,4	66	12,5	-	-

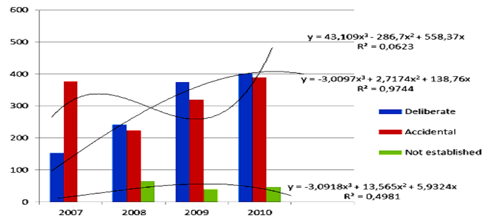


Fig. 1 The dynamics of information leakage for a given period of time

3 Problem Description

In the course of its commercial activities, various organizations are subject to economic crime, negligence of employees, which leads them to financial, physical, temporal, economic and other kinds of losses. Such activities of the staff are called insider ones. The problem of insider's detection has been considered in the report of Computer Security Institute in 2007 [10]. From this year (2011) this problem was stated as being in first place in world among all set of threats and vulnerabilities. Thus, the problem of insider's detection and defense preceded the problem of virus's defense. Especially susceptible to insider attacks are the enterprises of bank and those associated with the financial sector. Insider attacks have a very high level of latency (concealment) and the lowest level of detection. But, these methods are provided as only preventing the consequences of insider attacks, and are not providing for the detection of insiders within the enterprise.

Therefore, the problem of insider detection for the modern enterprises and their activity and financial position may be considered of one the most important task that are required to be undertaken. Because, for the author's opinion (on base the interpretation of the Pareto principle [11]), a leakage of 20% of commercial secrets of enterprise in 80% of cases leads to collapse of this enterprise. The many enterprises learned how to be on the defensive from external threats (cyber-attacks, intrusions, viruses and etc.) but before internal threats (the insiders), many enterprises may be considered almost completely defenseless!

4 The Mathematical Description of This Approach

Definition 1 (author: S.V. Kavun). An Insider is a fellow, whose work varies in time under the influence of external, internal, and individual causes [12] (it's the first distinction from other definitions). This work reflects a readiness of this fellow for actions. In addition, the socio-cultural environment of this fellow can be regarded as a violation of existing standards (disclosure of information with restricted access) and traditions (not doing the job, it's the second distinction).

Definition 2 (author: I.V. Sorbat). Insider Information is substantial undisclosed Public Service Information (PSI) for the enterprise [13]. This information if disclosed could lead to the loss of competitiveness of the company or into its collapse (it's the first distinction from other definitions). Employees who have this information are typically the system's administrators or the owners. Employees who received this information are called insiders. All these processes refer to the sphere of economic and information security [14] (it's the second distinction).

The aim of the paper is to provide a mathematical formalization of the task identifying insiders in the company based on graph theory for the possibility of using well-known mathematical techniques (such as the methods of linear and integer programming). Besides this, the authors show the possibility of solving the non-formalized problem by various methods of mathematical analysis. Lastly, the authors prove the theoretical possibility of the mathematical interpretation of non-formalized problems or tasks.

For the conduct of an objective study, in this approach what is needed is the consideration of the first types of organizational and functional structure of the enterprise [9]. What has been interpreted are the most common Organizational and Functional Structure of the Enterprise as a graphical model (Fig. 2).

The proposed model is accepted as standard model, but this model does not reflect fully the real structure of the enterprise. For this graph model introduction the following notation of vertices of graph **G** is made: **D** – Director (if the enterprise has a board of directors, then gets the set $\{\mathbf{D_w}\}$, where **w** is the number of directors (general, commercial, financial, etc.); **DD** – Deputy Director (if the enterprise has a board of directors, then get the set $\{\mathbf{CCO_w} \equiv \mathbf{DD_w}\}$, were $\mathbf{CCO_w}$ is the Chief Communications Officers); **S** – Secretary (if the enterprise has a board of directors, then gets the set $\{\mathbf{CCO_w}\} \equiv \{\mathbf{S_w}\}$); **L** – Lawyer (if the enterprise has a legal department, then gets the set $\{\mathbf{CLO_i}\} \equiv \{\mathbf{L_q}\}$, where **q** is the number of employees in the department, $\mathbf{CLO_i}$ is the Chief Legal Officers); **M** – Manager (if the enterprise has several departments, then gets the set $\{\mathbf{M_k}\}$, where **k** is the number of managers in the departments); **I** – Engineer (if the enterprise has several departments with engineers, then gets the set $\{\mathbf{I_z}\}$, where **z** is the number of engineers in the department); **T** – Technologist (if the enterprise has several departments, then gets the set $\{\mathbf{T_s}\}$, where **s** is the number of technologists at the department).

To construct a graph **G,** in which the vertices will determine the rank (a position). The ribs of this graph will be denoted by the Conditional Weights (**CW**) of Information with Restricted Access (**IRA**), which is suggested to be calculate by the formula

$$\mathbf{CW^{ij}_{IRA} = V^{ij} \times {}^{S}K^{ij} \times {}^{N}K^{i}_{C}}. \tag{1}$$

Here $\mathbf{V^{ij}}$ is the amount of the information with restricted access, which circulating within the electronic and paper documents of the enterprise between the i^{th} and j^{th} vertices of a graph **G**. The variable $\mathbf{V^{ij}}$ is determined by monitoring the traffic and dynamically updated within a given interval of time; $\mathbf{{}^{S}K^{ij}}$ is a coefficient of the importance of the information with restricted access, transmitted between the i^{th} and j^{th} vertices of the graph **G**; this coefficient is defined by experts or on the base of the statistical information on the enterprise; $\mathbf{{}^{N}K^{i}_{C}}$ is a normalized coefficient of coherence of the information with restricted access, which is connected to the i^{th} vertex of the graph **G**; **i** and **j** is an adjacent vertices of the graph **G**, between which circulates the information with restricted access with conditional weight.

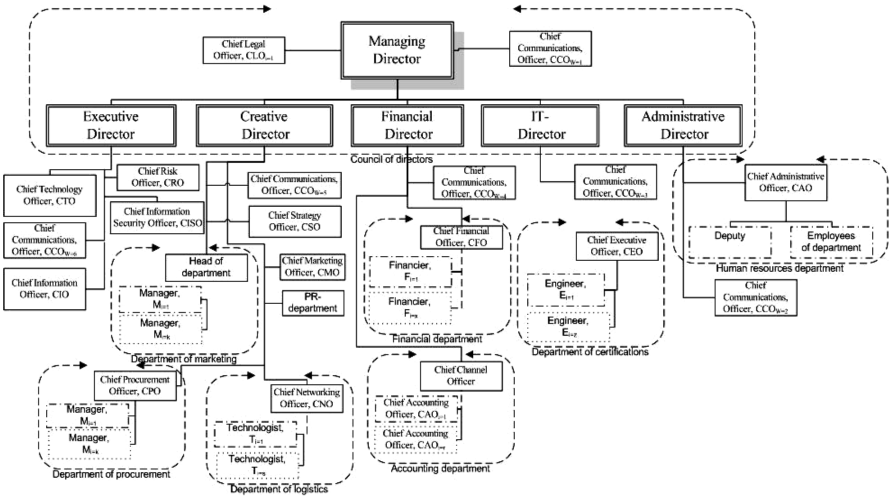


Fig. 2 Graph's model of the most common Organizational and Functional Structure of the Enterprise (Author: Kavun Sergii V.)

By using the concept of graph theory this introduces the following notation:

n is the number of vertices in the graph **G**, these vertices creates the set $\{\mathbf{v_i}\}$, where $\mathbf{i = 1,\dots, n}$; **m** is the number of ribs in graph **G**, these ribs creates the set $\{\mathbf{r_j}\}$, where $\mathbf{j = 1,\dots, m}$.

At the practical implementation step obtain the following relationships:

if $\mathbf{m > n}$, this means that in organizational and functional structure of the enterprise the relationship between employees are not cycled;

if $\mathbf{m \neq 2n}$, this means that within the organizational and functional structure of the enterprise, any worker while performing their official duties, may be contacted by more than two (usually a leader and a junior) other staff members.

Then graph **G** will describe the model the organizational and functional structure has of the enterprise which helps in the next set: $\mathbf{R = \{r_j\}}$, where $\mathbf{j = 1,\dots, (n - 1)}$ and denotes the number of ribs; $\mathbf{V = \{v_i\}}$, where $\mathbf{i = 1,\dots, n}$ is the number of vertices. Then the task of insider detection on the enterprise can be reduced to the task of finding the value of $\mathbf{CW^{ij}_{IRA}}$ which transmitted in the direction $\mathbf{i \to j}$.

Introduce the factor of significance $\mathbf{KS_{ij}}$ and a functional **F** that creates dependency and is defined with the following expression:

$$F = \max_{i,j} CW_{IRA}^{ij} = \max_{i,j}\left(V^{ij} \cdot K_{ij}^{S}\right), \tag{2}$$

where $\mathbf{i = 1,\dots, n, j = 1,\dots, n - 1}$.

Taking into account the introduced graph's model of the most common organizational and functional structure of the enterprise (was showed on Fig. 2), it could be assumed that the top management link always sends out some information (orders, instructions, orders, specification, etc.) to the lower level management. Following the implementation of this process the lower level management always returns reports with such information to the top management. Thus, in graph **G**, all ribs are bidirectional. For example construct a connectivity matrix $\mathbf{M_c}$ for the graph **G** (Table 2).

Each element of the matrix $\mathbf{M_c}$ may be equal to “1”, if the link between vertices **i** and **j** of graph **G** is exists, and equal to “0”, if the link is missing.

Axiom 1. All entries at the main diagonal of matrix $\mathbf{M_c}$ (recall that this matrix showing a links between all employees on the enterprise) are zero because communication (transmission) of the employee with himself does not exist. Otherwise this process will be a cyclic process that for the enterprise is not permissible. The main diagonal of matrix $\mathbf{M_c}$ shows in Table 2 with help gray color.

Table 2 Connectivity matrix M_c (presented by authors only for example)

i \ j	D_w	DD_w	L_q	S_w	M_k	I_z	T_s	$^iK_{CW}$	Δ,%
D_w	0	1	1	1	1	0	0	4	20
DD_w	1	0	1	1	1	0	0	4	20
L_q	1	1	0	0	0	0	0	2	10
S_w	1	1	0	0	1	0	0	3	15
M_k	1	1	0	1	0	1	1	5	25
I_z	0	0	0	0	1	0	0	1	5
T_s	0	0	0	0	1	0	0	1	5

Introduce the coefficient of coherence $\mathbf{^iK_{CW}}$ for each category of employees on base of the organizational and functional structure of the enterprise:

$$\mathbf{^iK_{CW} = \sum x_{ij}.} \tag{3}$$

Then the calculation can get the $\mathbf{K_{CW} = \{^iK_{CW}\}}$ and for the example, which is shown in Table 2, and then the calculation can get the $\mathbf{K_{CW} = \{4, 4, 2, 3, 5, 1, 1\}}$.

Here $\mathbf{^iK_{CW}}$ is the coefficient, which shows connectivity ratio as the importance of conventional unit (or department) personnel. The higher the ratio $\mathbf{^iK_{CW}}$, the greater is the $\mathbf{CW^{ij}_{IRA}}$.

For any further possible accounting, the coefficient $\mathbf{^iK_{CW}}$ must be normalized. Then one comes to the expression

$$^N K^i_{CW} = \frac{K^i_{CW}}{\sum_{j=1}^{n-1} K^j_{CW}} \tag{4}$$

Therefore, using the formula (4) obtain the following values (base on data from the example in Table 2) $\mathbf{^NK_{CW} = \{^NK^i_{CW}\} = \{0{,}2; 0{,}2; 0{,}1; 0{,}15; 0{,}25; 0{,}05; 0{,}05\}}$.

For the further use of the factor of importance of different indicators of physical nature, the expression (1) will be transformed into the following form

$$\mathbf{CW^{ij}_{IRA} = V^{ij} \times {}^NK^i_S \times {}^iK_{CW},} \tag{5}$$

where $^{N}K^{i}_{S}$ is the normalized importance factor of the information with restricted access, calculated by the following expression

$$^{N}K^{i}_{S} = \frac{\sum_{j=1}^{n-1} K^{ij}_{S}}{\sum_{i=1}^{n}\sum_{j=1}^{n-1} K^{ij}_{S}}. \tag{6}$$

Thus, based on the formulas (4) and (6), what can be considered is the possibility of using the dynamic coefficients obtained from the experts.

Axiom 2. The total volume of information with restricted access

$V_{IRA} = \sum_{i=1}^{n}\sum_{j=1}^{n-1} V^{ij}$ is always changing with time. On each node (i or j) in the graph **G,** the volume of information $\mathbf{V^{ij}}$ which is to be transferred from i^{th} to j^{th} node, this is changing, because in the work process, the employees makes adjustment in this information. This information with restricted access is always circulating within the enterprise. Then changing its total volume the information with restricted access $\mathbf{V_{IRA}}$ and its accompanying service's information are interrelated.

Theorem 1. Normalized importance factor the information with restricted access $\mathbf{^{N}K^{i}_{S}}$ for the i^{th} node on the graph **G** has no direct and no linear dependence on the total volume information $\mathbf{V_{IRA}}$ which is circulating on this graph.

Proof. Assume that inequalities $\mathbf{m > n}$ and $\mathbf{m \neq 2n}$ are valid, or else $\mathbf{K^{i}_{CW} \geq 3}$ for the i^{th} node on the graph **G**. Let **a** and **b** be the nodes of the graph **G**, which are adjacent to the i^{th} node on this graph. Consider the set $\mathbf{V = \{v_i\}}$, then $\mathbf{^{N}K^{i}_{S} = f\,(V^{ib} + V^{ia} + V^{bi} + V^{ai})}$ is a function depending on time. Further, consider a node for which nodes **c** and **d** are adjacent, while $\mathbf{c, d \in V = \{v_i\}}$. Then the right side of that the dependence $\mathbf{V^{a} = g\,(V^{ac} + V^{ad} + V^{ca} + V^{da})}$, is also a function depending on time. Since $\mathbf{f \neq g}$, then their dependence is not linear and not direct and has several variations of the distribution.

So, the problem of insider detecting on the enterprise can be represented as an integer programming problem:

$$CW_{IRA} = \min \sum_{i=1}^{n}\sum_{j=1}^{m}\left(V^{ij} \times {}^{N}K^{i}_{S} \times {}^{N}K^{i}_{CW} \times x_{ij}\right), \tag{7}$$

subject to

$$\begin{cases} V_{ij} \leq \sum_{k=1}^{i-1}\sum_{l=1}^{j-1} V_{kl}; \ \textit{for all } i, j; \\ x_{ij} = \begin{cases} 1, & \textit{if the criterion is applicable to the employee;} \\ 0, & \textit{otherwise;} \end{cases} \\ {}^{N}K^{i}_{CW} \in \{0;1\}; \end{cases}$$

where $\mathbf{x_{ij}}$ is the corresponding element of matrix $\mathbf{M_c}$.

Thus, what was presented is a possible of mathematical interpretation of the task of insider detection on the enterprise as the task of integer linear programming[15]. Based on the opinions of the authors, further mathematical interpretation in the use of a known mathematical method for solving this task and, on base of these methods, what is suggested is that an optimal solution is possible.

5 Discussion

The reduction in the number of accidental leaks of information during the year 2010 compared to the previous one may be due to the application of various methods, such as hardware and software systems to detect leaks [16]. The majority of these techniques and tools allow the enterprise to identify the past actions committed by insiders, but not to predict and prevent them. For further research and improvements, any refinements of the approach need to take into account the interpretation focus using graph theory [15].

For further research opportunities some of the input variable Δ (which will be measured in percent) in the matrix $\mathbf{M_c}$ (was showed in Table 2 as the rightmost column) need improvement. This variable will be determined at some bound or threshold which will be create a risk's zone for the enterprise. For example, on the some enterprise was accepted a risk's zone with the threshold which equal 65% (it's the example for Table 2). Then if this is done the ranking of the office's positions on this enterprise will have a decreasing order of the coefficient of coherence $^{\mathbf{i}}\mathbf{K}_{CW}$, then what will be seen is that this risk's zone applies to their manager, director and deputy director. Thus, it's the office's positions on the enterprise which could be the potential insider. Additionally, the variable Δ could be the number of employees on the enterprise that, would the insiders, then lead to a collapse. This proposition from the authors will need to be the subject of further research.

6 Conclusion

In this paper, we have reduced the task to identify insiders on the enterprise to the integer programming problem. The latter can be solved making use of the well-known mathematical programming methods, thus obtaining the optimal solution of the original problem. Also the proposed approach and its interpretation as a graph model have demonstrated the possibility of the use of optimization methods and further analysis tools [17]. In addition, the expression (7) can mathematically formalize the implementation of the proposed approach, which will further lead to a linear programming problem. The practical implementation of this approach (based on an example) has suggested that this method has solid mathematical bases and allows researchers to aggregate statistics.

Finally, the analysis of using this approach can be used as a methodological basis (considering the complexities of other approaches and methods) for a preliminary study of the applicant for a position within the enterprise. This will get (to a certain extent) an objective assessment of the relevance of the applicant for the position and would be of interest for the enterprise and there management.

References

1. Vereen, V.: Economic crimes, p. 215. Delo, Moscow (2002)
2. Kavun, S.: Life cycle of economic security. Development Management 6, 17–21 (2008)
3. Kavun, S.: Insider – a threat to economic security. Development Management 6, 7–11 (2008)
4. Geyets, V.: The modeling of economic security: power, region, enterprise: monograph. In: Geyets, V., Kizim, M., Klebanova, T., Chernyak, O. (eds.), p. 240. INZHEK, Kharkiv (2006)
5. Gurov, M., Kudryavtsev, Y.: The shadow economy and economic crime in the questions and answers: Textbook, p. 237. Saint-Petersburg State University of MIA of Russia (2002)
6. Oleynikov, E.: Economic and national security: the textbook for high schools, p. 768. AM: Exam (2005)
7. Kurkin, N.: Managing the economic security of the enterprise: monograph. In: Kurkin, N. (ed.) p. 452. Art-Press, Dnepropetrovsk (2004)
8. Shkarlet, S.: The economic security of enterprise: innovation aspect: monograph. In: Shkarlet, S. (ed.), p. 436. Books view of the NAU, Kiev (2007)
9. Kavun, S.: Statistical analysis in area of economic and information security (and part 2). ES INFECO. International Research Portal Information and Economic Security (2011-2012), `http://infeco.net`
10. Computer Security Institute, 2007 CSI Computer Crime and Security Survey (2007), `http://i.cmpnet.com/v2.gocsi.com/pdf/CSISurvey2007.pdf`
11. Juran, J.M.: Juran's Quality Control Handbook. McGraw-Hill (ASQC Quality Press), New York (1988)
12. Kavun, S.: Mathematical interpretation of the task of identifying insider in the organization (enterprise). In: Kavun, S., Sorbat, S. (eds.) Economic: Problems of Theory and Practice: BZ. Science Price 246, 862–869. Science Price, BZ (2008); At 5 parts. – Part IV. AUTHORITY, Dnepropetrovsk
13. Kavun, S.: Model of insiders in the company. In: Kavun, S., Sorbat, I. (eds.) Securitatea informațională: ed. a 5-a jubiliară Conf. Intern. (April 16-17), p. 76. ASEM, Chişinău (2008)
14. Kavun, S.: Informational security: textbook. In: Kavun, S. (ed.), p. 368. Pub. of INZHEK, Kharkiv (2099)
15. Christofides, N.: Graph Theory. Algorithmic approach. Trans. from English, Moscow: World, 432 (1978)
16. Kavun, S.: Insiding – the problem of economic security in conditions of reforming the Ukrainian economy. In: Kavun, S., Sorbat, I. (eds.) Actual Questions with Problems of Economics: Science Magazine, vol. 4 (94), pp. 91–97. National academy of management, Kiev (2009)
17. Gordon, L.A.: Economic aspects of information security: An emerging field of research. In: Gordon, A.L., Loeb, P.M. (eds.) Information Systems Frontiers, vol. 8(5), pp. 335–337 (2006), `http://portal.acm.org/citation.cfm?id=1196620&dl=GUIDE&coll=GUIDE&CFID=35748673&CFTOKEN=37314539`

ES-DM: An Expert System for an Intelligent Exploitation of the Large Data Set

Amel Grissa Touzi and Mohamed Amine Selmi

Abstract. In meeting the challenges that resulted from the explosion of collected, stored, and transferred data, Knowledge Discovery in Databases (KDD) or Data Mining has emerged as an important research area. However, the approaches studied in this area have mainly been oriented at highly structured and precise data. Thus, the problem of exploit these data is often neglected. In this paper, we propose an intelligent approach for exploitation of these data. For this, we propose to define an Expert System (ES) allowing the user to easily exploit the large data set. The Knowledge Base (KB) of our ES is defined by introducing a new KDD approach taking in consideration another degree of granularity into the process of knowledge extraction. This set represents a reduced knowledge of the initial data set and allows deducting the semantics of the data. We prove that, this ES can help the user to give semantics for these data and to exploit them in intelligent way.

1 Introduction

Nowadays, we notice a growing interest for the Knowledge Discovery in Databases (KDD) methods. One of the important reasons to that fact is the increasing volume of the accumulated data by organizations that are under-exploited extensively. Several solutions were proposed, they are based on neural networks, trees, concept lattices, association rules, etc. [1]. KDD is a research area that considers the analysis of large databases in order to identify valid, useful, meaningful, unknown, and unexpected relationships. Several approaches in this area were proposed in the literature. Unfortunately these data mining is generally restricted to the application of discovery and modeling techniques within the KDD process. Thus, the problem of exploit these data is often neglected. Hence the following questions are worth asking: What is the goal to propose several approaches of extractions of knowledge whereas the user cannot benefit to better exploit these data?

In our opinion, this problem was not really neglected but it was not sufficiently treated since the generated rules by these approaches, exceeding some times of

Amel Grissa Touzi · Mohamed Amine Selmi
Université de Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis,
Bp. 37, Le Belvédère 1002 Tunis, Tunisia
e-mail: amel.touzi@enit.rnu.tn, amine.selmi@gmail.com

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 291–301.
springerlink.com

thousand rules, are not easily exploitable [2], [3]. In this paper, we interest to the intelligent exploitation of these data. For this, we propose to define an Expert System, called Expert System for Data Mining (ES-DM) allowing the user to easily exploit the large data set. Indeed, this ES can help the user to give the semantics of data and to optimize the research of information. Our ES-DM is composed of a Knowledge Base (KB), an Inference Engine (IE) and two interfaces: acquirement of the knowledge and user-interface. Since the majority of KDD approach generates a big number of rules that are not easily assimilated, we propose for the definition of the KB a new KDD approach taking in consideration another degree of granularity into the process of knowledge extraction. We propose to define rules between classes resulting from a preliminary classification on the data. The idea of the utilization of the fuzzy cluster degree for data reduction has also been proposed by several authors in the areas of symbolic data analysis [4],[5].

The rest of the paper is organized as follows: Section 2 presents the basic concepts of discovering association rules, Conceptual scaling and Formal Concept Analysis (FCA). Section 3 presents problems and limits of the existing knowledge discovery approaches. Section 4 presents our Expert system for Data Mining. Section 5 describes our new KDD approach for defining the Knowledge Base for ES-DM. Section 6 validates the proposed approach. We finish this paper with a conclusion and a presentation of some future works.

2 Basic Concepts

In this section, we present the basic concepts of discovering association rules, Fuzzy conceptual scaling and FCA.

2.1 Discovering Association Rules

Association rules mining have been developed in order to analyze basket data in a marketing environment. Input data are composed of transactions: each transaction consists of items purchased by a consumer during a single visit. Output data is composed of rules. An example of an association rule is “90% of transactions that involve the purchase of bread and butter also include milk” [6]. Even if this method was introduced in the context of Market Business Analysis, it can also be used to search for frequent co-occurrences in every large data set.

The first efficient algorithm to mine association rules is *APriori* [7]. The first step of this algorithm is the research of frequent itemsets. The user gives a minimum threshold for the support and the algorithm searches all itemsets that appear with a support greater than this threshold. The second step is to build rules from the itemsets found in the first step. The algorithm computes confidence of each rule and keeps only those where confidence is greater than a threshold defined by the user. One of the main problems is to define support and confidence thresholds. Other algorithms were proposed to improve computational efficiency. Among them, we mention CLOSED, CHARM and TITANIC.

2.2 *Fuzzy Conceptual Scaling and FCA*

Conceptual scaling theory is the central part in Formal Concept Analysis (FCA). It allows to introduce for the embedding of the given data much more general scales than the usual chains and direct products of chains. In the direct products of the concept lattices of these scales the given data can be embedded [8].

FCA starts with the notion of a formal context specifying which objects have what attributes and thus a formal context may be viewed as a binary relation between the object set and the attribute set with the values 0 and 1. Wille's definition of a concept is a pair consisting of a set of objects (the extension) and a set of attributes (the intension) such that the intension consists of exactly those attributes that the objects in the extension have in common, and the extension contains exactly those objects that share all attributes in the intension[8]. In [9], an ordered lattice extension theory has been proposed: Fuzzy Formal Concept Analysis (FFCA), in which uncertainty information is directly represented by a real number of membership value in the range of [0,1]. In this case, the similarity of a fuzzy formal concept is defined as follow:

Definition. The similarity of a fuzzy formal concept $C_1 = (\varphi(A_1), B_1)$ and its subconcept $C_2 = (\varphi(A_2), B_2)$ is defined as:

$$S(C_1, C_2) = \frac{|\varphi(A_1) \cap \varphi(A_2)|}{|\varphi(A_1) \cup \varphi(A_2)|}$$

where $\cap$ and $\cup$ refer intersection and union operators on fuzzy sets, respectively.

In [10], we showed as these FFCA are very powerful as well in the interpretation of the results of the Fuzzy Clustering; in Optimization of the flexible query that in the definition of a summary of a Database.

Example: Let a relational database table presented by Table1 containing the list of AGE and SALARY of Employee. Table 2 presents the results of fuzzy clustering (using Fuzzy C-Means [11]) applied to *Age* and *Salary* attributes. For *Salary* attribute, fuzzy clustering generates three clusters (C1, C2 and C3). For *AGE* attribute, two clusters have been generated (C4 and C5). In our example, $\alpha - Cut$ (Salary) = 0.3 and $\alpha - Cut$ (Age) = 0.5, so, the Table 2 can be rewriting as show in Table 3. The corresponding fuzzy concept lattices of fuzzy context presented in Table 3, noted as TAH's are given by the line diagrams presented in the Figure 1 and 2.

Table 1 A relational database table

	SALARY	AGE
t1	800	30
t2	600	35
t3	400	26
t4	900	40
t5	1000	27
t6	500	30

Table 2 Fuzzy Conceptual Scales for age and salary attributes

	SALARY			AGE	
	C1	C2	C3	C4	C5
t1	0.1	0.5	0.4	0.5	0.5
t2	0.3	0.6	0.1	0.4	0.6
t3	0.7	0.2	0.1	0.7	0.3
t4	0.1	0.4	0.5	0.2	0.8
t5	-	0.5	0.5	0.6	0.4
t6	0.5	0.5	-	0.5	0.5

Table 3 Fuzzy Conceptual Scales for age and Salary attributes with $\alpha - Cut$

	SALARY			AGE	
	C1	C2	C3	C4	C5
t1	-	0.5	0.4	0.5	0.5
t2	0.3	0.6	-	-	0.6
t3	0.7	-	-	0.7	-
t4	-	0.4	0.5	-	0.8
t5	-	0.5	0.5	0.6	-
t6	0.5	0.5	-	0.5	0.5

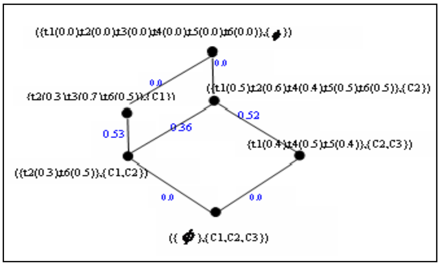


Fig. 1 Salary TAH

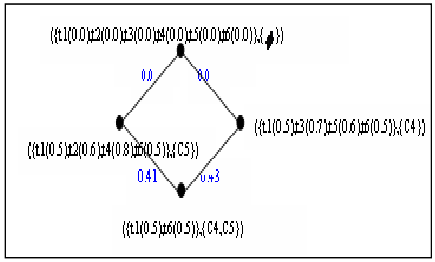


Fig. 2 Age TAH

3 Problems and Contributions

Generally the proposed approach in the KDD research area analysis the large databases, in order to identify valid, useful, meaningful, unknown, and unexpected relationships. Often these approaches are restricted to the application of discovery and modeling techniques within the KDD process. Thus, the goal to exploit these data is often neglected.

In our opinion, this problem was not really neglected but it was not sufficiently treated since the generated rules by these approaches, exceeding some times of thousand rules, are not easily exploitable [2], [3]. Indeed, this big number of rules is due to the fact that these approaches try to determine rules starting from the data or a data variety like the frequent item-sets or the frequent closed item-sets, which may be huge.

To remedy all these problems, we propose to define an ES allowing the end-user to easily exploit all generated knowledge. For the definition of the KB we propose a new KDD approach taking in consideration another degree of granularity into the process of knowledge extraction. We propose to define rules (Meta-Rules) between classes resulting from a preliminary classification on the data. Indeed while classifying data, we construct homogeneous groups of data having the same properties, so defining rules between clusters implies that all the data elements belonging to those clusters will be necessarily dependent on these same rules. Thus, the number of generated rules is smaller since one processes the extraction of the knowledge on the clusters which number is relatively lower compared to the initial data elements. We prove that we can easily deduce knowledge about the initial data set if we want more details.

4 Definition of an ES for Data Mining (ES-DM)

In this section, we present the definition of the ES for the Data Mining that has for goal to help the user to intelligent exploit these data. We, first of all, start with justifying the choice of the approach ES.

4.1 Why the Approach Expert System?

An ES is an intelligent program which is devoted to a specific domain of application where there is enough knowledge to infer one or several solutions, but not of precise or efficient algorithms which provide the same results. The power of an ES is characterized by the content of its KB and its capabilities to work as efficiently as possible.

Moreover, the choice of the ES for DM is justified by the following points:

1. The big volume of data to manage. The response to a request is not trivial. It needs an expertise to identify the relevant data at this request.
2. The update of the data (insert, delete or update of the data) is nowadays a very frequent operation in the Information systems. The taking into account of a new knowledge, or the modification of an existing knowledge, is a simple task in such a SE, it is enough to enrich or modify its KB.
3. Several approaches of KDD exist. Consequently, to topple an approach to another doesn't require a particular effect in the case of the ES.
4. The ES explains its reasoning and justifies its choice. This is very important for the user knowing that it starts from an enormous mass of data.

4.2 Objective of Expert System for Data Mining (ES-DM)

ES-DM has been designed as an intelligent tool for helping the user to give semantics of data, to seek information in this data and to satisfy his needs

The design of ES for DM must be directed by the following specific objectives:

- To constitute a KB composed of all useful concepts needed.
- To define an interactive methodological environment which permits to perform as far as possible the deduction steps with incomplete specifications and which permits to backtrack to any step in order to change some specifications or to integrate new information.

- To identify for each deduction step the general or specific principles of reasoning and to provide as detailed explanations as possible about these principles on which they are based.
- To build an open system of tools which enables, in the one hand, to integrate new theoretical concepts and deduction rules and, in the other hand, to transfer its expertise both via its usual use and via explanations and justifications of its results.
- To facilitate, interaction with the human designer by offering to him a semantically rich and easy to use interface.

4.3 Architecture of ES-DM

Like most ES, our ES for DM is composed of a Knowledge Base (KB), an Inference Engine (IE) and two interfaces: acquirement of the knowledge and user-interface.

The KB is composed of:

- The Rules base contains a set of rules. It models, in this case, the knowledge extracted of the initial data set.
- The Facts Base contains all the facts of the SE. It models, in this case, the initial data set data (after phase of data cleaning and Preprocessing and phase of Data Reduction)

The IE and two interfaces are supposed offered by the Expert system Shell which used for implemented this ES.

4.4 Definition of the User-Interface for ES-DM

A user doesn't have need that to permit him an easy exploitation and optimum of these data: (1) have an idea on the semantics of this large data set; (2) Seek of a data with an approximate response, therefore to model a flexible query.

According to the user's needs, we adopted four strategies of dialog:

First Strategy: The user specifies the number i of properties which he wants to check. The ES posts the list of the objects checking i properties.

Second Strategy: The user definite a property p, the ES posts the list of the properties which an object must check so that it can check the property p. The IE functions in this case in *Backward chaining.*

Third Strategy: The user defined a p property that is already satisfied by an object, the ES posts the list of property that can verify an object knowing that it verifies this property. The IE functions in this case in *Forward Chaining*

Fourth Strategy: The user can ask any queries for the ES. The IE functions in this case in mixed Chaining.

In the figure3, we summarize the interface of our ES.

5 Definition of the KB for our ES-DM

In this section, we present our new KDD approach for the definition of the KB.

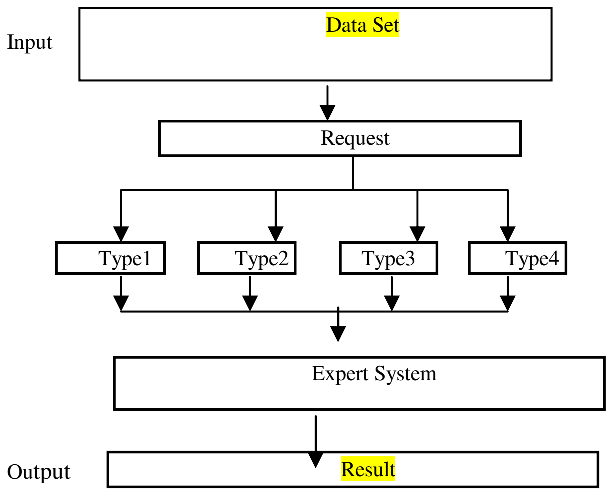


Fig. 3 Definition of the user-interface for ES-DM

5.1 Theoretical Foundation of Our Approach

In this part, we present the theoretical foundations of the proposed approach, based on the following properties:

Properties 1

- The number of clusters generated by a classification algorithm is always lower than the number of starting objects to which one applies the classification algorithm
- All objects belonging to one same cluster have the same proprieties. These characteristics can be deduced easily knowing the center and the distance from the cluster.
- The size of the lattice modeling the properties of the clusters is lower than the size of the lattice modeling the properties of the objects.
- The management of the lattice modeling the properties of the clusters is optimum than the management of the lattice modeling the properties of the objects.

Properties 2

Let C1, C2 be two clusters, generated by a classification algorithm and verifying the properties p1 and p2 respectively. Then the following properties are equivalent:

$$C1 \Rightarrow C2 \quad (CR) \quad \Leftrightarrow$$

- $\forall$ object O1 $\in$ C1 => O1 $\in$C2 (CR)
- $\forall$ object O1 $\in$ C1, O1 checks the property p1 of C1 and the property p2 of C2 with (CR)

Properties 3

Let C1, C2 and C3 be three clusters generated by a classification algorithm and verifying the properties p1, p2 and p3 respectively. Then the following properties are equivalent:

C1, C2 = > C3 (CR) $\Leftrightarrow$

- $\forall$ object O1 $\in$ C1 $\cap$ C2 = > O1 object $\in$C3 (CR)
- $\forall$ object O1$\in$ C1 $\cap$ C2 then O1 checks the properties p1, p2 and p3 with (CR)

The proof of this properties rises owing to the fact that all objects which belong to a same cluster check necessarily the same property as their cluster.

5.2 Discovering Meta Knowledge

In the first, we give a certain number of clusters for each attribute. Each tuple has values in the interval [0,1] representing these membership degrees according the formed clusters. Linguistic labels, which are fuzzy partitions, will be attributed on attribute's domain. This first step consists of TAH's and MTAH generation of relieving attributes.

Example. Let a relational database table presented by Table 1 containing the list of AGE and SALARY of Employee. Table 2 presents the results of fuzzy clustering applied to *Age* and *Salary* attributes.

The minimal value (*respectively.* maximal) of each cluster corresponds on the lower (*respectively.* higher) interval terminal of the values of this last. Each cluster of a partition is labeled with a ***linguistic labels*** provided by the user or a domain expert.

For example, the fuzzy labels *young* and *adult* could belong to a partition built over the domain of the attribute *AGE*. Also, the fuzzy labels *low*, *Medium* and *High,* could belong to a partition built over the domain of the attribute *Salary*.

This very simple sorting procedure gives us for each many-valued attribute the distribution of the objects in the line diagram of the chosen fuzzy scale. Usually, we are interested in the interaction between two or more fuzzy many-valued attributes. This interaction can be visualized using the so-called fuzzy nested line diagrams. It is used for visualizing larger fuzzy concept lattices, and combining fuzzy conceptual scales on-line. Figure 4 shows the Fuzzy Clusters Lattice constructed from Figure 1 and 2.

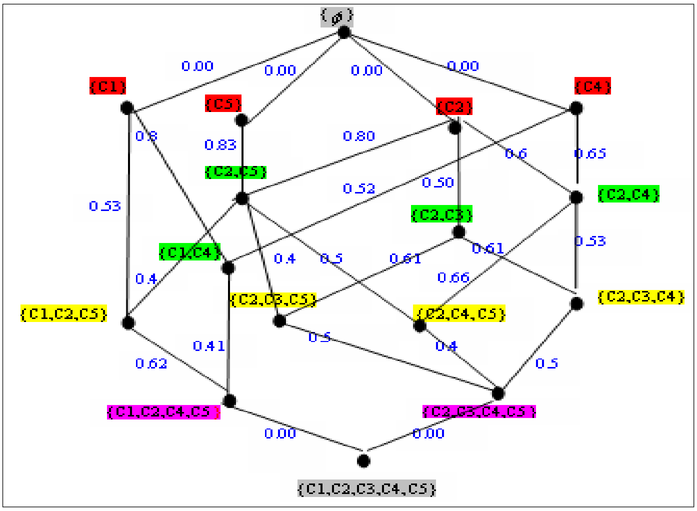


Fig. 4 The Fuzzy Clusters Lattice (FCL)

Considering the FCL in Figure 4, the derivation of fuzzy association meta-rules can be performed straightforwardly. Indeed, the meta-rule represent "inter-node" implications, assorted with the FC, between two adjacent comparable equivalence classes, i.e., from a set of clusters to another set of clusters immediately covering it. The confidence Factor will be equal to the weight of the arc binding the two nodes. Such an implication brings into participate two comparable equivalence classes, i.e. of a set of clusters towards another set of cluster including it in the partial order structure.

Example. The meta-rule C5 $\Rightarrow$ C2 (0,83), is generated starting from the two equivalence classes, whose their respective nodes are Clusters {C5}, {C2,C5} having as distance d=0.83.

Example. Let us consider the correspondence of the linguistic labels and their designations for the attributes Salary and Age (sub section 5.3).

The meta-rule C5=>C2 83% is transformed in Age(Adult)=>Salary(Medium) 83%

6 Validation of the Proposed Approach

To validate the approach proposed, we chose: 1) The FCM (Fuzzy C-Means) algorithm for a fuzzy classification of the data set, and 2) The *Ganter* algorithm for the construction of the lattice. We develop this approach with JAVA language. Concerning the ES we used the Expert system Shell '*JESS*' to implement it.

Our ES-DM present to the expert a main screen (Figure 5) similar to the standard usually used in the IDE (Integrated Development Environment).

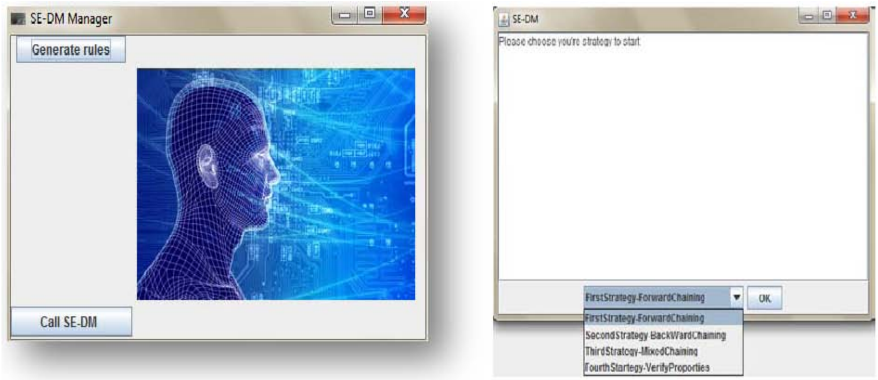


Fig. 5 Principal Window of the ES-DM

7 Conclusion

In this paper, we are interested of problem of exploitation of large data set which often neglected by the proposed approach in the area of data mining or KDD.

For this, we proposed to define an ES for Data mining enable to exploit these data with an intelligent approach. Like most ES, our ES-DM is composed of a Knowledge Base (KB), an Inference Engine (IE) and interfaces. The KB is composed of (1) *The Rules Base* models the knowledge extracted of the initial data set. the KB of our ES a new KDD approach taking in consideration another degree of granularity into the process of knowledge extraction. Indeed, we propose to define rules (Meta-Rules) between classes resulting from a preliminary classification on the data. Then, we automatically deduce knowledge about the initial data set. (2) The IE and two interfaces are supposed offered by the ESS which used for implemented this ES.

In the future, we propose to define an incremental method that permits to deduct the Knowledge Base generated by our model knowing the modifications carried out in the initial data base.

References

1. Goebel, M., Gruenwald, L.: A Survey of Data Mining and Knowledge Discovery Software Tools. In: SIGKDD, ACM SIGKDD, vol. 1(1), pp. 20–33 (June 1999)
2. Zaki, M.: Mining Non-Redundant Association Rules. Data Mining and Knowledge Discovery (9), 223–248 (2004)
3. Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L.: Intelligent Structuring and Reducing of Association Rules with Formal Concept Analysis. In: Baader, F., Brewka, G., Eiter, T. (eds.) KI 2001. LNCS (LNAI), vol. 2174, pp. 335–350. Springer, Heidelberg (2001)
4. Billard, L., Diday, E.: Symbolic Data Analysis: Conceptual Statistics and Data Mining. Wiley (2007)

5. Sato-Ilic, M.: Symbolic Clustering with Interval-Valued Data, Complex Adaptive Systems. Procedia Computer Sciences 6, 358–363 (2011)
6. Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules between sets of items in large Databases. In: Proceedings of the ACM SIGMOD Intl. Conference on Management of Data, Washington, USA, pp. 207–216 (June 1993)
7. Agrawal, R., Skirant, R.: Fast algoritms for mining association rules. In: Proceedings of the 20th Int'l Conference on Very Large Databases, pp. 478–499 (June 1994)
8. Ganter, B., Wille, R.: Formal Concept Analysis: mathematical foundations (translated from the German by Cornelia Franzke). Springer, Heidelberg (1999)
9. Thanh, T., Siu Cheung, H., Tru Hoang, C.: A Fuzzy FCA-based Approach to Conceptual Clustering for Automatic Generation of Concept Hierarchy on Uncertainty Data. In: CLA 2004, pp. 1–12 (2004) ISBN 80-248-0597-9
10. Grissa Touzi, A., Sassi, M., Ounelli, H.: An innovative contribution to flexible query through the fusion of conceptual clustering, fuzzy logic, and formal concept analysis. International Journal of Computers and Their Applications 16(4), 220–233 (2009)
11. Sun, H., Wang, S., Jiang, Q.: FCM-Based Model Selection Algorithms for Determining the Number of Clusters. Pattern Recognition 37, 2027–2037 (2004)

Evolution of Friendship Networks and Transition of Their Properties

Hiroshi Matsushima and Nobuhiro Inuzuka

Abstract. We analyze friendship networks using attendance records of students to classes. Acquire and analysis of networks are time-consuming. Hence analysis for the network evolution according time passed is highly cost. We have proposed an automated method to acquire friendship networks for students using a system which is becoming popular recently in universities. The method gives a score which measures a degree of friendship between two students based on the probability they are friends. By using these networks, we analyze how friend pairs evolve and how common properties of social networks, scale-free, cluster and small-world, change.

1 Introduction

Since the 1970s, the social network analysis (SNA) has been actively researched in sociology, and recently it gathers attention from computer scientists as well. SNA pays attention not to attributes of entities but relations among entities. For personal relationship, inter-business relationship, internation relationship and WWW, researchers have been studied how network forms and evolves, who has a primary role, what factions or communities exist and what structure exists behind.

In order to begin these investigations we first acquire network data, that is, we need qualitative information among entities. Usually for this purpose we use a questionnaire and it is always time-consuming. Accordingly study on formation and evolution of networks are difficult indeed. An automated method to collect network data is desired and is useful for these studies.

Inuzuka et al.[1] introduced an automated method to collect data for friendship networks using a system which collects attendance records to classes in school. The system or similar systems are becoming popular these

Hiroshi Matsushima · Nobuhiro Inuzuka
Nagoya Institute of Technology,
Gokiso-cho Showa, Nagoya 466-8555, Japan
e-mail: matsuhima@nous.nitech.ac.jp, inuzuka@nitech.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 303–312.
springerlink.com

days. It consists of student cards with IC-chips that all students keep as their IDs and card readers equipped in all lecture rooms. For each class students check their attendance with the cards at a card reader in the room. Inuzuka et al.[1] showed that the attendance records can be used to predict the friendship relation among students.

The purpose of this paper is to show that the method of friendship is useful to detect and analyze friendship network and their transitions. In this paper, first we introduce the system we used which collects date automatically. Section 3 reviews the method of friendship score using attendance records in university lecture class and generation of friendship network among students. Then we give analysis of temporal transition of network formation in Section 4. In Section 5 we see the common properties of social networks, that is, properties on scale-free, high cluster coefficient and small-world.

2 Class Attendance Record Management System and Friendship Relation

Nagoya Institute of Technology (NIT) installed a system to collect and manage class attendance records of students, which we call CARMS (Class Attendance Record Management System), in 2007. It aims to reduce tasks of instructors by collecting records automatically. The CARMS system consists of student ID-cards, card readers and a database management system (DBMS). A student ID card has a function of a wireless tag and keeps the information of the student ID of a card holder. A card reader has its own ID (reader-ID) and reads the information of an ID card when a holder places his/her card in the front of the reader. Each lecture room equips two or three readers near the entrances. Readers send the information as a tuple (SID, RID, ALT) to DBMS, where SID is a student ID, RID is a reader ID, and ALT is an attendance/leaving time, that is the time when the student puts close his/her card to the reader at his/her attendance to or leaving from a class. The DBMS collects and keeps all the information. CARMS gives lists of students who attended a class on a specified date. Although CARMS has also many other functions, we omit to describe the detail.

Table 1 shows the basic data of the system and records collected and used for our experiments. Student IDs for fourth year students appeared in records and their attendance records are fewer than other years. This is because the most of fourth year students participate only graduate research projects or take a small number of lectures. Graduate course students are also in the similar situations. Accordingly we used only the data of 1st, 2nd and 3rd year undergraduate students.

Table 1 Summary of records collected by CARMS system

The period of records	2007.10.1 – 2008.3.31
# students recorded	4,403
of which first year students	942
of which second year students	936
of which third year students	929
of which fourth year students	287
# readers equipped	129
# records	864,882
of which first year students	295,700
of which second year students	242,001
of which third year students	165,588
of which fourth year students	18,463

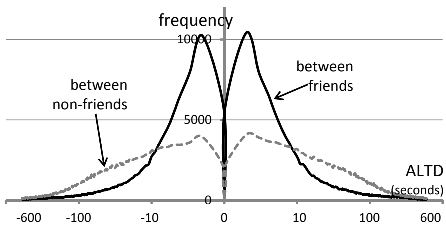


Fig. 1 Frequency distribution of ALTD

3 Friendship Scores Based on Attendance Records

Inuzuka et al.[1] utilized the character of frequency distributions of time differences of ALT (ALTD) between the each friend pair and non-friend pair(Fig. 1). The distribution of ALTD between friend pairs and one for non-friend pairs have a remarkable difference. Fig. 1 shows the distributions. The ALTD distributions have peeks at a small time length. Ones for friend pairs the peeks becomes acute.

Using the character of distributions a prediction method can be given. We take two students A and B. Let f is the event that A and B are friends each other, and $T = \{t_1, t_2, ..., t_n\}$ is a set of ALTD data between A and B. Here we assume that all t_i $(i = 1, \ldots, n)$ are independent. Then, conditional probability $p(f \mid T)$ can be written as follows,

$$p(f \mid T) = \frac{p(f) \cdot p(T \mid f)}{p(T)} = p(f) \prod_{t \in T} \frac{p(t \mid f)}{p(t)}. \tag{1}$$

We move to think of the ratio, r_t, of ALTD records which are of time length t and are given by friend pairs against all ALTD records of t, that is, represented as follows,

$$\begin{aligned} r_t &= \frac{\#(\text{all ALTD records of } t \text{ between friend pairs})}{\#(\text{all ALTD records of } t \text{ between all students pairs})} \\ &= \frac{X_f \cdot m_f \cdot p(t \mid f)}{X \cdot m \cdot p(t)}, \end{aligned} \tag{2}$$

where X is the number of all of students pairs, X_f is the number of all friend pairs, m is the expected number of ALTD records produced among randomly chosen two students, m_f is the expected number of ALTD records among friend pairs, $p(t)$ is the probability that randomly chosen two students have ALTD t, and $p(t|f)$ is the probability that randomly chosen friend pair have ALTD t. Then, we can reformulate $p(t \mid f)$ as follows.

$$p(t|f) = \frac{X \cdot m \cdot p(t) \cdot r_t}{X_f \cdot m_f} = \frac{m \cdot p(t) \cdot r_t}{p(f) \cdot m_f} \tag{3}$$

When we substitute this to Equation (1) we have the following equation.

$$p(f|T) = p(f) \prod_{t \in T} \frac{m \cdot r_t}{m_f \cdot p(f)} = p(f)^{(1-n)} \left(\frac{m}{m_f}\right)^n \prod_{t \in T} r_{tf}, \tag{4}$$

where $n =\mid T \mid$. When $\bar{f}$ denotes the event that two student are not friends. Then, $p(\bar{f} \mid T)$ is following.

$$p(\bar{f} \mid T) = p(\bar{f})^{(1-n)} \left(\frac{m}{m_0}\right)^n \prod_{t \in T} (1 - r_t), \tag{5}$$

where m_0 is the expected number of ALTD records produced non-friend pair.

Then we give the friendship score by the logit of $p(f \mid T)$ as follows.

$$\begin{aligned} \operatorname{logit} p(f|T) &= \log \left(\frac{p(f|T)}{1 - p(f|T)} \right) \\ &= \log(p(f|T)) - \log(p(\bar{f}|T)) \end{aligned} \tag{6}$$

In [1] the recall and precision of the friend estimation using friendship score are reported as around 70% compared with questionnaire. When considering the difficulties to take questionnaire the prediction is useful.

Table 2 Recall and Precision of each classes

@@@	class K	class J	class P
Recall	75.3	86.9	65.7
Precision	60.0	88.9	78.9

4 Friendship Networks

Using the friendship score method friendship networks can be generated and then we may analyze some properties among students and their friendship networks. The properties include basic information such as the number of friend pairs, and network formation and also the common properties which characterize such networks.

4.1 Basic Information of Friendship Networks

Friendship scores for freshman students in fiscal 2007 are calculated and we made their friendship networks. In order to evaluate clustering-coefficient, weights of links, i.e. friendship scores, are normalized into the range of 0 to 1. We gave weights by applying the sigmoid function to friendship scores. Fig. 2 illustrates a part of a friendship network generated. In addition, table3 summarizes the number of all students, the number of all student pairs between which there exist at least an ALTD datum less than ten minutes, which means they are potential candidates for friendship, and the number of friend pairs presumed by the friendship score.

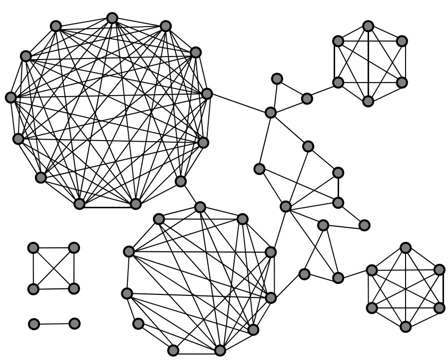
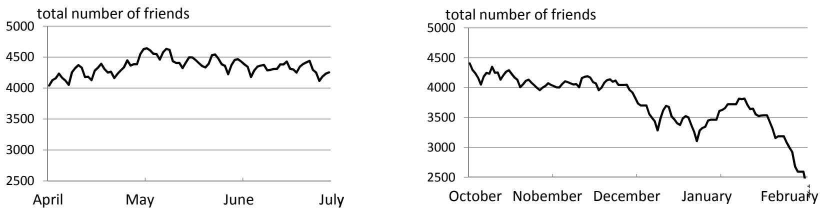


Fig. 2 A part of friendship network generated from the method

Table 3 Basic information of friendship network

term	number
Number of students	948
Number of student pairs	579552
Number of friend pair	5972



(a) First semester (b) Second semester

Fig. 3 A transition of number of friend pairs

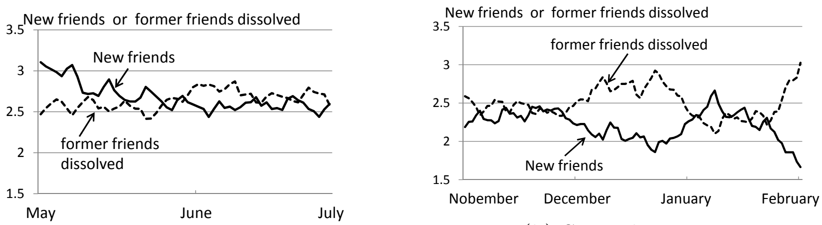
4.2 Analysis of Network Evolution

In order to see possibility of our method for study of evolution of friendship networks, we try to observe transition of network properties. First, we observes the transition of the number of links, the number of friendship relations in the networks.

For our purpose we generate networks from a fixed shorter period of the year and put off the period by a day. We may understand a network generated from data of a period as an averaged network along the period. By putting off the period used we can generated series of networks which may represent the evolution of the friendship.

In preliminary experiments we know that a sample of three weeks is enough to generate a networks. Accordingly we generated a series of networks from data of a month (30 days) and shift the period by a day. Fig. 3 shows the transition of the number of links in the series of networks. In NIT the first semester starts 1st April and ends 31st July. Accordingly the first period is from 1st April to 30th April and the last period is from 2nd July to 31st July. The x-axis shows the start date of the periods. In the first semester, there is no substantial change but friend pairs increase by around 600. Afterward, it is stable in both first and second semester until December. December and the following month include winter holidays and collection ALDT data are disturbed. The changes in the second half of the second semester may be caused from these reasons, although we need more investigation.

Using the same network series, we analyze increase and decrease of links according to the period. That is, we observed the number of newly made friend pairs and friendship disappeared compared to a networks of a fixed period, say a month, before. In Fig. 4 (a) shows the result for the first semester



(a) First semester

(b) Second semester

Fig. 4 Transitions of newly generated and dissolved friendships

and (b) shows one for the second semester. In the both graph, the solid line shows the number of new friends, that is, friendships which exist in the network after one month but do not exist before. The broken solid line shows the number of friends dissolved, that is, friendships which do not exist in the network after one month but exist before.

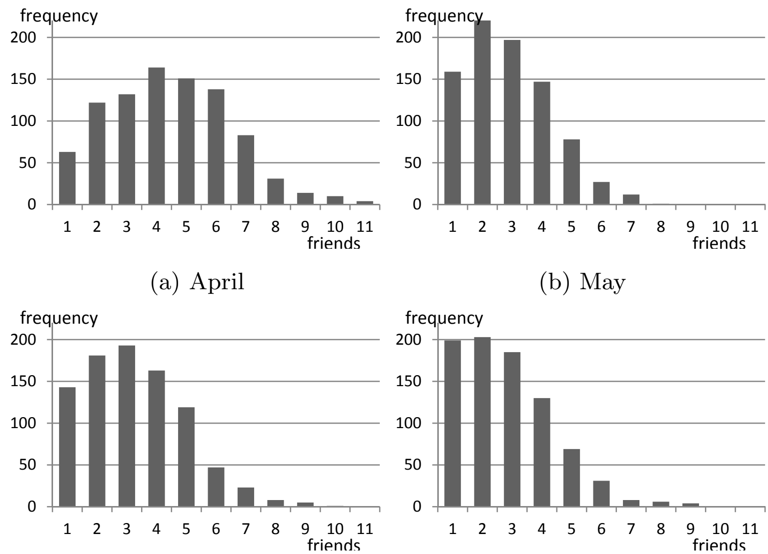
In early stage of first semester, newly generated friendships are larger than dissolved ones, so the number of friend pair increased in early stage of first semester. Afterward, although friendship keep changed, both of them were stable from 1.5 by 1.7. Thus, we can consider that a strength of friendship is dynamic. Of course we need carefully verify an effect of accuracy of friendship score.

5 Transitions of Common Properties of Social Network

By the previous literature, it has been clear that social networks have common properties, scale-free[2], strong cluster and small-world[3]. Scale-free shows that very few people have large number of friends but others have a few number of friends, and degree k's distribution $p(k)$ obeys $p(k) \propto k^{\gamma} (\gamma < 0)$, i.e. the power law. Strong cluster means that friends of a person are likely friend mutually. In order to measure the strength of cluster, we see the clustering-coefficient. Small-world means that everyone can reach an destination in relatively small steps, and means that average shortest path distances among nodes are short. Here we analyze transition of these properties.

5.1 Transitions of Scale-Free Property

Fig. 5 illustrates the degree distribution of the friendship networks. The degree of a node is the number of neighboring nodes of the node, that is, the number of friends in friendship networks.



(a) April (b) May

(c) June (d) July

Fig. 5 Degree distributions of friendship networks in the first semester

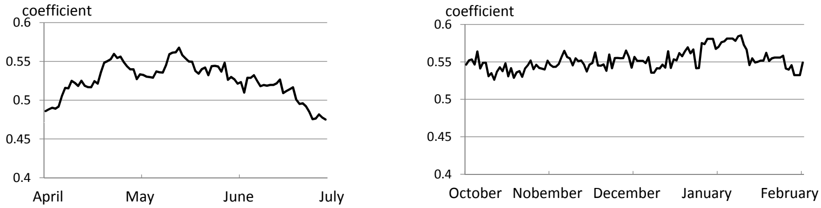
In April, the number of student who have four friends is the highest and the distribution may not obey power law. However, as time advances, the number of students who have less than three friends increases, while the number of students who have more than four friends decreases. The degree distribution in July may acquire the power law. Afterward, degree distributions become stable in second semester. Thus, friendship networks get scale-free property in around four months.

5.2 Transitions of Cluster

In order to analyze cluster, we use cluster-coefficient for weighted and directed graphs[6] Let $G = (V, E, c)$is a weighted and directed graph, where V is a set of vertexes, E is a set of edges and c is a weight function. The cluster-coefficient C of G can be calculated as follows[6].

$$C = avg_{v_i \in V}(C(v_i))$$

$$C(v_i) = \frac{1}{{}_{k|E(v_i)|}\mathrm{P}_2} \sum_{v_j \neq v_k \in E(v_i)} \frac{c(v_i, v_j) \cdot c(v_j, v_k) + c(v_i, v_k) \cdot c(v_k, v_j)}{c(v_i, v_j) + c(v_i, v_k)}$$



(a) First semester (b) Second semester

Fig. 6 Transitions of cluster-coefficient

Fig. 6 shows transitions of C calculated by the networks. In first semester, cluster-coefficient increases from April to early June but decreases afterward. In second semester, it is stable around 0.55 except span affected by vacation. In a whole year, about 0.06 increases. A popular SNS networks that is generally known that cluster is high, has its cluster coefficient about 0.36. Comparing it the friendship networks have remarkably high cluster coefficient. However, it may be affected by accuracy of friendship estimation by friendship score and we need to verify it.

5.3 Transitions of Small-World

In order to analyze small-world, we calculate average shortest path distances of friendship networks. At first, we divide the network of each strongly connected components. A strongly connected component is a subnetwork in which every node has a directed path to all other nodes. When a network has isolated parts the average of shortest path distance is not defined. However when a network include a big component which include almost all nodes, the average of shortest path distances should be defined for the components.

In general friendship networks consist of some small components that contain only a node, i.e. students, or a few nodes and a unique large component that contain almost all nodes. We observes the shortest path distances for the unique large components.

Table 4 shows the number N, of students contained in the unique largest components of each month, and the average shortest path distances, ASPD, of networks.

Each average shortest path distances does not change much more than properties and is around six. This is the almost same result of Milgram's experiment in the 1960's[4]. Also, in December and January, average shortest path distances are longer and number of vetexes are lower than others. This is because number of ALT date in December and January are lower than others due to long vacations and this reflects accuracy of friendship estimation of friendship score. The meaning of these value need to be verified.

Table 4 Shortest average path lengths of the networks

	April	May	June	July	Oct.	Nov.	Dec.	Jan.
N	895	909	898	885	892	866	822	802
ASPD	6.271	5.982	5.831	5.992	6.179	6.477	7.120	7.395

6 Discussions

In this paper, we made friendship networks using friendship score and analyzed transitions of network formation, transitions of three common properties of social networks and students' roles comparing questionnaires with students whose centralities are high.

A number of friendship links among students increased by end of May and network stabilized in early stage. Because strength of friendship consistently varies, friendship activities are dynamic. By analysis of transitions of three common properties, we find tendencies that degree distribution comes to acquire the power law in first semester, cluster coefficient increases in early stage and decreases before it stabilizes and shortest average path lengths do not change but are around six.

By analysis of students' role using centralities, using centralities on degrees, such as Degree and Pagerank, is predictably-effective to analyze roles on authorities.

Therefore, friendship network analysis using friendship score reflects well-known common properties of social networks and is useful to analyze transitions of friendship network and students' role.

Further interesting research topics include more detailed analysis of results, relation between friendship estimation and accuracy of friendship score and detection of new transitive properties. Also, we need to verify whether we can get the same transitional results from other networks. Moreover, it concludes suggestion of new graph models hold these properties.

References

1. Inuzuka, N., Nakano, T., Shimomura, K.: Friendship Analysis Using Attendance Records to University Lecture Classes. In: IASK International Conference Teaching and Learning, pp. 478–486 (2008)
2. Inuzuka, N., Kondo, T., Yamamoto, S.: Analysis of Asymmetric Friendship Among Students from Class Attendance Records. In: Nakamatsu, K., Phillips-Wren, G., Jain, L.C., Howlett, R.J. (eds.) New Advances in Intelligent Decision Technologies. SCI, vol. 199, pp. 393–403. Springer, Heidelberg (2009)
3. Albert, R., Barabasí, A.-L.: Statical mechanics of complex networks. Review of Modern Physics 74, 47–97 (2002)
4. Milgram, S.: The small world problem. Phycology Today 1, 67 (1967)
5. Watts, D.J.: Small Worlds. Princeton University Press (1999)
6. Suzuki, T.: Analysis for Directed and Weighted Complex Networks. JWEI07 (2006)

Feedback GMDH-Type Neural Network Algorithm Using Prediction Error Criterion Defined as AIC

Tadashi Kondo, Junji Ueno, and Shoichiro Takao

Abstract. In this study, a feedback Group Method of Data Handling (GMDH)-type neural network algorithm using prediction error criterion defined as AIC, is proposed. In this algorithm, the optimum neural network architecture is automatically selected from three types of neural network architectures such as sigmoid function type neural network, radial basis function (RBF) type neural network and polynomial type neural network. Furthermore, the structural parameters such as the number of feedback loops, the number of neurons in the hidden layers and useful input variables are automatically selected so as to minimize the prediction error criterion defined as Akaike's Information Criterion (AIC). Feedback GMDH-type neural network has a feedback loop and the complexity of the neural network increases gradually using feedback loop calculations so as to fit the complexity of the nonlinear system. This algorithm is applied to identification problem of the complex nonlinear system.

Keywords: Neural network, GMDH, AIC, Self-organization.

1 Introduction

The GMDH-type neural networks and their applications have been proposed in our early works [1],[2]. The GMDH-type neural networks can automatically organize the neural network architecture using the heuristic self-organization method which is a kind of evolutional computations and basic theory of the GMDH algorism [3]-[5]. The GMDH-type neural networks can also determine structural parameters such as the number of layers, the number of neurons in the hidden layers and the useful input variables. In the GMDH-type neural networks, the neural network architecture is organized so as to minimize the prediction error criterion defined as Akaike's Information Criterion (AIC) [6]. The conventional neural networks trained using the back propagation algorithm, can not automatically organize neural network architecture.

Tadashi Kondo · Junji Ueno · Shoichiro Takao
Graduate School of Health Sciences, The University of Tokushima,
3-18-15 Kuramoto-cho Tokushima, 770-8509, Japan

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 313–322.
springerlink.com

In this study, the feedback GMDH-type neural network algorithm using prediction error criterion defined as AIC is proposed. In this algorithm, the optimum neural network architecture is automatically selected from three types of neural network architectures such as the sigmoid function type neural network, the radial basis function (RBF) type neural network and the polynomial type neural network. Furthermore, the structural parameters such as the number of feedback loops, the number of neurons in the hidden layers and the useful input variables are automatically selected so as to minimize the prediction error criterion defined as AIC. The feedback GMDH-type neural network has a feedback loop and the complexity of the neural network increases gradually using feedback loop calculations so as to fit the complexity of the nonlinear system.

The feedback GMDH-type neural network algorithm is applied to the identification problem of the complex nonlinear system. The optimum neural network architecture fitting the complexity of the nonlinear system is selected from three types of the neural network architectures. The identification results of the feedback GMDH-type neural network are compared with those of the GMDH algorithm and the conventional multi-layered neural network trained using the back propagation algorithm. It is shown that the feedback GMDH-type neural network is a very useful identification method of the complex nonlinear system because the optimum neural network architecture is automatically organized so as to minimize the prediction error criterion defined as AIC.

2 Feedback GMDH-Type Neural Network Algorithm

The architecture of the feedback GMDH-type neural network proposed in this paper, has a feedback loop as shown in Fig.1.

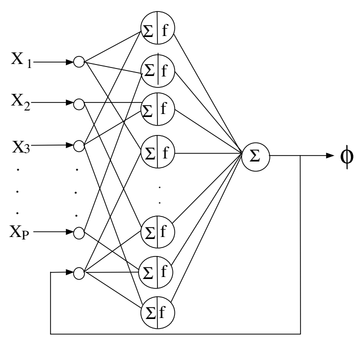


Fig. 1 Architecture of feedback GMDH-type neural network

The feedback GMDH-type neural network algorithm can select the optimum neural network architecture from three types of neural network architectures such as the sigmoid function type neural network, the RBF type neural network and the polynomial type neural network. The feedback GMDH-type neural network algorithm uses three types of neuron architectures such as the sigmoid function type neuron, the RBF type neuron and the polynomial type neuron. For each type of neuron architecture, we use two types of neurons called as the first and the second type neuron. In the feedback GMDH-type neural network, optimum neuron architectures fitting the characteristics of the nonlinear system are automatically selected using AIC.

2.1 First Loop Calculation

First, all data are set to the training data. In this algorithm, it is not necessary to separate the original data into the training and test sets because AIC can be used for organizing the network architectures.

2.1.1 Input Layer

$$u_j=x_j \qquad (j=1,2,\ldots,p) \tag{1}$$

where x_j $(j=1,2,\ldots,p)$ are the input variables of the system, and p is the number of input variables. In the first layer, input variables are set to the output variables.

2.1.2 Hidden Layer

All combinations of the r input variables are generated. For each combination, three types of neuron architectures which are the sigmoid function type neuron, the RBF type neuron and the polynomial type neuron, are generated and L neurons which minimize AIC values are selected for each type of neuron architectures.

Furthermore, for each combination, optimum neuron architectures fitting the characteristics of the nonlinear system are automatically selected using AIC.

a) Sigmoid function type neuron:

i) The first type neuron

Σ: (Nonlinear function)

$$z_k= w_1u_i+w_2u_j+w_3u_iu_j+w_4u_i^2+w_5u_j^2+w_6u_i^3+w_7u_i^2u_j+w_8u_iu_j^2+w_9u_j^3-w_0\theta_1 \tag{2}$$

f : (Nonlinear function)

$$y_k=\frac{1}{1+e^{(-z_k)}} \tag{3}$$

ii) The second type neuron

Σ: (Linear function)

$$z_k= w_1u_1+w_2u_2+w_3u_3+\ldots+w_ru_r-w_0\theta_1 \quad (r<p) \tag{4}$$

f : (Nonlinear function)

$$y_k=\frac{1}{1+e^{(-z_k)}} \tag{5}$$

b) RBF type neuron:

i) The first type neuron

Σ: (Nonlinear function)

$$z_k= w_1u_i+w_2u_j+w_3u_iu_j+w_4u_i^2+w_5u_j^2+w_6u_i^3+w_7u_i^2u_j+w_8u_iu_j^2+w_9u_j^3-w_0\theta_1 \tag{6}$$

f : (Nonlinear function)

$$y_k = e^{(-z_k^2)} \tag{7}$$

ii) The second type neuron

Σ: (Linear function)

$$z_k= w_1u_1+w_2u_2+w_3u_3+\ldots+w_ru_r-w_0\theta_1 \quad (r<p) \tag{8}$$

f : (Nonlinear function)

$$y_k = e^{(-z_k^2)} \tag{9}$$

c) Polynomial type neuron:

i) The first type neuron

Σ: (Nonlinear function)

$$z_k= w_1u_i+w_2u_j+w_3u_iu_j+w_4u_i^2+w_5u_j^2+w_6u_i^3+w_7u_i^2u_j+w_8u_iu_j^2+w_9u_j^3-w_0\theta_1 \tag{10}$$

f : (Linear function)

$$y_k= z_k \tag{11}$$

ii) The second type neuron

Σ: (Linear function)

$$z_k= w_1u_1+w_2u_2+w_3u_3+\ldots+w_ru_r-w_0\theta_1 \quad (r<p) \tag{12}$$

f : (Linear function)

$$y_k= z_k \tag{13}$$

In the first type neuron, $\theta_1=1$ and w_i (i=0,1,2,…,9) are the weights between the first and second layer and estimated by applying the stepwise regression analysis [7] to the training data. Only useful input variables u_i (i=1,2, …) are selected using AIC. The value of r, which is the number of input variables u in each neuron, is set to two for the first type neuron. The output variables y_k of the neurons are called as the intermediate variables.

In the second type neuron, $\theta_1=1$ and w_i (i=0,1,2,…,r) are the weights between the first and second layer and estimated by applying the stepwise regression analysis [7] to the training data. Only useful input variables u_i (i=1,2, …) are selected using AIC. The value of r, which is the number of input variables u in each neuron, is set to be greater than two and smaller than p for the second type neuron. Here p is the number of input variables x_i (i=1,2,…,p). The output variables y_k of the neurons are called as the intermediate variables.

The weights w_i (i=0,1,2,…) in each neuron are estimated by the stepwise regression analysis using AIC.

Estimation procedure of the weight w_i:
First, the values of z_k are calculated for each neuron architecture as follows.

i) Sigmoid function type neuron:

$$z_k = \log_e\left(\frac{\phi'}{1-\phi'}\right) \tag{14}$$

ii) RBF type neuron:

$$z_k = \sqrt{-\log_e \phi'} \tag{15}$$

iii) Polynomial type neuron:

$$z_k=\phi \tag{16}$$

where ϕ' is the normalized output variable whose values are between zero and one and ϕ is the output variable.

Then weights w_i are estimated by the stepwise regression analysis [7] which selects useful input variables using AIC. Only useful variables in Eq.(2), Eq.(4), Eq.(6), Eq.(8), Eq.(10) and Eq.(12) are selected by the stepwise regression analysis using AIC and optimum neuron architectures are organized by selected useful variables.

L neurons having the smallest AIC values are selected for three types of neuron architectures which are the sigmoid function type neuron, the RBF type neuron and the polynomial type neuron. The output variables y_k of L selected neurons for three types of neuron architectures, are set to the input variables of the neurons in the output layer.

2.1.3 Output Layer

For three types of neural network, the outputs y_k of the neurons in the hidden layer are combined by the following linear function.

$$\phi^* = a_0 + \sum_{k=1}^{L} a_k y_k \tag{17}$$

Here, L is the number of combinations of the input variables and y_k is the intermediate variables. The useful intermediate variables y_k are selected by using the stepwise regression analysis in which AIC is used as the variable selection criterion.

Values of AIC are calculated for three types of neural network architectures which are the sigmoid function type neural network, the RBF type neural network and the polynomial type neural network. Then, the neural network architecture which has smallest AIC value, is selected as the GMDH-type neural network architecture from three types of neural network architectures

Then, the estimated output values ϕ^* which is selected in the output layer, is used as the feedback value and it is combined with the input variables in the next loop calculation.

2.2 Second and Subsequent Loop Calculations

The optimum neural network architecture is selected from three types of neural network architectures in the output layer. Therefore, in the second and subsequent loop calculations, only one type of neuron architecture, which is the sigmoid function type neuron or the RBF type neuron or the polynomial type neuron, is used for the calculation.

First, the estimated output value ϕ^* is combined with the input variables and all combinations between the estimated output value ϕ^* and the input variables, are generated. The same calculation as the first feedback loop is carried out for each combination. Here, only one type of neuron architecture, which is selected in the first loop calculation, is used in the calculation. When AIC value of the linear function in (17) is increased, the loop calculation is terminated and the complete neural network architecture is organized using the L selected neurons in each feedback loop.

By using these procedures, the feedback GMDH-type neural network using prediction error criterion defined as AIC, can be organized.

3 Application to Nonlinear System Identification

The nonlinear system is assumed to be described by the following equations:

$$\phi_1 = (1.0+1.1x_1+1.2x_2+1.3x_3)^4 + \varepsilon_1 \tag{18}$$

$$\phi_2 = (1.0+1.4x_1+1.5x_2+1.6x_3)^4 + \varepsilon_2 \tag{19}$$

$$\phi_3 = (1.0+1.7x_1+1.8x_2+1.9x_3)^4 + \varepsilon_3 \tag{20}$$

$$\phi_4 = (1.0+2.0x_1+2.1x_2+2.2x_3)^4 + \varepsilon_4 \tag{21}$$

Here, $\phi_1 \sim \phi_4$ show output variables and $x_1 \sim x_3$ show input variables. $\varepsilon_1 \sim \varepsilon_4$ show noises. Furthermore, x_4 is added as the input variable of the neural network in order to check that the feedback GMDH-type neural network can eliminate the useless input variables. The neural network is organized using twenty training data. The prediction is carried out using twenty testing data so as to check the generalization ability.

3.1 Identification Results Obtained by Using the Feedback GMDH-Type Neural Network

(1) Input variables

Four input variables were used but the useless input variables x_4 was automatically eliminated.

(2) Number of selected neurons

Four neurons were selected in the hidden layer.

(3) Selection of the neural network architecture

Figure 2 shows AIC values of three kinds of neurons in the first feedback loop calculation. Polynomial type neuron had the smallest AIC value and polynomial type neural network architecture was selected as the feedback GMDH-type neural network architecture.

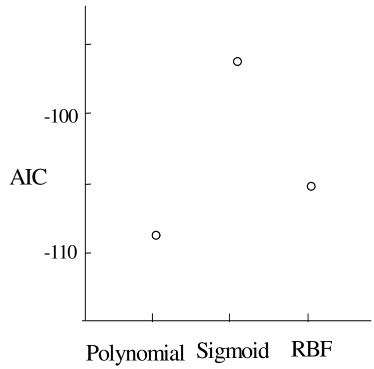


Fig. 2 AIC values of three kinds of neurons

(4) Structure of the neural network

The calculation of the GMDH-type neural network was terminated at the seventh feedback loop calculation.

(5) Estimation accuracy

The estimation accuracy was evaluated using the following equation:

$$J_1 = \frac{\sum_{i=1}^{20} |\phi_i - \phi_i^*|}{\sum_{i=1}^{20} |\phi_i|} \tag{22}$$

where ϕ_i (i =1,2,···,20) were the actual values and ϕ_i^* (i =1,2,···,20) were the estimated values by the feedback GMDH-type neural network. The values of J_1 for four output variables are shown in Table 1.

(6) The prediction accuracy

The prediction accuracy was evaluated using the following equation:

$$J_2 = \frac{\sum_{i=21}^{40} |\phi_i - \phi_i^*|}{\sum_{i=21}^{40} |\phi_i|} \tag{23}$$

where ϕ_i (i =21,22,···,40) were the actual values and ϕ_i^* (i =21,22,···,40) were the predicted values by the feedback GMDH-type neural network. The values of J_2 for four output variables are shown in Table 1.

(7) Variation of AIC and estimated values

The variation of AIC in the output variables ϕ_1 is shown in Fig.3. It decreased gradually by the feedback loop calculations and converged at the seventh feedback loop calculation. The estimated values of ϕ_1 by the GMDH-type neural network is shown in Fig.4. We can see that the estimated values are very accurate.

Table 1 Prediction and estimation accuracy

Models	J	ϕ1	ϕ2	ϕ3	ϕ4
GMDH-NN	J1	0.013	0.022	0.023	0.024
	J2	0.025	0.028	0.029	0.029
GMDH	J1	0.056	0.058	0.038	0.039
	J2	0.055	0.058	0.044	0.044
NN	J1	0.119	0.133	0.108	0.11
	J2	0.114	0.133	0.102	0.109

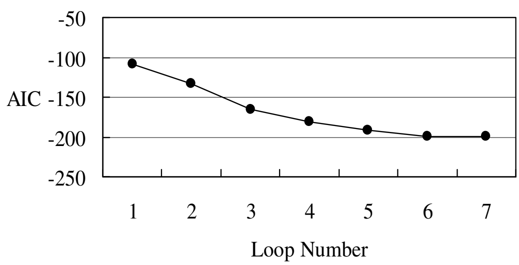


Fig. 3 Variation of AIC

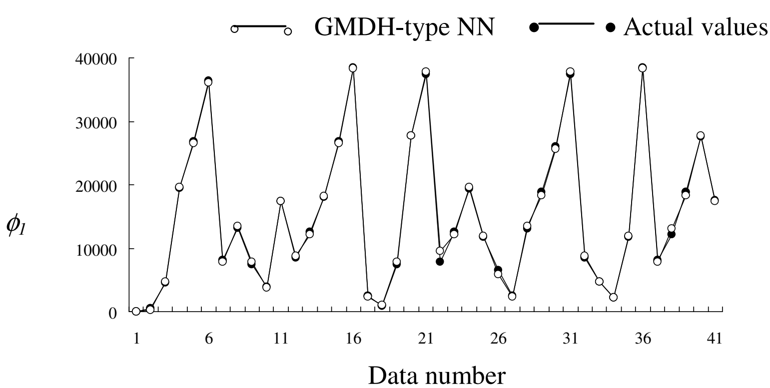


Fig. 4 Estimated values of ϕ_1 by feedback GMDH-type neural network

3.2 Comparison of the Feedback GMDH-Type Neural Network and Other Models

Identification results were compared with those of GMDH and conventional multilayered neural network trained using back propagation algorithm.

3.2.1 GMDH Algorithm

Four input variables were used but the useless input variable x_4 was automatically eliminated. Four intermediate variables were selected. The calculation was terminated at the fourth layer. The values of $J_1$and J_2 are shown in Table 1.

3.2.2 Conventional Multilayered Neural Network

The neural network had three layered structures. Four input variables were used in the input layer and eight neurons were used in the hidden layer. The estimated values of ϕ_1 by the conventional neural network are shown in Fig.5. The values of J_1 and J_2 are shown in Table 1.

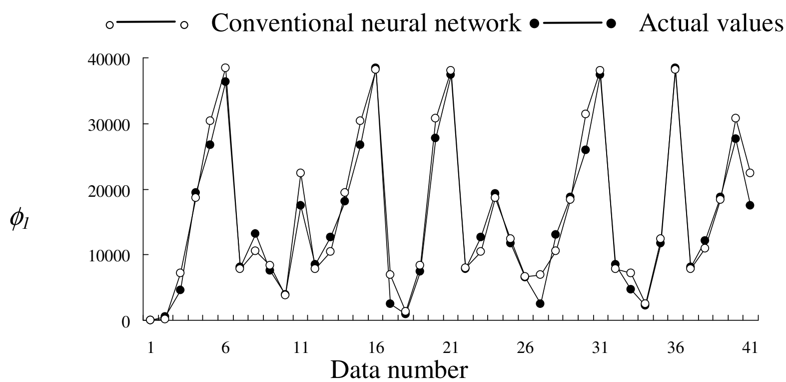


Fig. 5 Estimated values of ϕ_1 by the conventional neural network

4 Discussion

From the identification results, we can see the following:

(1) Both estimation and prediction errors of the feedback GMDH-type neural network were smallest in the three identified models. We can see that the feedback GMDH-type neural network was a very accurate identification method for the nonlinear system.

(2) In the feedback GMDH-type neural network, AIC value at the first loop calculation was not small but it was gradually decreased by the feedback loop calculations. So we can see that the feedback loop calculation plays a very important role in the feedback GMDH-type neural network.

(3) In the conventional neural network, the effects of high order terms of the input variables are not considered. Furthermore, it does not have the ability of self-selecting useful input variables. So the accuracy of the neural network was not good.

The feedback GMDH-type neural network can organize the conventional neural network architecture (sigmoid function type architecture) and the GMDH architecture (polynomial type architecture). This algorithm contains the characteristics of the conventional neural network and the GMDH algorithm and it is a very flexible method for the identification problem of the complex nonlinear system.

5 Conclusion

In this study, the feedback GMDH-type neural network using prediction error criterion defined as AIC, was proposed. This algorithm can automatically organize a multilayered neural network architecture fitting the complexity of the nonlinear system by using the heuristic self-organization method. It is very easy to apply this algorithm to the identification problem of the practical complex system because the optimum neural network architecture fitting the complexity of the nonlinear system, is automatically organized. This algorithm was applied to the nonlinear system identification problem and it was shown that this algorithm was accurate and a very useful method.

Acknowledgments. This work was supported by (JSPS) KAKENHI 22560403.

References

[1] Kondo, T., Ueno, J.: Medical image diagnosis of lung cancer by a revised GMDH-type neural network using various kinds of neurons. Artificial Life and Robotics 16(3), 301–306 (2011)
[2] Kondo, T., Ueno, J.: Medical image recognition of abdominal multi-organs by RBF GMDH-type neural network. International Journal of Innovative Computing, Information and Control 5(1), 225–240 (2009)
[3] Farlow, S.J. (ed.): Self-organizing methods in modeling, GMDH-type algorithm. Marcel Dekker, Inc., New York (1984)
[4] Ivakhnenko, A.G.: Heuristic self-organization in problems of engineering cybernetics. Automatica 6(2), 207–219 (1970)
[5] Tamura, H., Kondo, T.: Heuristics free group method of data handling algorithm of generating optimum partial polynomials with application to air pollution prediction. Int. J. System Sci. 11(9), 1095–1111 (1980)
[6] Akaike, H.: A new look at the statistical model identification. IEEE Trans. Automatic Control AC-19(6), 716–723 (1974)
[7] Draper, N.R., Smith, H.: Applied Regression Analysis. John Wiley and Sons, New York (1981)

GPGPU Implementation of Fuzzy Rule-Based Classifiers

Tomoharu Nakashima, Keigo Tanaka, Noriyuki Fujimoto, and Ryosuke Saga

Abstract. This paper presents a parallel implementation of fuzzy-rule-based classifiers using a GPGPU (General Purpose Graphics Processing Unit). There are two steps in the process of fuzzy rule-based classification: Fuzzy-rule generation from training data and classification of an unseen input pattern. The proposed implementation parallelizes these two steps. In the step of fuzzy-rule generation from training patterns, the membership calculation of a training pattern for available fuzzy sets is simultaneously processed. On the other hand, the membership calculation of an unseen pattern for the generated fuzzy if-then rules is simultaneously processed in the step of the classification of the pattern. The efficiency of the parallelization is evaluated through a series of computational experiments. The effect of the parallelization is evaluated for each step of fuzzy rule-based classifiers. The results of the computational experiments show that the proposed implementation successfully improve the speed of fuzzy rule-based classifiers.

1 Introduction

This paper presents parallel implementation of fuzzy rule-based classifiers that contain the complete set of fuzzy if-then rules. Each fuzzy if-then rule in the constructed fuzzy rule-based classifiers has all attributes available in a given training data set. The key idea is to employ a GPGPU (General Purpose computation on Graphics Processing Units) in order to reduce the computational time. GPUs, which were originally developed for graphics processing, have a large number of multiprocessors and thus allow us for high-speed parallel computation. At the time the GPGPU development was first made public, it was not easy for everyone to make full use of the parallel architecture due to the fact that the programmers needed to understand it deeply. Intel Corporation released a development tool set called CUDA (Compute Unified Device Architecture) in 2007 [1]. CUDA enables us to implement parallel processes in GPGPUs without any deep understandings of graphics processing.

Tomoharu Nakashima · Keigo Tanaka · Noriyuki Fujimoto · Ryosuke Saga
Osaka Prefecture University, Gakuen-cho 1-1, Naka-ku, Sakai, Osaka 599-8531, Japan
e-mail: nakashi@cs.osakafu-.ac.jp, mu301008@edu.osakafu-u.ac.jp, fujimoto@mi.s.osakafu-u.ac.jp, saga@cs.osakafu-u.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 323–332.
springerlink.com

Under CUDA, the GPU process can be implemented in C programming language. For example, Fujimoto [4] proposed a dense matrix-vector multiplication on the NVIDIA CUDA architecture.

In this paper, CUDA is employed for fuzzy rule-based classification systems. To implement with CUDA, it should be considered that a GPU has its own memories which are only accessible from the GPU. GPU's memories are composed of several types of devices with different access speeds and capacities. Thus the efficiency of parallel computing with GPUs depends on the optimality of the memory access. However, it is difficult to design the memory access optimally without understanding the details of the hardware architecture of GPUs. In this paper, we adapt existing implementation to fuzzy classification system.

The CUDA package includes libraries that help programmers develop softwares for GPGPUs. CUBLAS is one of the included libralies in the CUDA package that implements BLAS (Basic Linear Algebra Subprograms) computation for GPGPUs [2]. It allows us to develop parallel computing programs easily without heavily modifying source codes. While all algorithms in CUBLAS are published as binary files, some source codes of SGEMM (Single precision General Matrix Multiply) algorithms have been published by the developer.

In the proposed implementation, the calculation of the membership values is viewed as matrix calculation, using two matrices which represent antecedent fuzzy sets and training patterns. The published source codes of the matrix multiplication in SGEMM are modified so that the membership values of given training patterns with antecedent fuzzy sets are calculated in parallel. In a series of computational experiments, the computational time of the proposed method is compared with that of the traditional method that only uses a CPU. It is shown that the proposed method reduces the computational time for pattern classification problems that have high dimensionality and/or a large number of training patterns.

2 Fuzzy Classification System

2.1 Fuzzy If-then Rule

In a pattern classification problem with n dimensionality and M classes, we suppose that m labeled patterns, $\boldsymbol{x}_p = \{x_{p1}, x_{p2}, \cdots, x_{pn}\},\ p = 1, 2, \cdots, m$, are given as training patterns. We also assume that without loss of generality, each attribute of $\boldsymbol{x}_p$ is normalized to a unit interval $[0, 1]$. From the training patterns we generate fuzzy if-then rules of the following type:

$$R_q\text{: If } x_1 \text{ is } F_{q1} \text{ and } \cdots \text{ and } x_n \text{ is } F_{qn} \text{ then Class } C_q \text{ with } CF_q, \quad q = 1, 2, \cdots, N, \tag{1}$$

where R_q is the label of the q-th fuzzy if-then rule, $\boldsymbol{F}_q = (F_{q1}, \cdots, F_{qn})$ represents a set of antecedent fuzzy sets, C_q a the consequent class, CF_q is the confidence of the rule R_q, and N is the total number of generated fuzzy if-then rules.

Let us denote the compatibility of a training pattern $\boldsymbol{x}_p$ with a fuzzy if-then rule R_q as $\mu_{\boldsymbol{F}_q}(\boldsymbol{x}_p)$. The compatibility $\mu_{\boldsymbol{F}_q}(\boldsymbol{x}_p)$ is calculated as follows:

$$\mu_{\boldsymbol{F}_q}(\boldsymbol{x}_p) = \prod_{i=1}^{n} \mu_{F_{qi}}(x_{pi}),\ q = 1, 2, \cdots, N, \tag{2}$$

where $\mu_{F_{qi}}(x_{pi})$ is the compatibility of x_{pi} with the fuzzy set F_{qi} and x_{pi} is the the i-th attribute value of $\boldsymbol{x}_p$.

The number of fuzzy rules to be generated is L^n. That is, the number of rules increases exponentially for the division number and the dimensionality.

2.2 Generating Fuzzy If-then Rules

A fuzzy classification system consists of a set of fuzzy if-then rules. The fuzzy if-then rules are generated from the training patterns $\boldsymbol{x}_p$, $p = 1, 2, \ldots, m$. The number of generated fuzzy if-then rules is determined by the number of fuzzy partitions for each axis. That is, the number of generated fuzzy if-then rules is the number of combinations of fuzzy sets that are used for attribute axes. Although different numbers of fuzzy partitions can be used for different axes, in this paper we assume that it it the same for all axes. In this case, the number of fuzzy if-then rules is calculated as $N = L^n$ where n is the dimensionality of the pattern classification problem at hand. In this paper, it is supporsed that all attributes are divided in the same way (i.e., the same fuzzy partition). An illustrative example is shown in Fig. 1. In Fig. 1, a two-dimensional pattern space is divided into $3^2 = 9$ fuzzy subspaces as each attribute is divided into three fuzzy sets. Each subspace is labled with a rule label ($R_1 \sim R_9$). For example, the antecedent part of Rule R_6 has the fuzzy set A_3 for attribute x_1 and A_2 for attribute x_2. In this way, the total number of generated fuzzy if-then rules and the antecedent part of each fuzzy if-then rule is automatically determined after the number of fuzzy sets for each attribute is determined.

The consequent part of fuzzy if-then rules (i.e., C_q and CF_q in (1)) is determined from the given training patterns once the antecedent part is specified. The consequent class C_q of the fuzzy if-then rule R_q is determined as follows:

$$C_q = \arg \max_{h=1,\ldots,M} \beta_h^q, \tag{3}$$

where

$$\beta_h^q = \sum_{\mathbf{x_p} \in \text{Class } \mathbf{h}} \mu_{\mathbf{F_q}}(\mathbf{x_p}). \tag{4}$$

That is, the most matching class with the fuzzy if-then rule is selected considerng the given training patterns. If there is not any training pattern that is covered by the fuzzy if-then rules, the consequent class is set as empty. Also, in the case where

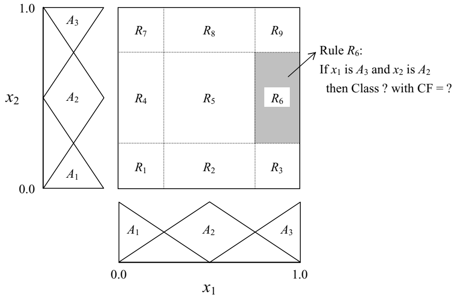


Fig. 1 Two-dimensional illustrative example of specifying the antecedent part of a fuzzy if-then rule (three fuzzy sets for both the two attributes)

multiple classes have the maximum value in (3), the consequent class is set as emply. The confidence CF_q is determined as follows:

$$CF_q = \frac{\beta_{C_q} - \bar{\beta}}{\sum_{h=1}^{m} \beta_h^q}, \tag{5}$$

where

$$\bar{\beta} = \frac{1}{M-1} \sum_{h \neq C_q} \beta_h^q. \tag{6}$$

There are other formulations for determining the confidence. Interested readers are referred to [5] for the discussion on the confidence calculation and the performance evaluation.

2.3 *Classification of Unseen Patterns*

Generated fuzzy if-then rules in the previous subsections are used to assign a class label to an unseen pattern which is not included in the set of training patterns. Let us denote an n-dimensional unseen pattern as $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. The fuzzy inference is employed to classify unseen patterns in the fuzzy classification system in this paper. The class of an unseen pattern **x** is classified as Class C that is determined by the following equation:

$$C = \arg \max_{h=1,\ldots,M} \{\alpha_h\}, \tag{7}$$

where

$$\alpha_h = \max_{\substack{q=1,\ldots,N \\ C_q = h}} \{\mu_{\boldsymbol{F}_q(\boldsymbol{x})} \cdot CF_q\}. \tag{8}$$

In the above equations, M is the number of classes and N is the number of geneated fuzzy if-then rules. In (7), if there are multiple classes that have the same maximum value of α_h, the classification of the unseen pattern is rejected.

3 GPGPU

3.1 CUDA

In CUDA, while functions that are executed on a CPU are compiled with a c-compiler, those that are executed on a GPU are compiled with NVCC (NVIDIA CUDA Compiler). As a c-compiler for developing CPU computing programs, GCC is used in this paper.

Since the data structures for GPU computing in CUDA is the same as that for CPU computing, it is easy to implement GPU computing if the programmer is familiar to C programming language. However, GPU memories are separate from CPU memories and only be accessible from the GPU. Thus data transfer between CPU memories and GPU memories is necessary before/after the GPU computation. The basic procedure of GPU computing is composed of four steps:

1. Data are transferred from CPU memories to GPU memories.
2. The CPU instructs the GPU to perform computation.
3. The GPU executes the computation.
4. The GPU transfers the result to the CPU memories.

Instructions sent to a GPU are composed of threads, blocks, and grids depending on the level of parallelization. A thread is an atomic-level execution of the instructions. A blocks is a set of threads, and a grid is a set of blocks. The hierarchical strucure of the instructions is shown in Fig. 2

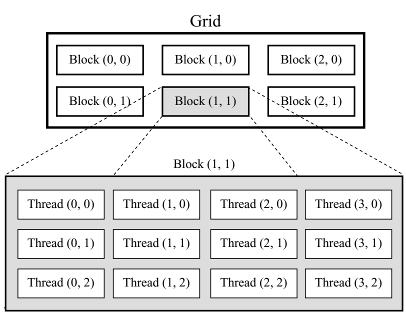


Fig. 2 Hierarchical structure of instructions in GPGPU

The CPU can only send instructions to grids. When instructions are sent to a grid, the instructions are sent to the blocks in the grid, then are executed by the threads in the blocks. Thus the threads in each block execute the calculations parallelly and the instructions of blocks in a grid are also executed parallelly. During the calculations, threads which belong to the same block can be synchronized and make use of a shared memory whose access speed is higher than the global memory.

The number of thread per block and the number of blocks per grid are design parameters by the programmer. Thus the efficiency of parallel computing with GPUs depends on the memory access and the composition of threads, blocks, and grids. However, it is difficult to design the optimal memory access without understanding the details of the hardware architecture of GPUs.

3.2 CUBLAS (CUDA Basic Linear Algebra Subprograms)

Although CUDA has enabled us to implement paralell computing with GPUs more easily than before, there is still an issue of designing the structure of instructions (i.e., the size of grids and blocks). Since the aim of this paper is to parallelize the calculation of compatibilities in (2) and (8), if some operations are similar to the compatibility calculation and the source codes of the operations for GPGPU are available, it is better to adapt them to our aim. In this paper,a library included in CUBLAS (CUDA Basic Linear Algebra Subprograms) [2] is used for this purpose.

CUBLAS is the CUDA version of BLAS (Basic Linear Algebra Subprograms) [3] and is included in the package of CUDA. BLAS is the optimized library for linear algebra operations. CUBLAS eases the burden of the programmers for re-implementing the parallel processing of linear algebra. SGEMM (Single precision General Matrix Multiply) and DGEMM(Double precision GEMM) are core functions of (CU)BLASS that perform general matrix multiplication.

CUBLAS was implemented and released by Volkov et al.[6]. Among the released library functions, the source codes of SGEMM are also published by them. Since the memory access of CUBLAS is designed efficiently, user can implement their algorithms without any concern about memory access diverting the memory access of it. In this paper, we modify the source code of SGEMM so that the calculation of membership values are parallelized in order to reduce the computational time. Representation and processing of floating point on GPUs follows IEEE754, and we suppose that real numbers on CPUs and GPUs are both single precision in this paper.

4 Parallel Implementation of Fuzzy Rule-Based Systems

This section describes the implementation of fuzzy rule-based systems using GPGPU. Since the process of calculating membership functions is common in Equation (2), this implies that the same procedure is iterated for calculating the compatibility of a training pattern with each fuzzy if-then rule: First calculating the compatibility for each attribute, and then multiplying them. Therefore, we can view this process as a function of two matrices. One matrix represents a set of fuzzy if-then rules. The size of this matrix is $N \times n$ and is composed of N row vectors whose lengths are n and elements are antecedent fuzzy sets $\boldsymbol{F}_q$. The other matrix represents a set of training patterns. This matrix is $n \times m$ and is composed of m column vectors whose lengths are n and each column is a training pattern $\boldsymbol{x_p}$. In the conventional matrix multiplication for two matrices, the (q, p) element of the product, r_{qp}, is represented as:

$$r_{qp} = \sum_{i=1}^{n} F_{qi} \times x_{pi}. \tag{9}$$

We adapt the above calculation to the calculation of membership value $\mu_{F_{qi}}(x_{pi})$ with the same access order as matrix multiplication. Thus the (q,p) element of the result, r'_{qp}, is represented as:

$$r'_{qp} = \prod_{i=1}^{n} F_{qi} \odot x_{pi}, \tag{10}$$

where $\odot$ denotes the membership calculation. That is, the membership calculation can be regarded as a matrix operation where product operation is replaced with a membership function and sum operation is replaced with a product operation.

As mentioned in Section 2, the formulation of membership values for fuzzy-rule generation is similar to that of matrix multiplication. We modify the SGEMM algorithm introduced in section 3 to be the algorithm to generate fuzzy if-then rules. Volkov et al.[6] published the SGEMM algorithm that calculates the following equation:

$$\boldsymbol{C}^{\text{new}} = \alpha \times \boldsymbol{A} \times \boldsymbol{B}^{\text{T}} + \beta \times \boldsymbol{C}^{\text{old}}, \tag{11}$$

where $\boldsymbol{A}$ is a $x \times y$ matrix, $\boldsymbol{B}^{\text{T}}$ is a $y \times z$ matrix, and $\boldsymbol{C}$ is a $x \times z$ matrix. α and β are scalar values. Equation (11) is calculated parallelly by a GPU after initialization by a CPU. In this paper we specify that $\alpha = 1$ and $\beta = 0$ to consider only the matrix multiplication. By representing elements of the matrices as $\boldsymbol{A} = (a_{ij})$, $\boldsymbol{B}^{\text{T}} = (b_{ij})$, and $\boldsymbol{C} = (c_{ij})$ and a temporal variable as t, the procedure to calculate an element of $\boldsymbol{C}^{\text{new}}$, c_{ij}, can be shown as the pseudocode in Fig. 3(a).

(a) Matrix multiplication	(b) Proposed method

$t = 0$
for($k = 1$ to y){
$\quad t$ += $a_{ik} \times b_{kj}$
}
$c_{ij} = t$
	$t = 1$
	for($k = 1$ to y){
	$\quad t \times= \mu(a_{ik}, b_{kj})$
	}
	$c_{ij} = t$

Fig. 3 Pseudocodes of the matrix multiplication and the proposed method

The parallel procedure to calculate compatibility is shown in Fig. 3(b), where a_{ik} is the label of the antecedent fuzzy set, and b_{kj} is the input value and $\mu(a_{ik}, b_{kj})$ is the membership function of the input value b_{kj} for the fuzzy set a_{ik}. Thus the order to access the elements of each matrices is the same as that of the original matrix multiplication. Therefore, the consistency of the parallel computation holds by replacing addition and multiplication of the elements in matrix multiplication to multiplication and membership calculation respectively. In addition, Volkov et al.[6] employs 16×4 threads per block and $(x/64) \times (y/16)$ blocks per grid to make the memory access efficient. However, this limitation has no effect on the calculation of the equation. Now we can parallelize the membership calculation on GPUs by applying the above procedure to matrices which represent antecedent fuzzy sets and training pattern sets.

We suppose that $x_{F_{qi}}$ is an element of $\boldsymbol{A}$, where $x_{F_{qi}}$ is the mode of the fuzzy set F_{qi}:

$$\boldsymbol{A} = \begin{bmatrix} x_{F_{11}} & \cdots & x_{F_{1n}} \\ \vdots & \ddots & \vdots \\ x_{F_{N1}} & \cdots & x_{F_{Nn}} \end{bmatrix}. \tag{12}$$

And for a set of training patterns, we set a transposed matrix $\boldsymbol{B}^{\mathrm{T}}$ as follows:

$$\boldsymbol{B}^{\mathrm{T}} = \begin{bmatrix} x_{11} & \cdots & x_{m1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \cdots & x_{mn} \end{bmatrix}. \tag{13}$$

By applying the calculation of compatibility modified from matrix multiplication to the above two matrices, a $N \times m$ matrix $\boldsymbol{C}$ is computed as:

$$\boldsymbol{C} = \begin{bmatrix} \mu_{\boldsymbol{F}_1}(\boldsymbol{x}_1) & \cdots & \mu_{\boldsymbol{F}_1}(\boldsymbol{x}_m) \\ \vdots & \ddots & \vdots \\ \mu_{\boldsymbol{F}_N}(\boldsymbol{x}_1) & \cdots & \mu_{\boldsymbol{F}_N}(\boldsymbol{x}_m) \end{bmatrix}, \tag{14}$$

where a row vector of $\boldsymbol{C}$ corresponds to the compatibility of rules for each patterns, i.e., equation (2). The procedure to parallelize fuzzy-rule generation with a GPU takes the following steps. First, by a CPU, $\boldsymbol{A}$ and $\boldsymbol{B}^{\mathrm{T}}$ are made, and transported to a GPU. Second, the GPU calculates $\boldsymbol{C}$ using the matrix operation for $\boldsymbol{A}$, $\boldsymbol{B}^{\mathrm{T}}$. Finally, $\boldsymbol{C}$ is transported to the CPU, and then it determines the consequent parts.

5 Computational Experiments

In order to verify the effect of parallelization with GPUs on computational cost, a series of computational experiments were conducted. In the experiments, the computational time to generate fuzzy rules and to infer the class of an unseen pattern is measured. For comparison purpose, the computational time was measured with and without the use of a GPU.

with a GPU is compared to that of a CPU. Table 1 shows the environment of the experiments. Although GeForce GTX 295 has a dual-chip structure, we use only one chip. In the experiments, the computational time for solving a two-class problem is compared. 100 classification problems with different number of training patterns and dimensionalities were used to evaluate the computational time. The number of fuzzy sets for each axis is fixed to two. The results were averaged to compare the efficiency of the parallelization. The results of the experiments are shown in Figs. 4-5. Figure 4 shows how dimensionality of the problem has an effect on the computational time when the number of training patterns is 64 and 816. In Fig. 4, the computational time with a CPU is shorter than that with a GPU when the dimensionality is small. As the dimensionality increased, the computational time with a CPU increased drastically while that with a GPU keeps short. Figures 5 shows that the number of patterns has an effect on the computational time when the dimensionality of the problem is 12 or 18 respectively. In Fig. 5, the computational time with a CPU is shorter than that with a GPU when the number of patterns is small. However, when the number of training patterns is large, the computational time with a GPU is shorter than that with a CPU. The computational time with a GPU is shorter than that with a CPU constantly. Thus the parallel computation for generating fuzzy rules with a GPU has the effect of reducing the computational time except in the case of low-dimensionality problems with small amount of training patterns.

Table 1 Environment of the experiments

CPU	Intel Core i7 Extreme 945
Clock Frequency	3.20 GHz
Memory Size	5.8 GB
Memory Clock	667 MHz
GPU	NVIDIA GeForce GTX 295
Processor Core	240
Processor Clock	1242 MHz
Memory Size	896 MB
Memory Clock	999 MHz
OS	Linux x86_64
Development Environment	CUDA(NVCC)2.2, GCC4.3

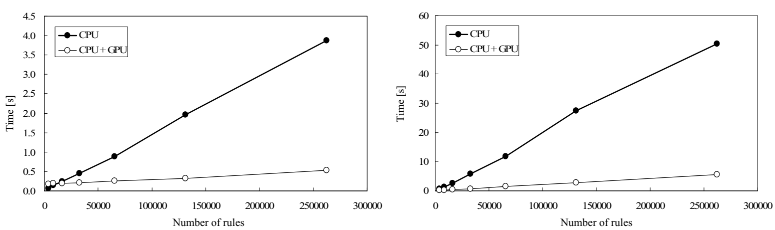


Fig. 4 Computational time for dimensions (64 and 816 training patterns)

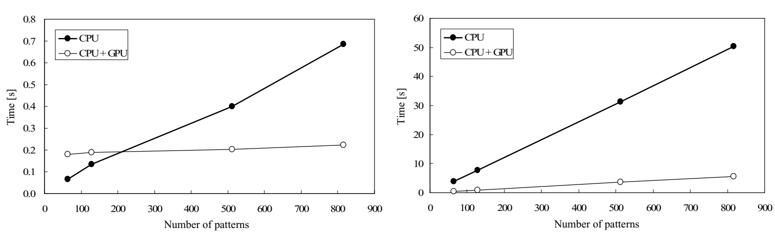


Fig. 5 Computational time for number of patterns (12 and 18 dimensionalities)

6 Conclusions

In this paper, we proposed a method to parallelize fuzzy-rule generation with a GPU using matrix multiplication that is optimized for CUDA. Computational experiments showed that the method reduced the computational time when the dimensionality of the problem and/or the number of training patterns were large. For future works, we will try to parallelize fuzzy inference with GPU, or resolve lack of memory on GPU when the method is applied to problems with further dimensionality and/or the number of training patterns.

References

1. Nvidia cuda, http://www.nvidia.com/object/cuda_home.html
2. nVidia Corporation: Cuda cublas library (2010), http://developer.download.nvidia.com/compute/cuda/3_2_prod/toolkit/docs/CUBLAS_Library.pdf
3. Foundation, N.S., of Energy, D.: BLAS (2010), http://www.netlib.org/blas/
4. Fujimoto, N.: Faster matrix-vector multiplication on GeForce 8800GTX. In: IEEE International Symposium on Parallel and Distributed Processing, IPDPS 2008, pp. 1–8 (2008), doi:10.1109/IPDPS.2008.4536350
5. Ishibuchi, H., Nakashima, T., Nii, M.: Classification And Modeling With Linguistic Information Granules: Advanced Approaches To Linguistic Data Mining, 1st edn. Springer-Verlag New York, Inc., Secaucus (2004)
6. Volkov, V., Demmel, J.: Lu, qr and cholesky factorizations using vector capabilities of gpus. Tech. Rep. UCB/EECS-2008-49, EECS Department, University of California, Berkeley (2008), http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-49.html

Granule Mining and Its Application for Network Traffic Characterization

Bin Liu, Yuefeng Li, and Kewen Wang

Abstract. Decision table and decision rules play an important role in rough set based data analysis, which compress databases into granules and describe the associations between granules. Granule mining was also proposed to interpret decision rules in terms of association rules and multi-tier structure. In this paper, we further extend granule mining to describe the relationships between granules not only by traditional support and confidence, but by *diversity* and *condition diversity* as well. *Diversity* measures how diverse of a granule associated with the other granules, it provides a kind of novel knowledge in databases. Some experiments are conducted to test the proposed new concepts for describing the characteristics of a real network traffic data collection. The results show that the proposed concepts are promising.

1 Introduction

Association rule mining[1] has been widely adopted in many real applications, to reveal hidden knowledge in databases. However, usually a lot of patterns or rules are generated, many of them are meaningless and user-unwanted [2, 4]. Currently, a lot of research has been done on generating useful patterns or association rules.

Rough set theory describes decision rules by decision table, which compresses databases into granules and reveals the associations between granules [9, 14]. Mining decision rules has been used for classifications, decision tree and flow graph [7, 10, 11, 12, 13, 15]. The advantage of using decision rules is to reduce the two-steps of association mining into one process. However, it lacks accuracy and flexibility to deal with the associations between data granules in databases [4].

Bin Liu · Yuefeng Li
School of Electrical Engineering and Computer Science
Queensland University of Technology, Brisbane, QLD 4001, Australia

Kewen Wang
School of Information and Communication Technology
Griffith University, Brisbane, QLD 4111, Australia

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 333–343.
springerlink.com

Granule mining [3, 5, 4, 6] is a novel theory that interprets decision rules in terms of association rules. It formally describes the process of finding interesting granules, as well as the corresponding associations between granules in a database. Granule mining also proposes multi-tier structure to build association mappings for efficiently discovering the interesting association rules in different size granules.

In this paper, we continue to extend the theory of granule mining. We use granule mining to reveal the relationships between granules not only by traditional support and confidence, but by *diversity* and *condition diversity* as well. *Diversity* measures how diverse of a granule associated with the other granules. For example, in computer network traffic, suppose host A sends 1,000 packets to only one host, and another host B sends 1000 packets to 1000 hosts(one packet for each host). Both hosts' support(or frequency) values are 1000, but the first one connects with only one host while the second one connects 1000 hosts. We propose *diversity* to measure host's this feature. Host A's support is 1000, diversity is 1; host B's support is 1000 but diversity is 1000. We also propose *condition diversity* to extract rules to interpret a granules's *diversity*.

We use the proposed method to describe the characteristics of network traffic, with a promising result. It can effectively discover the significant hosts in network which behave not only frequent but also diverse. The rules discovered by the proposed method can provide perfect interpretation of hosts' behaviors.

The remainder of the paper is structured as follows. Section 2 presents the concepts of decision table and decision rules. Section 3 presents the concept of association mapping. Multi-tier structure is presented in Section 4. In Section 5, we introduce the concept of *diversity* and *condition diversity*. The experimental result is presented in Section 6. It is followed by conclusion in the last section.

2 Decision Table and Decision Rules

In this section, we introduce the concept of granules, decision table and decision rules. These concepts and definitions are firstly introduced by Pawlak [8, 9], and are further developed by Li et al. [4, 5].

Formally, a transaction database can be described as an information table (T, V^T), where T is the set of transactions in which each transaction is a set of items, and $V^T = \{a_1, a_2, \ldots, a_n\}$ is a set of attributes for all transactions in T.

Let B be a subset of V^T. B determines a binary relation $I(B)$ on T, such that $(t_i, t_j) \in I(B)$ if and only if $a(t_i) = a(t_j)$ for every $a \in B$, where $a(t)$ denotes the value of attribute a for object $t \in T$. The family of all equivalence classes of $I(B)$, that is a partition determined by B, is denoted by T/B. The equivalence classes in T/B are referred to $B - granules$. The class in T/B induced by t is denoted by $B(t)$.

Definition 1. Let $g = B(t)$ be a *granule* induced by t. Its covering set

$$coverset(g) = \{t' | t' \in T, B(t') = B(t) = g\}$$

Definition 2. The *support* or *frequency* of granule g is

$$sup(g) = |coverset(g)|$$

where $|coverset(g)|$ is the cardinality of $coverset(g)$.

The *support* of a granule measures the number of the transactions that induce the granule.

We call the tuple (T, V^T, C, D) a decision table of information table (T, V^T) if $C \cap D = \emptyset$ and $C \cup D \subseteq V^T$, where C is a set of condition attributes and D is a set of decision attributes. The granules determined by C are called $C-granules$, granules determined by D are called $D-granules$, and the granules determined by $C \cup D$ are called basic $granules$.

Theorem 1. *Let granule $g = cg \wedge dg$, cg is a* C-granule *and dg is a* D-granule. *We have $coverset(g) = coverset(cg) \cap coverset(dg)$.*

Proof. Let g be induced by t, that is $g = (C \cup D)(t)$; therefore, we have $cg = C(t)$ and $dg = D(t)$. Assume transaction $t' \in T$. Because

$$(t', t) \in I(C \cup D) \Leftrightarrow (t', t) \in I(C) \; and \; (t', t) \in I(D).$$

Therefore, we have $t' \in coverset(g) \Leftrightarrow (t', t) \in I(C \cup D) \Leftrightarrow (t', t) \in I(C)$ *and* $(t', t) \in I(D) \Leftrightarrow t' \in (coverset(cg) \cap coverset(dg))$. □

A condition granule cg and a decision granule dg forms a decision rule $cg \to dg$.

Its *support* is

$$sup(cg \to dg) = sup(cg \wedge dg)$$

Its $confidence$ is

$$conf(cg \to dg) = \frac{sup(cg \wedge dg)}{sup(cg)}$$

We usually set a minimum support threshold *min_sup* and a minimum confidence threshold *min_conf* to select the interesting decision rules.

3 Association Mappings

The relationships between condition granules and decision granules can be described as basic association mappings(*BAMs*).

Definition 3. Let (T, V^T, C, D) be a decision table, T/C, T/D and $T/(C \cup D)$ are $C-granules$, $D-granules$ and basic $granules$. For each condition granule $cg \in T/C$, its association mapping is

$$\Gamma_{CD}(cg) = \{(dg, sup(cg \wedge dg)) | (cg \wedge dg) \in T/(C \cup D)\}$$

where $sup(cg \wedge dg)$ is the support of granule $cg \wedge dg$.

$\Gamma_{CD}(cg)$ includes all the decision granules that have relationships with the condition granule cg. Specially, $\Gamma_{CD}(cg) = \emptyset$ if $D = \emptyset$. The *support* of cg can be directly calculated from the association mapping.

Definition 4. Given two granules $cg \in T/C, g \in T/(C \cup D)$. If $coverset(cg) \supseteq coverset(g)$, we call cg a generalized granule of g, We use $cg \succ g$ to denote the generalized relationship between cg and g.

Theorem 2. *Let cg is a* C-granule*, the coverset of cg is*

$$coverset(cg) = \bigcup_{cg \succ g_i} coverset(g_i)$$

Proof. Assume $t \in coverset(cg)$, t also can induce a granule $g_i = cg \wedge dg_i$, hence we have $t \in coverset(g_i)$, therefore,

$$t \in \bigcup_{cg \succ g_i} coverset(g_i)$$

On the other hand, assume

$$t \in \bigcup_{cg \succ g_i} coverset(g_i)$$

t also can induce a granule cg, hence $t \in coverset(cg)$. □

Theorem 3. *Given condition granule cg and its association mapping $\Gamma_{CD}(cg)$. The support of condition granule cg is*

$$sup(cg) = \sum_{(dg_i, sup) \in \Gamma_{CD}(cg)} sup(cg \wedge dg_i) \tag{1}$$

where (dg_i, sup) is a short of $(dg_i, sup(cg \wedge dg_i))$. □

Proof. According to **Theorem 2.**, we have

$$coverset(cg) = \bigcup_{cg \succ g_i} coverset(g_i)$$

Since $g_i = (cg \wedge dg_i) \in T/(C \cup D)$, we have

$$coverset(cg) = \bigcup_{(cg \wedge dg_i) \in T/(C \cup D)} coverset(cg \wedge dg_i)$$

Hence,

$$sup(cg) = |coverset(cg)| = \sum_{(cg \wedge dg_i) \in T/(C \cup D)} sup(cg \wedge dg_i)$$
$$= |\bigcup_{(cg \wedge dg_i) \in T/(C \cup D)} coverset(cg \wedge dg_i)|$$
$$= \sum_{(dg_i, sup) \in \Gamma_{CD}(cg)} sup(cg \wedge dg_i) \qquad \square$$

We can set a minimum support min_sup to select the significant condition granules.

From association mapping $\Gamma_{CD}(cg)$, we also can easily calculate all the decision rules that condition part is cg.

For each $(dg, sup(cg \wedge dg)) \in \Gamma_{CD}(cg)$, cg and dg forms a decision rule $cg \to dg$.

Its support is

$$sup(cg \to dg) = sup(cg \wedge dg)$$

Its confidence is

$$conf(cg \to dg) = \frac{sup(cg \wedge dg)}{sup(cg)} = \frac{sup(cg \wedge dg)}{\sum_{(dg_i, sup) \in \Gamma_{CD}(cg)} sup} \tag{2}$$

where (dg_i, sup) is a short for $(dg_i, sup(cg, dg_i))$, $confidence$ of a rule $cg \to dg$ reveals an estimate of condition probability of dg given cg. It can be used to interpret why condition granule cg is significant. The higher the confidence value, the stronger association dg has with cg. We can set a minimum confidence threshold min_conf to select the interesting rules.

4 Multi-tier Structure

In this section, we first describe multi-tier structure to represent the relationships of different size granules. We then illustrate how to efficiently extract association rules in the multi-tier granules by deriving generalized association mappings from the basic association mappings.

Assume D_h is a subset of D, we can get a set of smaller decision granules $D_h - granules$. The relationships between $C - granules$, $D - granules$ and $D_h - granules$ can be represented by a 3-tier structure. Figure 1 illustrates an example of 3-tier structure. The 3-tier structure includes three kinds of association mappings, $\Gamma_{CD}, \Gamma_{CD_h}$ and $\Gamma_{(C+D_h)(D-D_h)}$.

Fortunately, Γ_{CD_h} can be derived from the basic association mapping Γ_{CD}. We call Γ_{CD_h} generalized association mappings of Γ_{CD}.

Theorem 4. *Let $\Gamma_{CD_h}(cg)$ and $\Gamma_{CD}(cg)$ are two association mappings, $cg \in T/C$, $D_h \subset D$. For each element $(d_h g, sup(cg \wedge d_h g)) \in \Gamma_{CD_h}(cg)$,*

$$sup(cg \wedge d_h g) = \sum_{(cg \wedge d_h g) \succ (cg \wedge dg_i)} sup(cg \wedge dg_i)$$

where $(dg_i, sup(cg \wedge dg_i)) \in \Gamma_{CD}(cg)$. □

Proof. According to Theorem 2,

$$coverset(cg \wedge d_h g) = \bigcup_{(cg \wedge d_h g) \succ (cg \wedge dg_i)} coverset(cg \wedge dg_i)$$

Hence,

$$\begin{aligned} sup(cg \wedge d_h g) &= |coverset(cg \wedge d_h g)| \\ &= |\bigcup_{(cg \wedge d_h g) \succ (cg \wedge dg_i)} coverset(cg \wedge dg_i)| \\ &= \sum_{(cg \wedge d_h g) \succ (cg \wedge dg_i)} sup(cg \wedge dg_i) \end{aligned}$$ □

From Γ_{CD_h}, we can generate a set of rules, which can reveal the knowledge of cg in a high level. For each element $(d_h g, sup(cg \wedge d_h g)) \in \Gamma_{CD_h}(cg)$, cg and $d_h g$ form a new rule $cg \to d_h g$,

$$sup(cg \to d_h g) = sup(cg \wedge d_h g) = \sum_{(cg \wedge d_h g) \succ (cg \wedge dg_i)} sup(cg \wedge dg_i)$$

$$conf(cg \to d_h g) = \frac{sup(cg \wedge d_h g)}{sup(cg)} = \frac{\sum_{(cg \wedge d_h g) \succ (cg \wedge dg_i)} sup(cg \wedge dg_i)}{\sum_{(dg_i, sup) \in \Gamma_{CD}(cg)} sup}$$

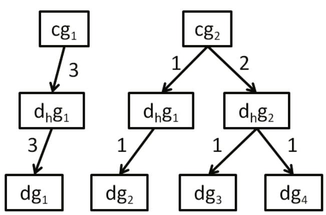


Fig. 1 A 3-tier structure of *C-granules,* $D_h - granules$ and *D-granules*

5 Diversity

In this subsection, we introduce the concept of *diversity* of a granule, we then present the concept of *diversity* and *condition diversity* of a rule.

5.1 Diversity of Granule

Definition 5. Given a condition granule cg and its basic association mapping $\Gamma_{CD}(cg)$, where $cg \in T/C, C \cap D = \emptyset$. The *diversity* of condition granule cg is defined as the cardinality of set $\Gamma_{CD}(cg)$.

$$divs_{CD}(cg) = |\Gamma_{CD}(cg)| \tag{3}$$

The *diversity* measures how diverse a condition granule connects with its decision granules. The higher the *diversity* value, the more diverse a condition granule is, the more significant a condition granule is. Specially, $divs(cg) = 1$ if $\Gamma_{CD}(cg) = \emptyset$.

Note that we always calculate diversity value of a condition granule according to its basic association mapping, since the cardinality of generalized association mapping of a condition granule can only reflect its diversity in the generalized tier.

We can set a minimum *diversity* min_divs, to select the significant condition granules.

5.2 Condition Diversity of Rule

Let $\Gamma_{CD}(cg)$ be a basic association mapping, for each element $(dg, sup(cg \wedge dg)) \in \Gamma_{CD}(cg)$, cg and dg form a rule $cg \rightarrow dg$,

Definition 6. Its *diversity* is defined as

$$divs(cg \rightarrow dg) = divs_{(C \cup D)\emptyset}(cg \wedge dg)$$

Definition 7. Its *condition diversity* is defined as

$$cond\_divs(cg \rightarrow dg) = \frac{divs_{(C \cup D)\emptyset}(cg \wedge dg)}{divs_{CD}(cg)} \tag{4}$$

The $cond\_divs(cg \rightarrow dg)$ is a ratio of granule $(cg \wedge dg)$'s *diversity* and granule cg's *diversity*. The higher the $cond\_divs$, the more likely the decision granule causes the condition granule behaving diverse.

We also can discuss *diversity* and *condition diversity* of a rule generated from generalized association mapping Γ_{CD_h}. For each element $(d_h g, sup(cg \wedge d_h g)) \in \Gamma_{CD_h}(cg)$, cg and $d_h g$ form a rule $cg \rightarrow d_h g$,

It's diversity is

$$divs(cg \rightarrow d_h g) = divs_{(C \cup D_h)(D - D_h)}(cg \wedge d_h g)$$

Its *condition diversity* is

$$cond\_divs(cg \rightarrow d_h g) = \frac{divs_{(C \cup D_h)(D - D_h)}(cg \wedge d_h g)}{divs_{CD}(cg)} \tag{5}$$

Note that the *diversity* value of condition granule cg is $|\Gamma_{CD}(cg)|$, instead of $|\Gamma_{CD_h}(cg)|$.

We can set a minimum *condition diversity* threshold min_conf to select the interesting rules that have strong contributions to the condition granule's diversity.

6 Experiments

To evaluate the proposed method, a set of experiments have been performed on several real network traffic datasets. There are several purposes in these experiments. The first is to evaluate the effect of minimum thresholds *min_sup* and *min_divs* to select significant hosts. The second is to evaluate the effect of minimum threshold *min_conf* to select the interesting rules. Finally, a result case study is conducted to show the effectiveness of the proposed method.

6.1 Dataset

The experimental datasets we used are *MAWI* data traces, which can be downloaded from http://tracer.csl.sony.co.jp/mawi/samplepoint-B/20060303/. The datasets are the four largest 15-minute data files on 03/03/2006. The characteristics of the datasets are listed in the table 1. An example of the data is illustrated in Table 2. We select five attributes to represent the features of packets, which are source IP address(SrcIP), source port(SrcPrt), destination IP address(DestIP), destination port(DestPrt) and protocol(Prot).

In the experiments, we designate *SrcIP* as condition attribute and the rest attributes as decision attributes, we try to discover significant hosts from the traffic data and extract interesting rules for the significant hosts.

Table 1 Characteristics of the selected datasets

ID	Time Captured	Packets Number	SrcIP Number
A	19:45–20:00	12,938,715	76,734
B	20:00–20:15	10,874,733	79,287
C	22:00–22:15	10,444,069	81,395
D	22:15–22:30	11,552,731	89,495

Table 2 An example of network traffic data

packets	*SrcIP*	*SrcPrt*	*Prot*	*DestIP*	*DestPrt*
t_1	95.94.90.25	2000	tcp	19.51.190.12	80
t_2	13.32.36.66	2002	tcp	19.51.190.12	80
t_3	13.32.36.66	2003	tcp	26.14.34.37	80
t_4	13.32.36.66	2004	udp	26.14.34.37	80

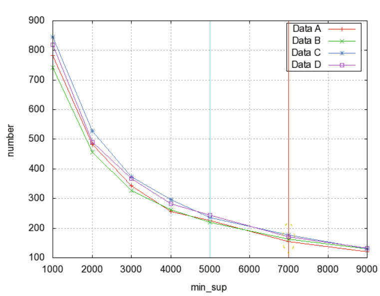


Fig. 2 Effect of *min_sup* to number of *SrcIP*

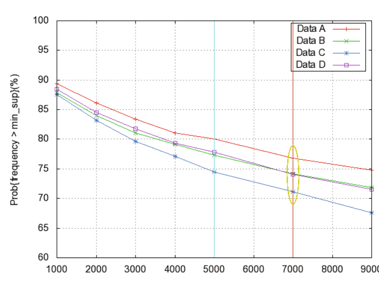


Fig. 3 Effect of *min_sup* to cumulative support(*SrcIP*)

6.2 *Effect of* min_sup *and* min_divs *Thresholds*

The effect of min_sup is illustrated in Fig 2 and Fig 3. When min_sup is set to 7000, less than 200 hosts are selected in each of the four datasets(see Fig 2), but at least 70% packets are generated by these hosts(see Fig 3).

The effect of min_divs is illustrated in Fig 4 and Fig 5. When min_divs is set to 400, less than 200 hosts are selected in each of the four dataset(see Fig 4), but these hosts are responsible for at least 55% of the whole diversity of the traffic(see Fig 5).

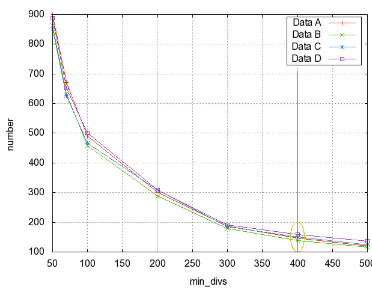


Fig. 4 Effect of min_divs to number of $SrcIP$

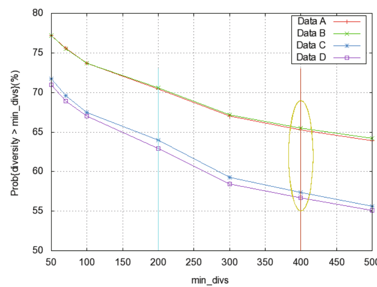


Fig. 5 Effect of min_divs to cumulative diversity($SrcIP$)

6.3 *Effect of Minimum Confidence* min_conf

Figure 6 and figure 7 show the effect of minimum confidence min_conf to select the rules. From the figure we can see, min_conf slightly effects the number of rules. It is because a number of rules' $confidence$ or $condition\ diversity$ value is 1.

6.4 *Number of Hosts and Rules*

Table 3 lists the number of significant hosts discovered by the proposed method. From the table we can see, more than 35% novel hosts are discovered by the $diversity$ feature.

Table 4 lists the number of significant rules discovered by the proposed method to interpret the significant hosts. From the table we can see, more than 40% novel rules discovered by the $diversity$ feature.

Table 3 Number of Hosts

Dataset	min_sup $= 7000$	$min\_sup = 7000$ or $min\_divs = 400$	Novel	ratio
A	152	255	103	0.40
B	161	255	94	0.37
C	174	285	111	0.38
D	168	286	118	0.41

Table 4 Number of Rules ($min\_conf = 0.3$)

Dataset	min_sup $= 7000$	$min\_sup = 7000$ or $min\_divs = 400$	Novel	ratio
A	219	387	168	0.43
B	233	393	160	0.41
C	238	419	181	0.43
D	257	453	196	0.43

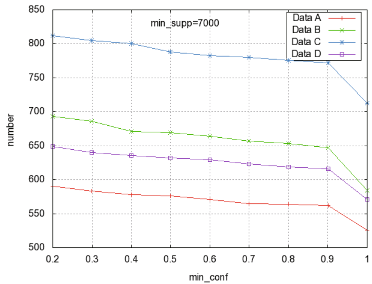


Fig. 6 Effect of *min_conf* to number of rules(support)

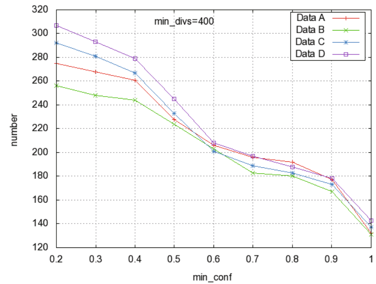


Fig. 7 Effect of *min_conf* to number of rules(diversity)

6.5 Case Study

Table 5 lists three results discovered from Dataset A. We briefly explain them to show their effectiveness to understand the behaviors of network traffic.

Table 5 Some interesting results

ID	Condition	support	diversity	Decision	*conf*	*cond_divs*
1	srcIP=215.35.248.109;	5,606,931	64,996	srcPrt=2893;prot=UDP; destIP=19.51.190.128;	1.00	1.00
2	srcIP=137.192.228.71;	408,675	4574	srcPrt=7700;proc=tcp;	0.63	0.16
				srcPrt=7700;prot=UDP;	0.01	0.76
3	srcIP=207.89.143.152;	58,816	58,816	srcPrt=6000;prot=tcp; destPort=1433;	1.00	1.00

The first host 215.35.248.109 and its rule discover a DoS attack. The host sent 5,606,840 *UDP* packets to host 19.51.190.128. Since *cond_divs* are 1, the host only attacks host 19.51.190.128. It attacks 64,996 ports of host 19.51.190.128.

The second host 137.192.228.71 and its rules reveal that rules that are responsible for support of a host sometimes are not responsible for *diversity* of the host. The host uses port 7700 and protocol *tcp* sent 63% of 408,675 of packets but uses port 7700 and protocol *UDP* connected with 76% of 4574 of destination hosts or ports.

The third host 207.89.143.152 is a scan host because it connects 58,816 hosts' port 1433. Note that the *support* and *diversity* value of the host are the same, 58,816, which implies the scan host only scans one time to each destination host.

7 Conclusion

This paper extends the theory of granule mining and applies the theory in network traffic characterization area. It characterizes the network traffic not only according to support and confidence, but also by diversity and condition diversity. Experiments performed in real network traffic have shown that the proposed method can effectively discover the significant hosts in networks. The rules discovered by the proposed method can provide perfectly interpretation of hosts' behaviors. Granule mining provides a promising methodology for knowledge discovery in databases.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of SIGMOD-93, pp. 207–216. Washington, DC, USA (1993)
2. Han, J., Cheng, H., Xin, D., Yan, X.: Frequent pattern mining: current status and future directions. Data Mining and Knowledge Discovery **15**(1), 55–86 (2007)
3. Li, Y.: Extended random sets for knowledge discovery in information system. In: Proceedings of the 9th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, pp. 524–532. Chongqing, China (2003)
4. Li, Y., Yang, W., Xu, Y.: Multi-tier granule mining for representations of multidimensional association rules. In: Proceedings of 6th IEEE International Conference on Data Mining, pp. 953–958 (2006)
5. Li, Y., Zhong, N.: Interpretations of association rules by granular computing. In: IEEE International Conference on Data Mining, pp. 593–596. Melbourne, Florida, USA (2003)
6. Li, Y., Zhong, N.: Mining rough association from text documents for web information gathering. Springer Transactions on Rough Sets **VII**, 103–119 (2006)
7. Pawlak, Z., , Skowron, A.: Rough sets and boolean reasoning. Information Sciences **177**, 41–73 (2007)
8. Pawlak, Z.: Rough sets. International Journal of Comput. Inf. Sci. **11**, 341–356 (1982)
9. Pawlak, Z.: Rough sets and intelligent data analysis. Information Sciences **147**(1-4), 1–12 (2002)
10. Pawlak, Z.: Decison trees and flow graphs. In: Proceedings of 5th International Conference on Rough Sets and Current Trends in Computing, pp. 1–11. Kobe, Japan (2006)
11. Skowron, A.: Rough sets and vague concepts. Fundamenta Informaticae **64**(1-4), 417–431 (2005)
12. Skowron, A.: Approximate reasoning in mas: rough set approach extended abstract (keynote talk). In: Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence (2006)
13. Yao, J.T., Yao, Y.Y.: Induction of classification rules by granular computing. In: Proceedings of the Third International Conference on Rough Sets and Current Trends in Computing, TSCTC '02, pp. 331–338. Springer-Verlag, London, UK, UK (2002)
14. Yao, Y.Y., Zhao, Y., Maguire, R.B.: Explanation oriented association mining using rough set theory. In: Proceedings of RSFD GrC 2003, pp. 165–172 (2003)
15. Zhao, Y., Yao, Y., Yao, J.: Level-wise construction of decision trees for classification. International Journal of Software Engineering and Knowledge Engineering IJSEKE **16**(1), 103–126 (2006)

How to Deal with Overgood and Underbad Alternatives in Bipolar Method

Tadeusz Trzaskalik and Sebastian Sitarz

Abstract. Bipolar is one of the Multiple Attribute Decision Making methods, based on the concept of Bipolar reference objectives, proposed by Konarzewska-Gubała. The essence of the analysis in the Bipolar method consists in a fact that the decision alternatives are not compared directly to each other, but they are confronted to the two sets of reference objects: desirable and non-acceptable. Practical application of the method showed some its shortcomings. It may happen that a decision alternative can be evaluated as better than a desirable reference object and simultaneously as worse than a non-acceptable object. The aim of the paper is to formulate modifications of the classical Bipolar approach to overcome such difficulties.

1 Introduction

Multiple Attribute Decision Making (MADM) is the approach which helps making decisions in terms of choosing, ranking or sorting alternatives, taking into account the pros and cons of a plurality of points of view. Considerable research has been devoted to support the decision maker in the evaluation of discrete set of alternatives. The state-of-art surveys have been made by many authors, recently by Tzeng and Huang [14] and Figueira, Greco and Ehrgott [2]. There are many well established methodologies which help the decision maker in her/his final choice.

One of the MADM methods is Bipolar, proposed by Konarzewska-Gubała [4,5,6,7,8]. The essence of the analysis in the Bipolar method consists in a fact that the alternatives are not compared directly to each other, but they are confronted to the two sets of reference objects: desirable (good) and non-acceptable (bad). These two separate sets constitute Bipolar reference system.

The Bipolar method consists of three phases. In the first phase alternatives are compared to reference objects and as a result outranking indicators and preference structure in the reference system are established. In the second phase position of

Tadeusz Trzaskalik
University of Economics in Katowice, Poland

Sebastian Sitarz
University of Silesia in Katowice, Poland

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 345–354.
springerlink.com

each alterative with regard to Bipolar reference system is established. In the third phase, on the basis of two mono-sortings of alternatives into specified categories and two partial preorders (mono-orders) introduced independently into the set of alternatives, the intersection of these two preorders, creating the Bipolar partial preorder is obtained. In the first phase we can recognize the ideas of concordance and veto thresholds, introduced in Electre methodology (see Roy [11], in the second phase - the idea of Merighi [9] algorithms of confrontation.

The Bipolar method has already been used in applications (Jakubowicz, Konarzewska-Gubała [4], Dominiak [1], Konarzewska-Gubała [5]. Moreover, the method has also been applied to model multi-stage multi-criteria decision processes (Trzaskalik [13]). The Bipolar method belongs to a group of methods that involve reference objects while compare alternatives. Other examples of this approach include for example bi-reference method (Michałowski and Szapiro [10]) and the method proposed by Skulimowski [12]. More recently learning methods, for instance DRSA, based on rough set methodology to derive classification rules has been developed (Greco, Matarazzo, Słowiński [3]).

Applying Bipolar method, some alternatives can be evaluated as “overgood”, i.e. better than at least one good object or “underbad”, i.e. worse than of one bad object. A question arises is it possible for an alternative to be overgood and underbad simultaneously? Konarzewska-Gubała [7] claims that if none bad object dominates any good object, such a situation cannot occur. Practical applications of Bipolar method (Dominiak [1] showed that condition described above is not sufficient for elimination such a possibility.

This paper aims at formulating modifications of the source Bipolar approach to avoid problems with overgood and underbad alternatives. Selection of the presented possibilities should be made together with a decision maker willing to apply the Bipolar approach.

The paper consists of four parts. Section 2 is a new, formal presentation of the source Bipolar method. Such a description seems to be necessary for presentation of proposed modifications. In section 3 modifications of Bipolar categories and reference system are proposed. Also an illustrative example is given. Concluding remarks in section 4 end the paper.

2 The Source Bipolar Method

It is assumed, that there are given: the set of decision alternatives $\mathbf{A} = \{\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^m\}$ and the set of criteria functions $F = \{f_1, \ldots, f_n\}$, where $f_k\colon \mathbf{A} \to \mathrm{K}_k$ for $k=1,\ldots n$, and K_k is a cardinal scale.

Criteria are defined in such a way that higher values are preferred to lower. Description of remaining types of criteria is given in Konarzewska-Gubała [8]. It is possible to transform them to the form, considered in the present paper. For each criterion the decision maker establishes weight w_k of relative importance (it is assumed, that $\sum_{k=1}^{n} w_k = 1$ and $w_k \geq 0$ for each $k=1, \ldots, n$), equivalence threshold q_k and veto threshold v_k. The decision maker also establishes minimal criteria values concordance level s as the outranking threshold. It is assumed, that condition $0.5 \leq s \leq 1$ holds.

The decision maker establishes a bipolar reference system $\mathbf{R} = \mathbf{G} \cup \mathbf{B}$, which consists of the set of good objects $\mathbf{G} = \{\mathbf{g}^1,\dots, \mathbf{g}^g\}$ and the set of bad objects $\mathbf{B} = \{\mathbf{b}^1,\dots,\mathbf{b}^b\}$, where g and b denote the number of good and bad objects, respectively. It is assumed, that $\mathbf{G} \cap \mathbf{B} = \varnothing$. The number of elements of the set $\mathbf{R}$ is equal to $g+b$. Elements of the set $\mathbf{R}$ are denoted as $\mathbf{r}^h$, $h=1,\dots,g+b$. Values $f_k(\mathbf{r}^h)$ for k=1,...,n and h=1,...,r are known. We assume, that the following condition holds:

$$\forall_{k=1,\dots,n} \forall_{\mathbf{g}\in \mathrm{G}} \forall_{\mathbf{b}\in \mathrm{B}} \quad f_k(\mathbf{g}) \geq f_k(\mathbf{b}) \tag{1}$$

2.1 Comparison of Alternatives to Reference Objects

2.1.1 Outranking Indicators

For the pair $(\mathbf{a}^i, \mathbf{r}^j)$, where $\mathbf{a}^i \in \mathbf{A}$, $\mathbf{r}^j \in \mathbf{R}$, the following values:

$$c^+(\mathbf{a}^i,\mathbf{r}^j) = \sum_{k=1}^{n} w_k \varphi_k^+(\mathbf{a}^i,\mathbf{r}^j) \quad \text{where} \quad \varphi_k^+(\mathbf{a}^i,\mathbf{r}^j) = \begin{cases} 1, & \text{if } f_k(\mathbf{a}^i) - f_k(\mathbf{r}^j) > q_k \\ 0, & \text{otherwise} \end{cases}$$

$$c^-(\mathbf{a}^i,\mathbf{r}^j) = \sum_{k=1}^{n} w_k \varphi_k^-(\mathbf{a}^i,\mathbf{r}^j) \quad \text{where} \quad \varphi_k^-(\mathbf{a}^i,\mathbf{r}^j) = \begin{cases} 1, & \text{if } f_k(\mathbf{r}^j) - f_k(\mathbf{a}^i) > q_k \\ 0, & \text{otherwise} \end{cases}$$

$$c^=(\mathbf{a}^i,\mathbf{r}^j) = \sum_{k=1}^{n} w_k \varphi_k^=(\mathbf{a}^i,\mathbf{r}^j) \quad \text{where} \quad \varphi_k^=(\mathbf{a}^i,\mathbf{r}^j) = \begin{cases} 1, & \text{if } \left|f_k(\mathbf{r}^j) - f_k(\mathbf{a}^i)\right| \leq q_k \\ 0, & \text{otherwise} \end{cases}$$

are calculated. Sets of indices:

$$I^+(\mathbf{a}^i,\mathbf{r}^j) = \{k : \varphi_k^+(\mathbf{a}^i,\mathbf{r}^j) = 1\} \qquad I^-(\mathbf{a}^i,\mathbf{r}^j) = \{k : \varphi_k^-(\mathbf{a}^i,\mathbf{r}^j) = 1\}$$

are determined.

Let v_k be threshold values given for k=1,...,n by the decision maker. Condition

$$\underset{k\in I^-}{\forall} f_k(\mathbf{a}^i) > v_k$$

is called veto test. Conditions

$$\underset{k\in I^-}{\forall} \quad f_k(\mathbf{a}^i) > v_k \qquad \underset{k\in I^+}{\forall} \quad f_k(\mathbf{a}^i) > v_k$$

are called non-discordance tests.

<u>Case 1</u>: $c^+(\mathbf{a}^i, \mathbf{r}^j) > c^-(\mathbf{a}^i, \mathbf{r}^j)$

- If for the pair $(\mathbf{a}^i, \mathbf{r}^j)$ veto test is positively verified, then outranking indicators are defined as follows:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = c^+(\mathbf{a}^i, \mathbf{r}^j) + c^=(\mathbf{a}^i, \mathbf{r}^j), \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = 0$$

- If for the pair $(\mathbf{a}^i, \mathbf{r}^j)$ veto test is not positively verified, then outranking indicators are defined as follows:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = 0, \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = 0$$

Case 2: $c^+(\mathbf{a}^i, \mathbf{r}^j) < c^-(\mathbf{a}^i, \mathbf{r}^j)$

•If for the pair $(\mathbf{a}^i, \mathbf{r}^j)$ veto test is positively verified, then outranking indicators are defined as follows:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = 0, \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = c^-(\mathbf{a}^i, \mathbf{r}^j) + c^=(\mathbf{a}^i, \mathbf{r}^j)$$

•If for the pair $(\mathbf{a}^i, \mathbf{r}^j)$ veto test is not positively verified, then outranking indicators are defined as follows:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = 0 \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = 0$$

Case 3: $c^+(\mathbf{a}^i, \mathbf{r}^j) = c^-(\mathbf{a}^i, \mathbf{r}^j)$.

• If for the pair $(\mathbf{a}^i, \mathbf{r}^j)$ two non-discordance tests are positively verified, then outranking indicators are defined as follows:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = c^+(\mathbf{a}^i, \mathbf{r}^j) + c^=(\mathbf{a}^i, \mathbf{r}^j) \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = c^-(\mathbf{a}^i, \mathbf{r}^j) + c^=(\mathbf{a}^i, \mathbf{r}^j)$$

• For the pair $(\mathbf{a}^i, \mathbf{r}^j)$ at least one of non-discordance tests is not positively verified, then outranking indicators are defined as follows:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = 0, \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = 0.$$

2.1.2 Preference Structure

By means of outranking indicators three relationships: large preference L_s, indifference I_s and incomparability R_s are defined as follows:

$\mathbf{a}^i\, L_s \mathbf{r}^h$ iff $d^+(\mathbf{a}^i, \mathbf{r}^h) > s \wedge d^-(\mathbf{a}^i, \mathbf{r}^h) = 0$ $\qquad$ $\mathbf{r}^j L_s \mathbf{a}^i$ iff $d^+(\mathbf{a}^i, \mathbf{r}^h) = 0 \wedge d^-(\mathbf{a}^i, \mathbf{r}^h) > s$
$\mathbf{a}^i\, I_s \mathbf{r}^h$ iff $d^+(\mathbf{a}^i, \mathbf{r}^h) > s \wedge d^-(\mathbf{a}^i, \mathbf{r}^h) > s$ $\qquad$ $\mathbf{a}^i\, R_s\, \mathbf{r}^j$ otherwise

2.2 *Position of an Alternative in Relation to the Bipolar Reference System*

2.2.1 Success Achievement Degree

For a given $a^i \in A$ auxiliary sets of indices are defined as follows:

$L_s(\mathbf{a}^i, \mathbf{G}) = \{h: \mathbf{a}^i L_s \mathbf{g}^h, \mathbf{g}^h \in \mathbf{G}\}$ $\qquad$ $I_s(a^i, \mathbf{G}) = \{h: \mathbf{a}^i I_s \mathbf{g}^h, \mathbf{g}^h \in \mathbf{G}\}$
$L_s(\mathbf{G}, \mathbf{a}^i) = \{h: \mathbf{g}^h\, L_s\, \mathbf{a}^i, \mathbf{g}^h \in \mathbf{G}\}$

In the set $L_s(\mathbf{a}^i, \mathbf{G})$ there are included these indices of good objects, for whom the statement $\mathbf{a}^i L_s \mathbf{g}^h$ is true. The two remaining sets are defined similarly.

Defining the position of an alternative $\mathbf{a}^i$ in relation to the set $\mathbf{G}$ we consider three possibilities:

Case S1. $L_s(\mathbf{a}^i, \mathbf{G}) \cup I_s(\mathbf{a}^i, \mathbf{G}) \neq \emptyset$.

The value $d_{\mathbf{G}}^+(a^i) = \max\{d^+(\mathbf{a}^i, \mathbf{g}^h): h \in L_s(\mathbf{a}^i, \mathbf{G}) \cup I_s(\mathbf{a}^i, \mathbf{G})\}$
is calculated. The success achievement degree $d_S(\mathbf{a}^i)$ is defined to be equal to $d_G^+(\mathbf{a}^i)$.

Case S2. $L_s(\mathbf{a}^i, \mathbf{G}) \cup I_s(\mathbf{a}^i, \mathbf{G}) = \emptyset \wedge L_s(\mathbf{G}, \mathbf{a}^i) \neq \emptyset$.

The value $d_{\mathbf{G}}^-(\mathbf{a}^i) = \min \{d^-(\mathbf{a}^i, \mathbf{g}^h): h \in L_s(\mathbf{G}, \mathbf{a}^i)\}$
is calculated. The success achievement degree $d_S(\mathbf{a}^i)$ is defined to be equal to $d_{\mathbf{G}}^-(\mathbf{a}^i)$.

Case S3.

If conditions described in Cases S1 and S2 are not fulfilled, then the success achievement degree $d_S(\mathbf{a}^i)$ is defined to be equal to 0.

2.2.2 Failure Avoidance Degree

For a given $\mathbf{a}^i \in \mathbf{A}$ auxiliary sets of indices are defined as follows:

$$L_s(\mathbf{B}, a^i) = \{h: \mathbf{b}^h L_s \mathbf{a}^i, \mathbf{b}^h \in \mathbf{B}\}$$
$$I_s(\mathbf{B}, a^i) = \{h: \mathbf{b}^i I_s \mathbf{a}^h, \mathbf{b}^h \in \mathbf{B}\}$$
$$L_s(a^i, \mathbf{B}) = \{h: \mathbf{a}^i L_s \mathbf{b}^h, \mathbf{b}^h \in \mathbf{B}\}$$

In the set $L_s(\mathbf{B}, \mathbf{a}^i)$ there are included these numbers of bad objects, for whom the statement $\mathbf{b}^h L_s \mathbf{a}^i$ is true. The two remaining sets are interpreted similarly.

Defining the position of an alternative $\mathbf{a}^i$ in relation to the set $\mathbf{B}$ we consider three possibilities:

Case N1. $L_s(\mathbf{B}, \mathbf{a}^i) \cup I_s(\mathbf{B}, \mathbf{a}^i) = \emptyset \wedge L_s(\mathbf{a}^i, \mathbf{B}) \neq \emptyset$.

The value $d_{\mathbf{B}}^+(\mathbf{a}^i) = \min \{d^+(\mathbf{a}^i, \mathbf{b}^h): h \in L_s(\mathbf{a}^i, \mathbf{B})\}$ is calculated. The failure avoidance degree $d_F(a^i)$ is defined to be equal to $d_B^+(a^i)$.

Case N2. $L_s(\mathbf{B}, \mathbf{a}^i) \cup I_s(\mathbf{B}, \mathbf{a}^i) \neq \emptyset$.

The value $d_{\mathbf{B}}^-(\mathbf{a}^i) = \max \{d^-(\mathbf{a}^i, \mathbf{b}^h): h \in L_s(B, a^i) \cup I_s(a^i, D)\}$ is calculated. The failure avoidance degree $d_F(a^i)$ is defined to be equal to $d_B^-(a^i)$.

Case N3.
If conditions described in Cases S1 and S2 are not fulfilled, then the failure avoidance degree $d_F(\mathbf{a}^i)$ is defined to be equal to 0.

2.2.3 Mono-sortings and Mono-rankings

According to the success achievement degree the alternatives from the set $\mathbf{A}$ are sorted to the three categories:

Category S1 consists of the „overgood" alternatives, for which condition, formulated in Case S1 is fulfilled.

Category S2 consists of the alternatives, for which condition, formulated in Case S2 is fulfilled.

Category S3 consists of the alternatives, for which condition, formulated in Case S3 is fulfilled (decision variants non-comparable with $\mathbf{G}$).

A way of building above categories implies that each alternative from the Category S1 should be preferred to any alternative from the Category S2. Alternatives are ordered as follows:

$\mathbf{a}^i$ is preferred to $\mathbf{a}^j$ (i≠j, i,j=1,...,m), iff
$[d_S(\mathbf{a}^i) = d_G^+(\mathbf{a}^i) \wedge d_S(\mathbf{a}^j) = d_G^+(\mathbf{a}^j) \wedge (d_G^+(\mathbf{a}^i) > d_G^+(\mathbf{a}^j)] \vee [d_S(\mathbf{a}^i) = d_G^+(\mathbf{a}^i) \wedge d_S(\mathbf{a}^j) = d_G^-(\mathbf{a}^j)] \vee [d_S(\mathbf{a}^i) = d_G^-(\mathbf{a}^i) \wedge d_S(\mathbf{a}^j) = d_G^-(\mathbf{a}^j) \wedge d_G^-(\mathbf{a}^i) < d_G^-(\mathbf{a}^j)]$

a^i is equivalent to a^j, iff
$[d_S(\mathbf{a}^i) = d_G^+(\mathbf{a}^i) \wedge d_S(\mathbf{a}^j) = d_G^+(\mathbf{a}^j) \wedge d_G^+(\mathbf{a}^i) = d_G^+(\mathbf{a}^j)] \vee [d_S(\mathbf{a}^i) = d_G^-(\mathbf{a}^i) \wedge d_S(\mathbf{a}^j) = d_G^-(\mathbf{a}^j) \wedge d_G^-(\mathbf{a}^i) = d_G^-(\mathbf{a}^j)]$

According to the failure avoidance degree the alternatives from the set **A** are sorted to the three categories:

Category F1 consists of the alternatives, for which condition, formulated in Case N1 is fulfilled.

Category F2 consists of underbad alternatives, for which condition, formulated in Case N2 is fulfilled.

Category F3. consists of the alternatives, for which condition, formulated in Case N3 is fulfilled (alternatives non-comparable with **B**).

A way of building above categories implies that each alternative from the Category N1 should be preferred to any alternative from Category N2. Alternatives are ordered as follows:

$\mathbf{a}^i$ is preferred to $\mathbf{a}^j$, iff
$[d_F(\mathbf{a}^i) = d_B^+(\mathbf{a}^i) \wedge d_F(\mathbf{a}^j) = d_B^+(\mathbf{a}^j) \wedge d_B^+(\mathbf{a}^i) > d_B^+(\mathbf{a}^j)] \vee [d_F(\mathbf{a}^i) = d_B^+(\mathbf{a}^i) \wedge d_F(\mathbf{a}^j) = d_B^-(\mathbf{a}^j)] \vee [d_F(\mathbf{a}^i) = d_B^-(\mathbf{a}^i) \wedge d_F(\mathbf{a}^j) = d_B^-(\mathbf{a}^j) \wedge d_B^-(\mathbf{a}^i) < d_B^-(\mathbf{a}^j)]$

$\mathbf{a}^i$ is equivalent to $\mathbf{a}^j$, iff
$[d_F(\mathbf{a}^i) = d_B^+(\mathbf{a}^i) \wedge d_F(\mathbf{a}^j) = d_B^+(\mathbf{a}^j) \wedge d_B^+(\mathbf{a}^i) = d_B^+(\mathbf{a}^j)] \vee [d_F(\mathbf{a}^i) = d_B^-(\mathbf{a}^i) \wedge d_F(\mathbf{a}^j) = d_B^-(\mathbf{a}^j) \wedge d_B^-(\mathbf{a}^i) = d_B^-(\mathbf{a}^j)]$

2.2.4 Bipolar-Sorting and Bipolar-Ranking

Considering jointly evaluation of success achievement degree and failure avoidance degree, three categories of alternatives are defined:

Category B1 consists of such alternatives $\mathbf{a}^i$, that $d_G^+(\mathbf{a}^i) > 0 \wedge d_B^+(\mathbf{a}^i) > 0$

Category B2 consists of such alternatives $\mathbf{a}^i$, that $d_G^-(\mathbf{a}^i) > 0 \wedge d_B^+(\mathbf{a}^i) > 0$

Category B3 consists of such alternatives $\mathbf{a}^i$, that $d_G^-(\mathbf{a}^i) > 0 \wedge d_B^-(\mathbf{a}^i) > 0$

Assuming, that each alternative from the Category B1 is preferred to any alternative from Category B2 and each alternative from Category B2 is preferred to any alternative from Category B3, linear order is given in each category as follows:

For $\mathbf{a}^i, \mathbf{a}^j \in$ B1 $\quad \mathbf{a}^i$ is preferred to $\mathbf{a}^j$, iff $d_S(\mathbf{a}^i)+ d_F(\mathbf{a}^i) > d_S(\mathbf{a}^j)+d_F(\mathbf{a}^j)$
$\mathbf{a}^i$ is equivalent to $\mathbf{a}^j$, iff $d_S(\mathbf{a}^i)+d_F(\mathbf{a}^i) = d_S(\mathbf{a}^j)+d_F(\mathbf{a}^j)$

For $\mathbf{a}^i, \mathbf{a}^j \in$ B2 $\quad \mathbf{a}^i$ is preferred to $\mathbf{a}^j$, iff $1-d_S(\mathbf{a}^i)+d_F(\mathbf{a}^i) > 1-d_S(\mathbf{a}^j)+d_F(\mathbf{a}^j)$
$\mathbf{a}^i$ is equivalent to $\mathbf{a}^j$, iff $1-d_S(\mathbf{a}^i)+d_F(\mathbf{a}^i) = 1-d_S(\mathbf{a}^j) +d_F(\mathbf{a}^j)$

For $\mathbf{a}^i, \mathbf{a}^j \in$ B3 $\quad \mathbf{a}^i$ is preferred to $\mathbf{a}^j$, iff $d_S(\mathbf{a}^i) + d_F(\mathbf{a}^i) < d_S(\mathbf{a}^j) + d_F(\mathbf{a}^j)$
$\mathbf{a}^i$ is equivalent to $\mathbf{a}^j$, iff $d_S(\mathbf{a}^i) + d_F(\mathbf{a}^i) = d_S(\mathbf{a}^j)+ d_F(\mathbf{a}^j)$.

3 Modifications of Reference Sets and Categories

Up to now we have assumed that condition (1) holds. Let D be classical domination relation, defined for any $\mathbf{x}$, $\mathbf{y}$ in the following way:

$$f(\mathbf{x})\ \mathrm{D}\ f(\mathbf{y}) \Leftrightarrow \forall_{k=1,..n} f_k(\mathbf{y}) \le f_k(\mathbf{x}) \wedge \exists_{l=1,..n} f_l(\mathbf{y}) < f_l(\mathbf{x}).$$

Now we will assume, that the following condition holds:

$$\sim \exists_{\mathbf{g}\in \mathbf{G}} \sim \exists_{\mathbf{b}\in \mathbf{B}} f(\mathbf{b})\ \mathrm{D}\ f(\mathbf{g}) \tag{2}$$

Example 1

Let us consider the set of alternatives $\mathbf{A}=\{\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3, \mathbf{a}^4\}$ such, that $f(\mathbf{a}^1)=(1,5)$, $f(\mathbf{a}^2)=(3,5)$, $f(\mathbf{a}^3)=(4,3)$, $f(\mathbf{a}^4)=(5,2)$ and the reference system, which consists of the sets: $\mathbf{G} = \{\mathbf{g}^1, \mathbf{g}^2, \mathbf{g}^3\}$: $f(\mathbf{g}^1)=(0,6)$, $f(\mathbf{g}^2)=(5,6)$, $f(\mathbf{g}^3)=(5,0)$ and $\mathbf{B} = \{\mathbf{b}^1, \mathbf{b}^2\}$: $f(\mathbf{b}^1)=(0,2)$, $f(\mathbf{b}^2)=(2,0)$. Condition (1) does not hold, because $f_1(\mathbf{g}^1)<f_1(\mathbf{b}^2)$ and $f_2(\mathbf{g}^3)< f_2(\mathbf{b}^1)$, but condition (2) is fulfilled. Assuming, that veto thresholds are equal to $v_1= 0$, $v_2 = 0$, weights are equal to $w_1 = 0{,}6$, $w_2 = 0{,}4$, and the concordance threshold and the equivalence threshold are equal to $s=0{,}5$ and $q_1=0{,}25$, $q_2=0{,}25$, we apply classical Bipolar procedure, and obtain the following Bipolar ranking: 1: $\mathbf{a}^4$ (category B1), 2: $\mathbf{a}^2$, $\mathbf{a}^3$ (category B1). We have $d_{\mathbf{G}}^+(\mathbf{a}^1) > 0$ and $d_{\mathbf{B}}^-(\mathbf{a}^1) > 0$. Alternative $\mathbf{a}^1$ is both overgood and underbad and cannot be classified to any previously defined bipolar categories. That is why it is not included in the final bipolar ranking.

3.1 Modification of the Reference System

We want to modify the reference system to assure that condition (1) will be fulfilled.

3.1.1 Modification of the Reference Set of Good Objects

Let $\hat{f}^{\mathrm{B}}$ denote the ideal vector in the reference set of bad objects, hence

$$\hat{f}_k^{\mathrm{B}} = \max \{ f_k(\mathbf{b}):\ \ \mathbf{b}\in \mathbf{B}\}$$

We replace the set $\mathbf{G}$ by the set $\overline{\mathbf{G}} = \{ \overline{\mathbf{g}}^1, \overline{\mathbf{g}}^2 ..., \overline{\mathbf{g}}^g\}$, changing these evaluations, which are too low according to ideal solutions in the set $\mathbf{B}$, that is

$$f_k\left(\overline{\mathbf{g}}\right)=\begin{cases} f_k(\mathbf{g}) & \text{if } f_k(\mathbf{g})\geq \hat{f}_k^{\mathrm{B}} \\ \hat{f}_k^{\mathrm{B}} & \text{if } f_k(\mathbf{g})< \hat{f}_k^{\mathrm{B}} \end{cases}$$

Let us notice, that if $f(\mathbf{g})$ dominates all bad objects, then $f_k(\overline{\mathbf{g}})) = f_k(\mathbf{g})$ for k=1,…,n. In the opposite case we increase the appropriate components of vector $f(\mathbf{g})$.

Example 2

Let us consider a problem described in the Example 1. Applying proposed modification, we obtain:

$$\hat{f}_1^{\mathrm{B}} = \max\{0, 2\} = 2 \qquad\qquad \hat{f}_2^{\mathrm{B}} = \max\{2, 0\} = 2$$

We create $\mathbf{G} = \{\overline{\mathbf{g}}^1, \overline{\mathbf{g}}^2, \overline{\mathbf{g}}^3\}$. The evaluations of new reference good objectives are as follows: (2,6), (5,6), (5,2). We apply classical Bipolar method with modified categories and we obtain the following ranking: 1. a_4 (category B1), 2. a_2, a_3 (category B1), 3. a_1 (category B2).

3.1.2 Modification of the Reference Set of Bad Objects

Let $\check{f}^{\mathbf{G}}$ denote nadir vector in the reference set of good objects, hence

$$\check{f}_k^{\mathbf{G}} = \min\{f_k(\mathbf{g}):\ \ \mathbf{g}\in\mathbf{G}\}$$

We replace the set $\mathbf{B}$ by the set $\overline{\mathbf{B}} = \{\overline{\mathbf{b}}^1,\ldots, \overline{\mathbf{b}}^b\}$, changing these evaluations, which are too high according to nadir solution in the set D, that is

$$f_k\left(\overline{\mathbf{b}}\right)=\begin{cases} f_k(\mathbf{b}) & \text{if } f_k(\mathbf{b})\geq \check{f}_k^{\mathrm{G}} \\ \check{f}_k^{\mathrm{G}} & \text{if } f_k(\mathbf{b})< \check{f}_k^{\mathrm{G}} \end{cases}$$

Let us notice, that if $f(\mathbf{b})$ is dominated by all bad objects, then $f(\overline{\mathbf{b}}) = f(\mathbf{z})$. In the opposite case we decrease the appropriate components of vector $f(\mathbf{b})$.

Example 3

Let us consider a problem described in Example 1. Applying proposed modification, we obtain: $\check{f}_1^{\mathbf{G}} = \min\{0,5,5\} = 0$, $\check{f}_2^{\mathbf{G}} = \min\{0, 6, 6\} = 0$. We create $\mathbf{Z} = \{\overline{\mathbf{z}}^1, \overline{\mathbf{z}}^2\}$. The evaluation of new reference bad objects are as follows: $\check{f}^{\mathrm{G}} = f(\overline{\mathbf{b}}^1) = f(\overline{\mathbf{b}}^2)=(0,0)$. We apply classical Bipolar method with modified categories and we obtain the following ranking: 1. a_4 (category B1), 2. a_2, a_3 (category B1), 3. a_1 (category B2).

3.2 Modification of Categories in Bipolar Ranking

The aim of this modification is to create additional category B2' including all overgood and underbad alternatives. Alternative $\mathbf{a}^i$ belongs to the category B2', iff

$$d_G^+(\mathbf{a}^i) > 0 \qquad \wedge \qquad d_B^-(\mathbf{a}^i) > 0$$

Category B1 includes alternatives better than decision variants classified to remaining categories. Category B3 includes alternatives worse than alternatives classified to remaining categories. Categories B2 and B2' are not comparable.

Treating categories B2 and B2' symmetrically, we propose the following procedure of ranking in the set B2 $\cup$B2':

For $\mathbf{a}^i, \mathbf{a}^j \in$ B2: $\mathbf{a}^i$ is preferred to $\mathbf{a}^j$ iff $1-d_S(\mathbf{a}^i)+d_F(\mathbf{a}^i) > 1-d_S(\mathbf{a}^j)+d_F(\mathbf{a}^j)$
a^i is equivalent to $\mathbf{a}^j$ iff $1-d_S(\mathbf{a}^i)+d_F(\mathbf{a}^i) = 1-d_S(\mathbf{a}^j)+d_F(\mathbf{a}^j)$

For $\mathbf{a}^i, \mathbf{a}^j \in$ B2': $\mathbf{a}^i$ is preferred to $\mathbf{a}^j$, iff $1+d_S(\mathbf{a}^i) - d_F(\mathbf{a}^i) > 1+d_S(\mathbf{a}^j)-d_F(\mathbf{a}^j)$
$\mathbf{a}^i$ is equivalent to $\mathbf{a}^j$ iff $1+d_S(\mathbf{a}^i)-d_F(\mathbf{a}^i) = 1+d_S(\mathbf{a}^j)-d_F(\mathbf{a}^j)$

For $\mathbf{a}^i \in$ B2, $\mathbf{a}^j \in$ B2': $\mathbf{a}^i$ is preferred to $\mathbf{a}^j$ iff $1-d_S(\mathbf{a}^i) + d_F(\mathbf{a}^i) > 1+d_S(\mathbf{a}^j)-d_F(\mathbf{a}^j)$
$\mathbf{a}^i$ is equivalent to $\mathbf{a}^j$ iff $1-d_S(\mathbf{a}^i)+d_F(\mathbf{a}^i) = 1+d_S(\mathbf{a}^j)-d_F(\mathbf{a}^j)$

For $\mathbf{a}^i \in$ B2', $\mathbf{a}^j \in$ B2: a^i is preferred to $\mathbf{a}^j$ iff $1+d_S(\mathbf{a}^i) - d_F(\mathbf{a}^i) > 1-d_S(\mathbf{a}^j)+d_F(\mathbf{a}^j)$
a^i is equivalent to $\mathbf{a}^j$ iff $1+d_S(\mathbf{a}^i)-d_F(\mathbf{a}^i) = 1-d_S(\mathbf{a}^j)+d_F(\mathbf{a}^j)$.

Example 4

Let us consider the problem described in Example 1 ($f(\mathbf{g}^1)=(0,6)$, $f(\mathbf{g}^2)=(5,6)$, $f(\mathbf{g}^3)=(5,0)$, $f(\mathbf{b}^1)=(0,2)$, $f(\mathbf{b}^2)=(2,0)$). We apply classical Bipolar method with modified categories and obtain the following ranking: 1.a_4 (category B1), 2. a_2, a_3 (category B1), 3. a_1 (category B2').

It is seen that alternative $\mathbf{a}^1$ which cannot be classified applying classical Bipolar procedure belongs to class B2 or B2' after proposed modifications.

4 Concluding Remarks

Modifications of the source version of the Bipolar method, suggested in section 3 allow both for rationalizing Bipolar incomparability of some alternatives and elaborating of a ranking.

The above presented modifications of the Bipolar method refer to a situation when all the considered criteria are presented on the cardinal scale. Further research may be directed towards subsequent modifications of the method in case the criteria in question are presented on the ordinal or binary scales.

Acknowledgement. This work is supported by the grant from Polish Ministry of Science and Higher Education, project number NN 111 2350 36.

References

1. Dominiak, C.: Portfolio selection using the idea of reference solution. In: Fandel, G., Gal, T. (eds.) Multiple Criteria Decision Making, pp. 593–602. Springer, Heidelberg (1997)
2. Figueira, J., Greco, S., Ehrgott, M. (eds.): Multiple criteria de-cision analysis: States of the art surveys. Springer (2005)
3. Greco, S., Matarazzo, B., Słowinski, R.: Rough set methodol-ogy for sorting problems in presence of multiple attributes and criteria. EJOR 138, 247–259 (2002)
4. Jakubowicz, S., Konarzewska-Gubała, E.: Work characteris-tics of a physics teacher. University of Wrocław, Copied manuscript (1989) (in Polish)
5. Konarzewska-Gubała, E.: Multiple criteria company ben-chmarking using the BIPOLAR Method. In: Trzaskalik, T., Michik, J. (eds.) Multiple Objective and Goal Programming. Recent Developments, pp. 338–350. Springer, Heidelberg (2002)
6. Konarzewska-Gubała, E.: Supporting an effective perfor-mance appraisal system. Argumenta Oeconomica 1, 123–125 (1996)
7. Konarzewska-Gubała, E.: Multiple criteria decision aid: system Bipolar. University of Economics in Wrocław (1991) (in Polish)
8. Konarzewska-Gubała, E.: BIPOLAR: Multiple criteria decision aid using Bipolar reference system. LAMSADE, Cahier et Documents 56, Paris (1989)
9. Merighi, D.: Un modello di valutazione rispetto insiemi di riferimento assegnati. Ricerca Operativa 13, 31–52 (1980)
10. Michałowski, W., Szapiro, T.: A bi-reference procedure for interactive multiple criteria programming. Operations Research 40(2), 247–258 (1992)
11. Roy, B.: Methodologie multicritere d'aide a la decision. Economica, Paris (1985)
12. Skulimowski, A.: Decision support systems based on reference sets. AGH, Krakow (1996)
13. Trzaskalik, T.: Model of multistage multicriteria decision processes applying reference sets, vol. 413, pp. 73–93. University of Economics in Wrocław (1987) (in Polish)
14. Tzeng, G.-H., Huang, J.-H.: Multiple attribute decision making. Methods and applications. Taylor & Francis (2011)

Improved Defect Classification of Printed Circuit Board Using SVM

Yuji Iwahori, Deepak Kumar, Takuya Nakagawa, and M.K. Bhuyan

Abstract. This paper proposes an analytical approach to make correct recognition between the true defect and the pseudo defect with visual inspection of the electronic board. Some classification approaches have already been proposed for the limited kinds of defects but there have been incorrect recognitions for the defects which is difficult to handle with the visual inspection. This paper proposes the approach to reduce the incorrect recognition for the defects using Support Vector Machine. Real electronic board image data are tested and evaluated with the proposed approach. It is shown that the proposed approach gives efficient classification with some new features with a proper analysis on histogram of image according to the proposed evaluation criteria for the performance.

1 Introduction

Basically, PCB is a piece of phenolic or glass epoxy board with copper clad on one or both sides. The portion of copper that are not needed are etched off, leaving 'printed' circuits which connects the components. It is used to mechanically support and electrically connect electronic components using conductive pathways, or traces, etched from copper sheets and laminated onto a non-conductive substrate. PCBs are rugged, inexpensive and highly reliable and so it is used in virtually all but the simplest commercially produced electronic devices.

It has become one of the basic components of electronic devices. It provides the electrical connections between the electronic or IC components mounted on it. In recent years, the demand of electronic devices with more compact design and more sophisticated functions has forced the PCBs to become smaller and denser with circuits and components. As it is crucial part of electronic device it needs to be

Yuji Iwahori · Takuya Nakagawa
Dept. of Computer Science, Chubu University, 1200 Matsumoto-cho,
Kasugai 487-8501, Japan
e-mail: {iwahori,tnakagwa}@cs.chubu.ac.jp

Deepak Kumar · M.K. Bhuyan
Indian, Institute of Technology, Guwahati - 781039, India
e-mail: {kumar,mkb}@iitg.ernet.in

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 355–363.
springerlink.com

properly investigated before get launched. Automatic inspection systems are used for this purpose but due to more complexity in circuits, PCB inspections are now more problematic. This problem leads to new challenges in developing advanced automatic visual inspection systems for PCB.

Automatic Optical Inspection (AOI) has been commonly used to inspect defects in Printed circuit board during the manufacturing process. An AOI system generally uses methods which detects the defects by scanning the PCB board and analysing it. AOI uses methods like Local Feature matching, image Skeletonization and morphological image comparison to detect defects and has been very successful in detecting defects in most of the cases but production problems like oxidation, dust, contamination and poor reflecting materials leads to most inevitable false alarms. To reduce the false alarms is the concern of this paper.

Previous approach [2] classifies the defects using neural network and [3] proposes a method to classify the defects using the intensity at the pixels around the defects region. These approaches classify the defect under the condition that kinds of the defects are previously known. There are some defects whose recognitions are difficult even with the visual inspection. These defects cause the problem. The problem includes the case of misjudgement where a true defect is recognized as a pseudo defect and it is included in the products as a result. Another approach [4] has been proposed for the distinction of defect classification by determining the features at random. This approach [4] classifies the kinds of defect with selecting the appropriate features with classifiers, but there are still incorrect classification cases where a true defect is classified into a pseudo defect. This paper proposes a new classification approach to the recognition which considerably improves the correct recognition to 94%. The paper improves the work done in the paper [6] and purposes a new approach to improve the performance by proper feature selection.

2 Prerequisites

2.1 Kinds of Defect

PCB manufacturing processes are based on chemical and mechanical actions, so sometimes it gets into a situation where these actions damage the intended design. The common defects that PCB encounters are excessive current leakage (caused by conducting paths between traces), open circuit (caused by excessive strain like the vibration and torsion), fatigue cracks and corrosion, purity (of semiconductor) based defects, electrical faults etc. These defects can be further classified into many tiny defects. Some of those are cuts, protrusions,oxidation, weak or wear rust. The defects can be true or pseudo depending on the cause of having defects. When some dust gets aggregated on the surface, it looks like a defect in real but it is not actually a defect, hence categorised as pseudo defect.

True defects and pseudo defects need to be distinguished. The distinction is important as we cannot allow the items with true defect to get launched in the market and if we know items with pseudo defect, we can allow it after cleaning. This improves the quality of product.

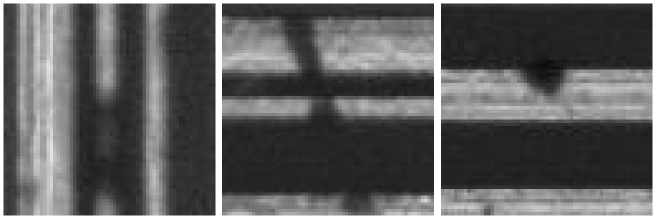


Fig. 1 True Defects

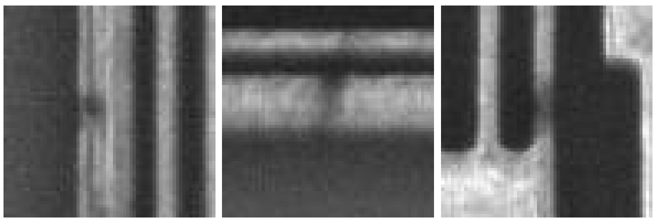


Fig. 2 Pseudo Defects

2.2 Learning Data

The data used for the approach were gathered manually by the visual inspection of defects. The data set were divided into two parts 1) True defect 2) pseudo defect. There were 93 samples corresponding to true defect and 57 samples corresponding to pseudo defect. The Classifier were trained using these data set.

2.3 Support Vector Machine

SVM has an ability to classify the data into two classes with the maximization of the margin. It determines the support vector from the learning sample data of each class and determine the discrimination boundary in the multiple dimensional feature space. Two classes are defined as true defect class and pseudo defect class in this case. The data which is input into the region inside the margin of SVM are treated as the difficult judgement data which may belong to either of two classes.

The discrimination function is written as

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \tag{1}$$

Transforming Eq.(1) in the form of kernel function gives

$$y(\mathbf{x}) = \sum_{n \in S} a_n t_n K(\mathbf{x}, \mathbf{x}_n) + b \tag{2}$$

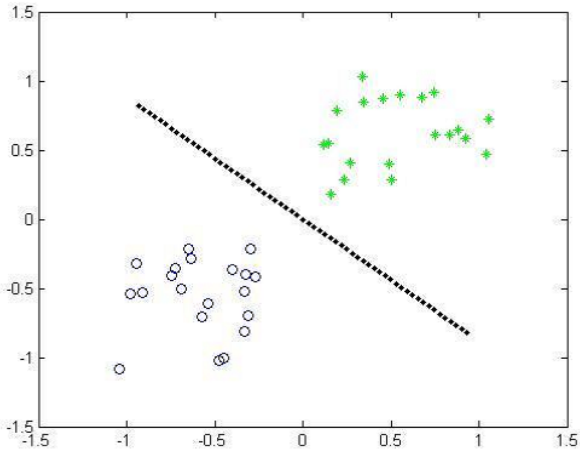


Fig. 3 Classification with SVM

$\mathbf{x}$ represents the input data, $\mathbf{x}_n$ represents the support vector and a represents Lagrange multipliers. t represents the teaching signal, K represents the Kernel function, b represents the threshold value and S represents the set of support vector.

When the output value takes larger value than $+1$, the data is classified into the class A, while when the output value takes less than -1, the data is classified into class B. While the value is between -1 and $+1$, the data is projected to the region inside the margin.

Kernel trick makes it possible to realize the nonlinear classification and the equation used in the approach is as follows.

$$K(\mathbf{x}, \mathbf{x}_n) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{\sigma^2}\right) \tag{3}$$

Gaussian kernel is generally used and it is applied to Eq.(2).

3 Improved Classification

3.1 Analysis of Defect

In this section,we subtracted the defect image with the corresponding reference image. As the defected and defect free image is available, the subtraction does not need any prior processing. This is done considering the fact that these two images are properly aligned. The approach of getting defect image was mentioned in paper [6].

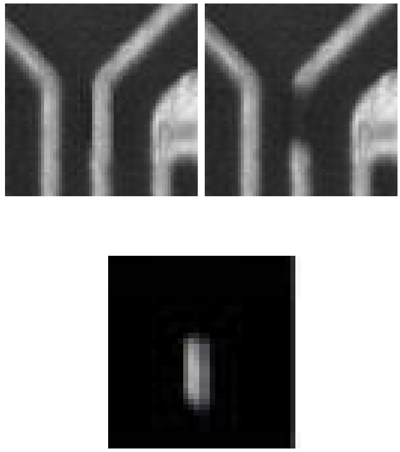


Fig. 4 Reference Image (top Left side) and Defected Image (top right side) and Candidate Region (bottom)

After subtraction, cleaning is performed on the image then the statistical analysis of the image histogram is done. It has been observed during the experiment that the gray level ratio, i.e. the ratio of number of pixels having pixel value greater than 70 to the number of pixels having pixel value lying between 20 and 70 is more in true defect as compared to pseudo defect. This property helped to improve the recognition accuracy in the further experiment. Here are some samples of the histogram.

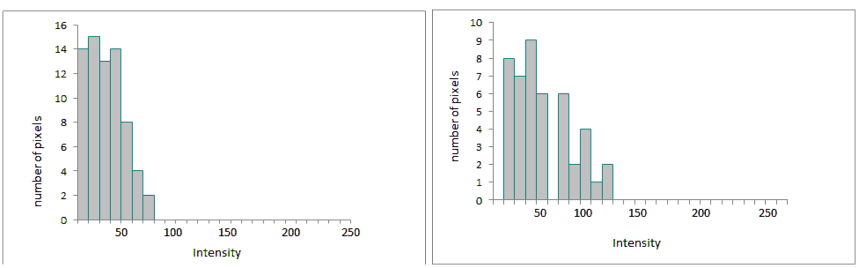


Fig. 5 (a) Histogram of True Defect (b) Histogram of Pseudo Defect

3.2 Feature Extraction

On the image obtained after the subtraction, some spatial domain and intensity domain feature extraction were implemented. We extracted nine shape features or spatial features and seven texture features of the defected image. The shape features that we used are 1) complexity; 2) area of the defect; 3)circularity; 4) aspect ratio; 5) minor axis length; 6) major axis length; 7) perimeter of the defect; 8) diameter of the defect; 9) Euler number. While the texture features were 1) average gray level; 2) standard deviation; 3) smoothness; 4) third moment; 5) uniformity; 6) entropy; 7) gray level ratio. After extracting the features, the features were organised in a matrix where each row represents an observation and each column represents a feature and then it is mean normalized to remove any biasing in the data. The data set is now cross validated to randomly create train data and test data. The classifier was modelled with the training set. We used here SVM classifier. The test data generated through cross validation were passed to the classifier for classification. The features used for the classification were selected sequentially using sequential feature selection.

3.3 Parameter Selection

Here we used RBF kernel function for the SVM classification. The kernel function has a parameter Y which depends on σ. The effectiveness of the SVM depends on the selection of the kernel function, kernel parameters and soft margin parameter C. Best combination of C and Y is selected by a grid search with exponentially growing sequences of C and Y.As Y depends on sigma, we used combinations of C and σ and we took C=(2^{-5},2^{-3},2^{-1},1,2,8,32) and σ=(8,4,2,1,0.5,.25,.125). The best combination were obtained through the iterative procedure. In this experiment, we got the better combination of C and sigma as σ=(1,1);(8,1.2);(2,0.7);(1.2,0.6).

Out of these combinations,the best result were obtained when we chose the value (C, σ) as (1, 1) and then parameter C is fixed and values of σ is varied from 0.1 to 1.5 with a step of 0.1 as to see the influence of sigma parameter. During the experiment, almost more than 94% accuracy were achieved. Here we evaluated our performance using different combination of features and results revel that classification accuracy depends more on texture information rather than shape information. The wrongly detected defect were analysed with the optimum feature combination.

4 Experiment

We took 93 samples of true defect and 57 samples of pseudo defects for our experiments. The choice of box-constraint and parameter σ was made through iterative procedure. We used forward sequential selection method to decide the optimum features for experimentation. Considering the box-constraint C equal we varied the parameter of σ for 0.1 to 1.5 with the step of 0.1 and calculated the accuracy of the classification. The Fig.6 illustrates the variation of recognition accuracy with parameter σ. Here we can see that the maximum accuracy obtained when the sigma were close to 0.8.

The combination that gave the maximum accuracy were a) gray level ratio; b) smoothness of the defect; c) standard deviation; d) entropy. In Table 1, we can see the top optimum combinations with the corresponding recognition accuracy. Here, AGL=Average gray level ratio, SD=standard deviation, SM=smoothness, GLR=gray level ratio, E=entropy.

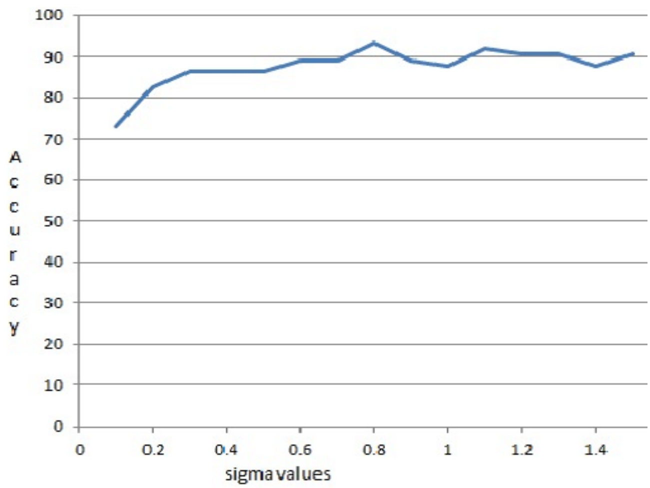


Fig. 6 Accuracy Graph

Table 1 Accuracy of Classifier at Different Values of σ

Features	Accuracy Obtained(%)
AGL	82.43
SD	85.13
SM	83.78
AGL,GLR	83.78
TM,GLR	86.48
E,GLR	87.83
TM,U,GLR	89.18
SM,U,GLR	90.54
GLR,SD,SM,E	94.59
SD,SM,E	91.89

Further we used our detection process to calculate the performance. The purpose of this parameter is to evaluate the efficiency of the inspection process.This parameter is required to know how much percentage of bare PCB produced can be used further for component fabrication. The performance has been defined as

$$\mathbf{performance} = (1 - \mathbf{TP}/(\mathbf{TP} + \mathbf{TT} + \mathbf{PP})) * 100 \quad (4)$$

Here TP denotes the number of true defect that were recognised as pseudo defect, PT denotes the pseudo defect recognised as true defect, TT denotes the true defect that were correctly recognised and PP denoted the pseudo defect recognised correctly as pseudo defect.

Table 2 Performance of Classifier at Different Values of σ

Sigma	TP	PT	TT	PP	Performance(%)
0.1	1	19	92	38	99.3
0.2	4	10	89	47	97.2
0.3	7	3	86	54	95.3
0.4	1	9	92	48	99.3
0.5	6	4	87	53	96
0.6	2	6	91	51	98.7
0.7	5	3	86	54	96.6
0.8	1	4	92	53	99.4
0.9	5	2	88	55	96.8
1	4	5	89	50	97.3
1.1	2	4	91	53	98.7
1.2	3	3	90	54	98
1.3	1	5	92	52	99.4
1.4	3	6	90	51	98
1.5	2	5	91	52	98.7

Table 2 explains that the Performance is lesser in case of sigma equals 0.8 and 1.3, however we got almost 94% accuracy that sigma value. Hence the optimal parameter was decided to be around 0.8 with + or -0.001 variance. Finally, the graph of performance Vs σ has shown in Fig.7.

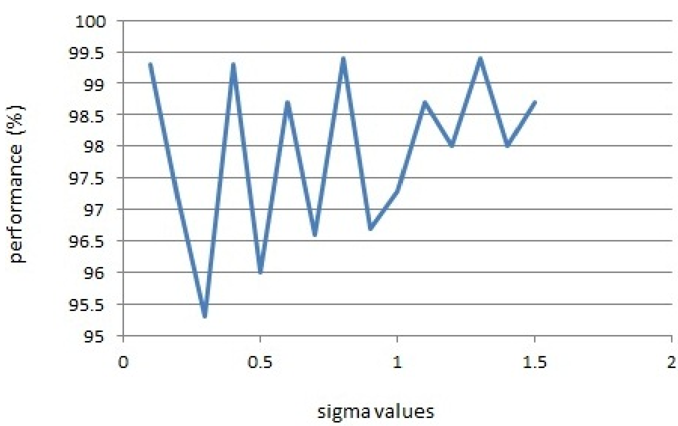


Fig. 7 Performance Graph

The performance explains about the incorrect recognition of true defect into pseudo defect. After evaluating the performance we can see that with optimum parameter we can decrease the incorrect recognition to 0.6% which is very less. Higher the performance, higher the satisfaction of the customer and hence higher the growth of the industry. The proposed approach gives good performance in the evaluation.

5 Conclusions

This paper proposed a new approach to increase the recognition accuracy of true defect and pseudo defect using the SVM parameter selection and proper feature extraction. During feature extraction, we analysed features on statistical basis which helped in improving the accuracy. The experiment showed that the recognition accuracy can be improved by proper choice of features. The texture based features proved more beneficial as compared to statistical features. The accuracy obtained while using only statistical features were around 88% but when the texture features were introduced, the accuracy was improved to 94% when both types of features were used. Hence it seems obvious that texture features are more important in the case of defect classification. In comparison with paper [6], this paper reduced complexity in determining the defect candidate region.It was simply obtained by performing the XOR operation on reference image and defective image.Paper [6] does not highlight the most prominent distinguishing feature.In this paper,a histogram analysis was done to extract out the eminent distinguishing feature which is gray level ratio,the region for comparison were selected through proper observation of histogram.Also,this paper introduced some new features which helped to decide the best combination of features. With slightly different values of sigma, more change in performance were found. The proper choice of σ is very important.

Further approach involves more generalized approach to find the parameters of the SVM.

Acknowledgements. This research was done while Deepak Kumar was visiting Iwahori Lab. for his research internship. Iwahori's research is supported by JSPS Grant in Aid for Scientific Research(C)(23500228) and Chubu University Research Grant.

References

1. Roh, B., et al.: A Neural Network Approach to Defect Classification on Printed Circuit Boards. Journal of the Japan Society of Precision Engineering 67(10), 1621–1626 (2001) (in Japanese)
2. Tanaka, T., Hotta, S., Iga, T., Nakamura, T.: Automatic Image Filter Creation System: To Use for a Defect Classification System. IEICE Technical Report 106(448), 195–198 (2007)
3. Rau, H., Wu, C.-H.: Automatic Optical Inspection for Detecting Defects on Printed Circuit Board Inner Layers. The International Journal of Advanced Manufacturing Technology 25(9-10), 940–946 (2005)
4. Kondo, K., et al.: Defect Classification Using Random Feature Selection and Bagging. The Journal of the Institute of Image Electronics Engineers of Japan 38(1), 9–15 (2009) (in Japanese)
5. Vapnik, V.N.: Statistical Learning Theory. John Wiley & Sons, Inc. (1998)
6. Iwahori, Y., Futamura, K., Adachi, Y.: Discrimination of True Defect and Indefinite Defect with Visual Inspection Using SVM. In: König, A., Dengel, A., Hinkelmann, K., Kise, K., Howlett, R.J., Jain, L.C. (eds.) KES 2011, Part IV. LNCS, vol. 6884, pp. 117–125. Springer, Heidelberg (2011)

Improving the Strategic Level Performance Measurement in Warehousing Processes

Péter Németh, Péter Földesi, and János Botzheim

Abstract. In our paper we propose a new method for the strategic level performance measurement at a warehouse in Hungary. The growing challenge of remaining successful in the warehousing forced the given warehouse to introduce and use a new tool for monitoring and forecasting strategic level indicators. Our proposed tool uses computational intelligence to establish connection between basic operational level data and important strategic level indicators. In this article we present the idea behind this tool and the process of programming and learning with company data. Finally, an evaluation is presented based on the results of the program run.

1 Introduction

The main purpose of the joint research between our University and the warehouse was the development of a computational model and a computer program that is able to predict future strategic level indicators based on forecasted operational level data.

The participant of this project is a central warehouse of a global company acting also as a distribution center for Central Europe and as a finished goods warehouse for the Hungarian factory. An additional task of the warehouse is to serve other warehouses with full truck loads. This makes the warehouse a true logistic service provider despite that it is situated beside a factory. This status comes with additional challenges and problems like as follows:

- different performance factors at outbound flows regarding directions
- different requirements for the different outgoing flows.

These facts will have serious effects on performance measurement that is to be discussed later.

The company management uses Balanced Scorecard (BSC) as strategic planning and monitoring tool. Local warehouse performance targets are derived from strategic level objectives. Since the appearance of the BSC in 1992, it has gained widespread acceptance as an important tool for performance measurement and

Péter Németh · Péter Földesi · János Botzheim
Department of Logistics and Forwarding, Széchenyi István University,
1 Egyetem tér, Győr, 9026, Hungary
e-mail: {nemethp,foldesi,botzheim}@sze.hu

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 365–374.
springerlink.com

strategic management in the for-profit sector. The balanced scorecard model was developed as a means for addressing the strategy development process, continuing in monitoring strategy achievement and performance measurement. It does this by dividing measures into four different, inter-related perspectives:

1. Financial Perspective: How do we look to the shareholders? The strategy for the growth, profitability and risk viewed from the shareholder.
2. Customer Perspective: How do our customers see us? The strategy for creating value and differentiation from the perspective of the customer.
3. Internal Business Processes Perspective: What must we excel at? The strategic priorities for various business processes that create customer and shareholder satisfaction.
4. Innovation and Learning Perspective: Can we continue to improve and create value? The priorities to create a climate that supports organizational change, innovation and growth.

By integrating objectives, measures, targets and initiatives of each of the four perspectives to support the overall vision and strategy, the BSC demonstrates its value as a strategic management instrument that goes beyond mere financial indicators by emphasizing the importance of non-financial perspectives. To ensure financial performance, the other perspectives act as drivers and need to be given balanced weighting [1].

More important is the tactical level performance measurement that is linked to the Balanced Scorecard measurement. This tool is able to show the results for the workers about the previous month. The following areas are covered with that tool:

- Customer complaints
- Damaged products at the warehouse
- Warehouse inventory differences
- Warehouse loading delays
- Trailer utilization
- Equipment damages
- Warehouse productivity.

All the above mentioned measurements are carried out each month to show some results from the warehouse performance within the given year. It is linked with the BSC measurement numbers, and thus it is a suitable tool to monitor the changes in BSC indicators within the given year.

The tool is a method for linking operative processes to the strategic performance measurement numbers of the BSC method. Without a proper operative level performance measurement the feedback to strategic level will be missing thus making evasive actions against problems and bottlenecks impossible.

As for operational level performance measurement the following method is used. At the end of all shifts the shift leader records the subsequent data about the shift workers:

- tasks fulfilled at the given day
- time consumption of these tasks.

At the end of the month – based on the daily recorded data, the actual performance in piece or in m^3 and a pre-defined norm – the operative performance of the warehouse can be calculated.

The major problem with this performance measurement system is that it relies on the pre-defined norms. If the performance is high for a given amount of time the management will raise the norm values. However, if the complexity of the tasks is raising the norm values will not be lowered. This one-direction inelasticity can make the operative feedback about performance inaccurate.

The main idea behind the use of the proposed model was that we needed a link between complex Balance Scorecard (BSC) indicators and easy-to-measure and easy-to-obtain indicators. After visiting the warehouse we found that the operating SAP-based business software is frequently used to obtain operational level data and is capable to obtain other data not used yet. Based on the above information our first task was to obtain a logical link between some of the already available data and a BSC indicator.

For the pilot test we have chosen the warehouse efficiency BSC indicator. In the next step we identified operational level data that is most likely being in logical connection with this BSC indicator. We have chosen the following operational level data and with the help of the warehouse staff we received the required number of them:

- percentage of not full pallets among outgoing goods
- work load (actual/norm)
- warehouse storage capacity utilization (in percentage).

Work load represents the ratio between the actual working hours performed and the working hours based on a norm for the performance (handled goods) on any given week.

With the exact logical connection described and with the use of hypothetical or forecasted operational level data the approximate BSC indicator can be calculated. For this purpose we used artificial neural networks. After the learning and testing phase the program is able to calculate the number of damaged goods used the aforementioned operational level indicators as input [4, 11, 13].

2 Our Proposed Approach

We propose a computational model that is able to predict future BSC indicators based on forecasted operational level data.

2.1 Connection between Operational Level and Strategic Level Performance Measurement

It is without doubt that strategic level performance is based on operational level actions. However the connection between them depends on various factors – among them the various types of complexity [5, 10]. The exact relation between them cannot be calculated with traditional mathematical and statistic tools such as function approximation, or regression analysis. Our proposal is to use computational intelligence techniques to identify the connection and describe it in a way that can be used as future strategic planning.

At the warehouse the strategic level performance measurement is based on various BSC-based results such as damaged goods, customer complaints, inventory differences among others. Our goal was to develop a tool that is able to create a direct connection between those indicators and various operational level indicators such as incoming goods, delivered goods, transshipment, and placement within the warehouse, capacity usage of the warehouse and overtime of the workers.

As described before the connection between the above mentioned indicators is very complex and inter-related. Simple calculations and regression analysis is not able to model the inter-relations and complexity of the various input parameters and so the prediction of strategic level performance indicators based on operational level actions will be inaccurate.

2.2 Artificial Neural Networks

The tool to be chosen should be able to predict the strategic level result based on operational level inputs. The operational level inputs can be "manipulated" in short term (for example not allowing overtime for the workers) but strategic level outcomes must be accepted and altering them in short term is not possible.

Based on the above mentioned complex and inter-related connection between the inputs and outputs our proposal is to use artificial neural networks as a computational model for establishing connection between strategic and operational level indicators.

Artificial neural networks (ANN) are inspired by the structure and functional aspects of biological neural networks [7, 14]. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. This learning phase is important for the use of the model. With enough data from the past the ANN is able to learn the connection between the input and output data. Parallel with the learning phase a test phase is taking place where the network is testing itself with other data from the same problem. As a result after the learning the ANN can be used to predict output data from input data.

Lots of different types of artificial neural networks are proposed in the literature [7, 14]. In our application we use Multi-layer Perceptron (MLP) which is one of the most widely known types of ANNs. The topological structure of the MLP type neural network is illustrated in Figure 1.

The network contains three layers. The input layer contains the inputs x_i. In the hidden layer m hidden neurons are applied. Between the input and hidden layers there are connection weights w_{ij} connecting the i-th input and j-th hidden neuron. There is also a bias input $x_0 = 1$ with weight w_{0j} to the j-th hidden neuron. There is an output layer with output y. Between the hidden and output layers there are weights v_j connecting the j-th hidden neuron with the output.

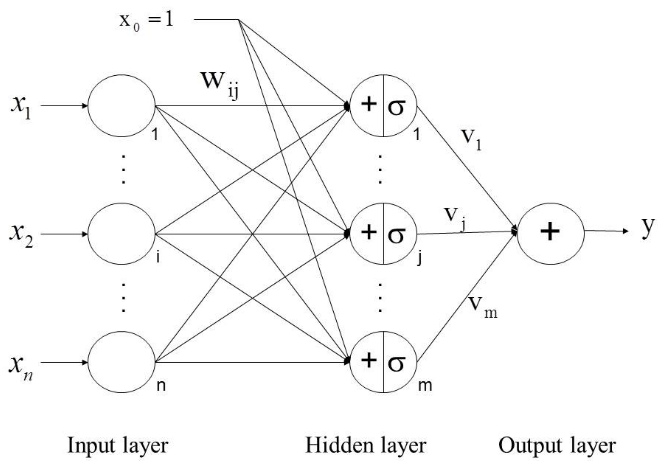


Fig. 1 Topological structure of artificial neural network (MLP)

The output of the neural network can be computed as:

$$y = \sum_{j=1}^{m} v_j \cdot \sigma \left(\sum_{i=0}^{n} x_i \cdot w_{ij} \right). \tag{1}$$

In Equation (1) σ is the sigmoid function:

$$\sigma(s) = \frac{1}{1 + e^{-K \cdot s}}, \tag{2}$$

where K is the slope parameter of the sigmoid function.

2.3 Collection of the Data

In our case the output of the model is one of the BSC indicators hence we cannot affect it in short term. The inputs are operational level data that can be affected in short term. With an ANN that has been taught before the BSC indicator can be predicted by entering possible operational level inputs. And with the manipulation of these inputs the required BSC indicator level can be reached.

For the BSC indicator we have chosen warehouse efficiency. Warehouse efficiency shows the time needed to handle one cubic meter of goods. As input we have chosen:

- percentage of not full pallets among outgoing goods
- work load (actual/norm)
- warehouse storage capacity utilization (in percentage).

For the best training and test results of the model a sufficient number of input data is required. If we give only the incoming and outgoing goods as input to this model the training will perhaps more simple but the future use will be questionable: the more input we have the more we are able to re-model the complex inter-relations that build up goods damage in the warehouse.

Other important criterion for the success of the model is to have enough number of data from the various inputs. In our case we used the weekly data of two years (August 2009 until August 2011) from the above listed input parameters. 51 data were used for each year thus 102 for two years. We have used the odd weeks and even weeks from both years separately for training and testing the model (Table 1).

Table 1 A part of the collected data

Year	Week	Not full pallets percentage	Warehouse capacity percentage	Work load (actual/norm)	Warehouse efficiency (min/m^3)
2009	32	54.09%	84.56%	98.87%	23.6837366
	33	55.77%	85.12%	92.97%	20.6139805
	34	60.25%	84.32%	97.70%	22.9354754
	35	53.02%	82.32%	97.59%	19.0998566
	36	48.72%	79.95%	86.97%	17.8602165
	37	53.72%	77.40%	93.30%	18.8291707
	38	49.90%	73.53%	89.05%	17.8092029
	39	49.18%	71.81%	96.30%	19.7998139
	40	42.27%	66.20%	91.28%	15.5085854
	41	48.99%	66.70%	92.17%	18.6296776
	42	47.85%	63.12%	93.04%	17.6430505
	43	48.47%	63.62%	92.65%	19.0160061
	44	45.37%	51.72%	98.39%	16.9874691
	45	50.75%	51.10%	115.80%	22.2704345
	46	56.09%	52.06%	104.75%	24.4182601
	47	49.89%	51.67%	95.24%	21.1554346
	48	58.36%	51.61%	111.28%	22.9276136
	49	44.21%	53.70%	134.18%	24.5774920
	50	44.83%	58.36%	120.13%	27.5615552
	51	47.11%	59.75%	116.62%	21.7125210

The data were exported in a simpler number structure for the use of the model separately for training and testing. The parameters related to the dataset are:

- number of inputs: The number of operational level performance indicators used for the program (in our case 3)
- number of training patterns: The number of data for each operational level indicators used for training (in our case 51 hence we used weekly data for two years and only the even weeks for training)

- number of test patterns: The number of data for each operational level indicators used for testing (in our case 51 hence we used weekly data for two years and only the odd weeks for testing).

2.4 Training Algorithm

Training or learning is the method of modifying the parameters (e.g. the weights) of the neural network in order to reach a desired goal. We can classify the learning with respect to the learning mechanism, with respect to when this modification takes place, and according to the manner how the adjustment takes place. In this paper the artificial neural network is trained in supervised, off-line manner by bacterial memetic algorithm [3].

Nature inspired evolutionary optimization algorithms are often suitable for global optimization of even non-linear, high-dimensional, multi-modal, and discontinuous problems. Bacterial Evolutionary Algorithm (BEA) [9] is one of these techniques. BEA uses two operators, the bacterial mutation and the gene transfer operation. These operators are based on the microbial evolution phenomenon. The bacterial mutation operation optimizes the chromosome of one bacterium, the gene transfer operation allows the transfer of information between the bacteria in the population.

Evolutionary algorithms are global searchers, however in most cases they give only a quasi-optimal solution to the problem, because their convergence speed is low. Local search approaches can give a more accurate solution, however they are searching for the solution only in the neighborhood of the search space. Local search approaches might be useful in improving the performance of the basic evolutionary algorithm, which may find the global optimum with sufficient precision in this combined way. Combinations of evolutionary and local-search methods are usually referred to as memetic algorithms [8].

A new kind of memetic algorithm based on the bacterial approach is the bacterial memetic algorithm (BMA) proposed in [3]. The algorithm consists of four steps. First, a random initial population with N_{ind} individuals has to be created. Then, bacterial mutation, a local search, and gene transfer are applied, until a stopping criterion (number of generations, N_{gen}) is fulfilled.

BMA can be applied for training neural networks [2, 6]. In this case the parameters to be optimized which are encoded in the bacterium are the w_{ij} and v_j weights of the neural network (see Fig. 1). The details of BMA can be found in [2, 3, 6].

2.5 Experimental Results

In the experiments we performed 10 simulations using 7 hidden neurons and the mean relative error (MRE) of the train and test set was investigated which is defined as:

$$MRE = \frac{1}{p}\sum_{i=1}^{p}\left|\frac{y_i - t_i}{y_i}\right|, \tag{3}$$

where y_i is the output of the network for the i-th pattern, t_i is the desired output for the i-th pattern, and p is the number of patterns.

Table 2 shows the best, the average, and the worst results based on the 10 simulations.

Table 2 Experimental results

	best	average	worst
train	15.56%	15.82%	16.01%
test	17.17%	17.42%	17.61%

Given the fact that our input parameters and the warehouse efficiency are not closely related and the only computation method for this task is human logic, the results are promising.

3 Evaluation of the Results

In our previous paper [12] we presented one possible use of the model. In that case the strategic level indicator was the percentage of damaged goods at the warehouse. This model was further developed with the evaluation feature.

For further testing and to promote the further possibilities of the ANN we prepared five scenarios based on the test (Fig. 2).

The first scenario describes a best-case scenario. The warehouse storage capacity utilization is at a low level; with lowered percentage for the not full pallets and low workload.

The fifth scenario is a worst-case scenario, the warehouse storage capacity utilization at a high level, high work load and large percentage for not full pallets.

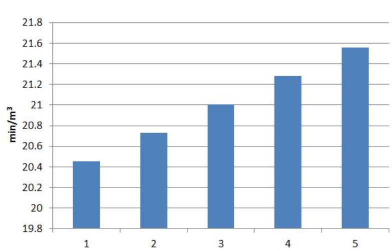


Fig. 2 Warehouse efficiency (min/m^3) as result of the five scenarios

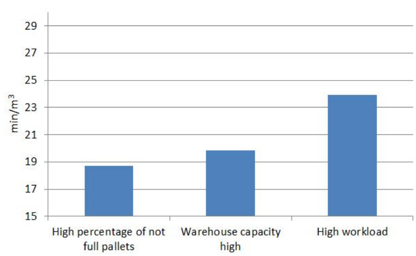


Fig. 3 The effect on warehouse efficiency (in min/m^3) with the "ceteris paribus" raise of various inputs

For further evaluation we created more scenarios leaving all but one input unchanged. The results showed that "ceteris paribus" – leaving all other inputs unchanged – a rise in the workload level affect the warehouse productivity the most (Fig. 3). By this way we can eliminate all inputs that are not relevant enough for the model. After that – besides the aforementioned parameter – we can insert new inputs for a repeated learning of the ANN. That will further improve the model and help to find all operational level data that matters the most for the warehouse efficiency.

4 Conclusions and Future Works

The main area of use should be the prediction of important BSC indicators based on inputs that can be forecasted or guessed. With the altering of those inputs an optimal solution for the fulfillment of the BSC indicator can be chosen.

The result showed that with this first model a usable tool was created that can be further used for other outputs. Other models can be built for other important BSC indicators with completely different input parameters that were chosen by us for the pilot phase.

Our evaluation can help further improve the existing model by eliminating irrelevant inputs for a higher accuracy of the model.

Acknowledgements. The paper is supported by the European Union and co-financed by the European Regional Development Fund TAMOP-4.2.1/B-09/1/KONV-2010-0003: Mobility and Environment: Researches in the fields of motor vehicle industry, energetics and environment in the Middle- and West-Transdanubian Regions of Hungary.

References

[1] Bhagwat, R., Sharma, M.K.: Performance measurement of supply chain management: A balanced scorecard approach. Computers & Industrial Engineering 53(1), 43–62 (2007)

[2] Botzheim, J., Földesi, P.: Fuzzy neural network with novel computation of fuzzy exponent in the sigmoid functions. In: Proceedings of the 8th International Symposium on Management Engineering, Taipei, Taiwan, pp. 285–291 (2011)

[3] Botzheim, J., Cabrita, C., Kóczy, L.T., Ruano, A.E.: Fuzzy rule extraction by bacterial memetic algorithms. In: Proceedings of the 11th World Congress of International Fuzzy Systems Association, Beijing, China, pp. 1563–1568 (2005)

[4] Bozarth, C.C., Warsing, D.P., Flynn, B.B., Flynn, E.J.: The impact of supply chain complexity on manufacturing plant performance. Journal of Operations Management 27(1), 78–93 (2009)

[5] Dubois, A., Hulthén, K., Pedersen, A.C.: Supply chains and interdependence: a theoretical analysis. Journal of Purchasing and Supply Management 10(1), 3–9 (2004)

[6] Gál, L., Botzheim, J., Kóczy, L.T., Ruano, A.E.: Applying bacterial memetic algorithm for training feedforward and fuzzy flip-flop based neural networks. In: Proceedings of the 2009 IFSA World Congress and 2009 EUSFLAT Conference, IFSA-EUSFLAT 2009, Lisbon, Portugal, pp. 1833–1838 (2009)

[7] Hecht-Nielsen, R.: Neurocomputing. Addison-Wesley (1990)
[8] Moscato, P.: On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms. Tech. Rep. Caltech, Pasadena, USA (1989)
[9] Nawa, N.E., Furuhashi, T.: Fuzzy system parameters discovery by bacterial evolutionary algorithm. IEEE Transactions on Fuzzy Systems 7(5), 608–616 (1999)
[10] Németh, P.: Flexibility in supply chains. Acta Technica Jaurinensis Series Logistica 1(2), 371–379 (2008)
[11] Németh, P.: Ellátási láncok hatékony irányítása multi-kritériumos teljesítményméréssel Ph.D. thesis. Széchenyi István University, Győr, Hungary (2009) (in Hungarian)
[12] Németh, P., Földesi, P., Botzheim, J.: Enhancing warehouse performance at a global company. In: Proceedings of Knowledge Globalization Conference 2011, Boston, Massachusetts, USA (2011)
[13] Németh, P., Földesi, P., Csík, Á.: The concept of logistic space in the modelling of supply chain performance. In: Proceedings of the 22nd Annual Production and Operations Management Society Conference, Reno, Nevada, USA (2011)
[14] Zurada, J.M.: Introduction to Artificial Neural Systems. West Publishing Co., St. Paul (1992)

Informatics Medicine as an Intervention That Uses Medicine Informatics Data Systems

John Ronczka

Abstract. This paper aims to provide an introductory emphasis on the use of Informatics for medical purposes, that is designated 'Informatics Medicine'. Traditionally the focus of medical data management ('Medical informatics') has been supported by intelligent decision technologies within the health system, but with the level of drug resistance viruses increasing with overall health management costs, an alternative approach may be plausible. Medical interventions using data ('Informatics Medicine') might be considered to be on the flip side of traditional 'Medicine informatics' but yet, is likely to enable options in the medicine practitioner's intervention Toolkit. These Toolkits could be used via Biorheology interfaces that may use Near-Field Communication or Artificial Intelligent plasmas, foams, gels or glues. As a wayforward, an 'Informatics Medicine' conceptualisation has been put forward within the context of Intelligent Decision Technologies.

1 Introduction

In the journey within the context of Intelligent Decision Technologies (IDT) and medical practice, consideration needs to be given to the relationship between the Patient and Doctor–medical practitioner. Therefore as an introductory process to 'Informatics medicine' what has been utilised is an Occam's Razor (simplicity) approach [1]. The initially step, is to have an understanding of the key words:

- **Medicine:** "the scientific study or practice of diagnosing, treating, and preventing disease or disorders of the body or mind of a person or animal" [2]; and
- **Informatics (medical context):** "the collection, classification, storage, retrieval, and dissemination of recorded knowledge" [3].

A nexus is suggested between practice and knowledge. The next step involves the combination of the two words as 'Medical informatics. This has been undertaken by a review of a sample of academic papers (Table 1) to distil key Themes (conjectures) that are likely to assist in developing the 'Informatics Medicine' conceptualisation and paradigm shift from the data driven Medicine informatics.

John Ronczka
Adjunct to National Centre for Ports and Shipping (NCPS), Maritime Logistics Management, Australian Maritime College, University of Tasmania, Launceston, Tasmania, Australia
e-mail: ronczkaj@ozemail.com.au

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 375–384.
springerlink.com

Table 1 Medicine Informatics Definitions to Themes

Definition	Themes (Conjectures)
Discusses four areas: structures to represent complex data and knowledge relationships, acquiring and presenting data to lessen overload, managing change to information use and integrating information [4].	Structures, relationships; acquiring; managing (C^5M); optimization; integration, (SIAN)
Is used as an overarching term both to describe any informatics efforts related to health care and also to describe a distinct specialty in the discipline of medicine [5].	Efforts (Causality); relationships; distinction; (SIAN); (BRI)
Is the application of computers, communications and information technology and systems to all fields of medicine - medical care, medical education and medical research [6].	Control, Computers, communications (C^5M); information technology; systems; (BRI)
Can be defined as the branch of science concerned with the use of computers and communication technology to acquire, store, analyze, communicate, and display medical information and knowledge to facilitate understanding and improve the accuracy, timeliness, and reliability of decision-making [7].	Computers; communication technology; acquire, store, analyze; communicate (C^5M); display; decision-making (SIAN); (BRI)
It deals with the resources, devices, and methods required to optimize the acquisition, storage, retrieval, and use of information in health and biomedicine [8].	Resources, optimize; acquisition; storage; retrieval, use; (SIAN); (BRI)
Informatics is an emerging discipline that has been defined as the study, invention, and implementation of structures and algorithms to improve communication, understanding and management of medical information [9].	Communication, understanding; management (C^5M), (SIAN)

Note: SIAN (Synergy, integration (Interoperability), assimilation and narratives); C^5M ('Command—communication—control' with *'Management—causalities—consequences'*) (Ronczka, 2009, 2010; 2011); BRI (Biorheological informatics: use of mathematics, statistics, and computer science approaches to study biological interfaces of life sciences (e.g. biology, genome etc.) [10], [11].

The outcomes sought, in the context of IDT, medical practice and the relationship between the Patient and Doctor–medical practitioner are to achieve a:

- Emphasis on the use of informatics for medical purposes;
- Focus that is only on what may be the enablers to developing intelligent decision making tools (e.g. intelligent agents, artificial neural networks, expert systems);

- Suggest the plausibility of 'Informatics Medicine' to medical practitioners; disease management professionals and the health administrative system (e.g. must assist an aging population with increasing impairments and costs); and
- Use of key Themes shown in Fig. 1 may suggest that 'Informatics Medicine' might have an application using the basis of 'Medical Informatics'.

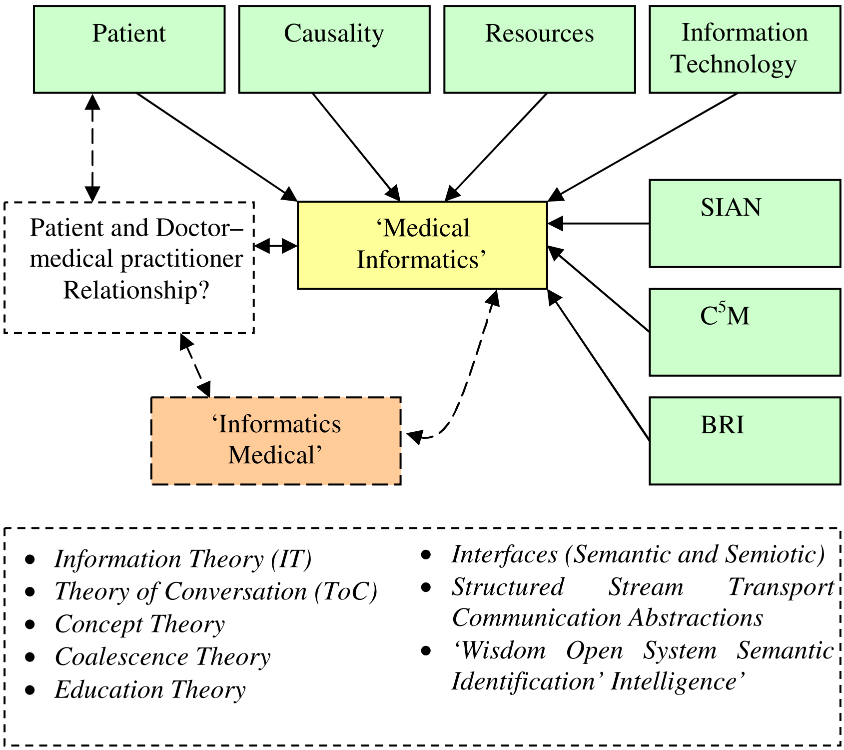


Fig. 1 'Medical Informatics'–'Informatics Medicine' sides of the same coin?

2 Domains

To establish the Domains, a Semantic based mapping using 'Medicine Informatics Definitions to Themes' (Table 1; Fig. 1) assists in their discovery. When Semantic based mapping (Domain [Entity], Model [Representation] and Ontology [constructs]) in the IDT context, there is likely to be entangled single–to–multiple chain 'knowledge—information—learning domains (KILD's) that have a tendency to drive 'Domains' [10]. These entangled–single–multiple chains might exist in multiple operating KILD dimensions. What may be suggested are Semantic relationships (Domain [Medicine and Informatics], Model [Theory's] and Ontology []) that might have a nexus. Such nexus reinforces the use of 'Medical Informatics Domains' (Table 2) to demonstrate a paradigm shift to 'Informatics Medicine'.

Table 2 'Medical Informatics' – 'Informatics Medicine' – Domains

Themes (Conjectures)	Medical Informatics Activity [12]	Semantic Domain (Entity)
Patient	Medical Sociology	Medicine
Resources	Biostatistics; Computational medicine	Informatics
Relationships	Relationship between the Patient and Doctor–medical practitioner	Medicine
Information Technology	Health information; Pathology informatics; Nursing informatics	Informatics
C^5M	Bioinformatics (computational genomics, computational genetics);	Informatics
SIAN	Medical and Health Administration	Informatics
BRI	Medical imaging informatics	Informatics

Note: The developed 'Informatics Medicine' Toolkits could be used via Biorheology interfaces that may use Near-Field Communication (NFC) or Artificial Intelligent (AI) plasmas, foams, gels or glues. It may further be suggested that many disease in the 'Informatics Medicine' context are likely 'Artificial Wisdom Intelligence' (AWI's) as they adapt and reengage.

Medical Themes and domains assist in intergradations and adaptations with conventional model-based medical diagnosis and proposed interventions–countermeasures. Such an approach, within an IDT–medical–health environment may align to a hybrid Semantic and Semiotic theory of control. These controls are likely to have their typologies highlighted via causality patterning. This would not be restrictive to a single language, but would be able to adapt and transcend across Semantic and Semiotic based languages (e.g. dialect, Sign language; Braille alphabet; Quick Response Code [QR]; Morse Code, Hieroglyphs), hardware and firmware associated with various Biorheological interfaces [13]. Therefore, it may be suggested, the prerequisites for developing causality evidence based nexus might align to Domains based on Scientific method (Observation; Question; Hypothesis; Predication Test; decision [Intelligent Decision Technologies {e.g. support; does not support}] [14] and Understanding of the bi-level timescales of physiological mechanisms [15].

3 Causality

In causality evidence based nexus development, what are highlighted are evidence based relationships and processes that exist between the entities and events, in the context of controls such as Biorheological human–machine interface within:

- C^5M [10], [13], [16];
- SIAN [10], [13], [16]; and
- BRI [11], [10], [13], [16].

Within Table 3 is the causality evidence based nexus between 'Medical Informatics'– 'Informatics Medicine' as well as drawing from Fig. 1.

Table 3 'Medical Informatics' – 'Informatics Medicine' – Causality

Themes (Conjectures) (Domains [*Medicine; #Informatics])	Symptoms	Immediate cause	Remote Cause
Patient [*]		YES	
Resources [#]	YES		
Relationships [*]		YES	
Information Technology [#]			YES
C^5M [#]			YES
SIAN [#]			YES
BRI [#]			YES

Causality: Symptoms: effects—factors specific to an event or occurrence. Immediate cause explains why the event or occurrence has occurred. Remote cause may explain why the event or occurrence has occurred [17, [18]. **Relationship** means between the Patient and Doctor–medical practitioner.

Only two Themes drive the 'Immediate cause', that is the Patient and Relations with the set of strong 'Remote cause' that may assist further the development of 'Informatics Medicine'. Conceivably, the fusion of human—machine with interfaces (e.g. hardware, Biorheology) might be the demonstration initiative for 'Informatics Medicine' within the context of Cloud based *'Internet Technologies and Applications'* evolution. A Toolkit could use Biorheology interfaces that may be via NFC or AI plasmas, foams, gels and glues.

4 What Are the 'Informatics Medicine' Examples?

Contained by the context of IDT, early development examples of 'Informatics Medicine' Toolkit from the literature may be suggested to be:

- **Smart phone applications** (Apps):
  - directly monitor the Patient–user: "monitor important health indicators such as glucose and blood pressure" [19];
  - teach your brain to overcome ageing effects on eyesight: improve vision by pick out when a fuzzy pattern is in a white circle" [20]; and
  - EEG scanner: to study people's brain signals in more natural environments such as at home or in the workplace" [21].

- **Computerized Pathologist** to grade breast cancer, predicting which tumors are associated with aggressive treatment [21];
- **The Wireless Cardiac Stimulation** (WiCS): implanted near the heart, to pace the left ventricle synchronously with the right" [22];

- **Biosensor Chip:**
  - Pick's up "heart signals, Electric Potential Integrated Circuit (EPIC's) to assisting the disabled and quadriplegics to control devices [23]; and
  - A computer chip has replaced brain tissue to being an artificial cerebellum to restored lost brain functions [24].
- **Photographic Memory Wearable Cam**: deploying "nana technologies" as electronic memories to compensate for failing neural memories [25].

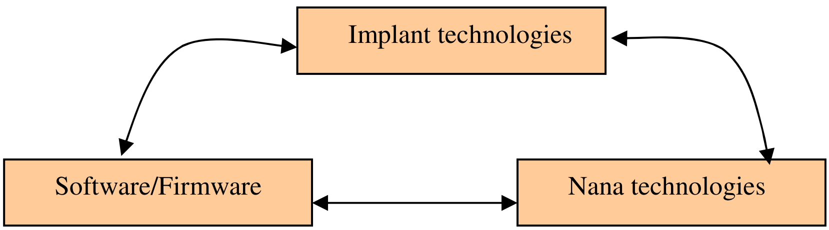


Fig. 2 Informatics medicine contextual paradigm

5 Discussion

Causality?

As 'Informatics Medicine' might be considered to be on the flip side of traditional data driven 'Medicine informatics', an 'Informatics Medicine' conceptualisation has been put forward in a causality framework (Table 4).

Table 4 Early Forms of 'Informatics Medicine' – Causality

Forms of Intervention within Toolkit	Symptoms	Immediate cause	Remote Cause
Smart phone applications	YES		
Computerized Pathologist	YES		
WiCS	YES		
Biosensor Chip			YES
Photographic Memory			YES

The closest to Informatics Medicine in the Toolkit sample (Table 4) appears to be both 'Biosensor chips' and 'Photographic Memory Wearable Cam'. The Chips and Cam links as a Remote cause to the Patient to overcome functional impairments that could be bio-firmware based. An alignment to 'Informatics Medicine' Themes (Conjectures) which in turn form the Domain's appears evident by using Tables 2 and 3 that demonstrate a nexus (Themes to Domains to paradigm).

Flip side of traditional?

In today's medical operating environment, the early forms of 'Informatics Medicine' Toolkits appear to align to an interpretation of 'Informatics' based on Fig. 1. This may suggest the following mechanisms–-relationships:

- Both **sender and receiver** understand single to multi entanglements are likely to enable bundled or unbundled packets that are likely;
- **Nexus** might exist between, with, over, and within events, actions, reactions and interactions over time, space, pace or place at different scales;
- Actions, reactions, and interactions **meshed engagements** that possibly have nexus outcomes;
- **Prima** that might guide tempo and exist between engagements of entity, events, actions, reactions and interactions of packets;
- There could be higher-order **conditional** probabilities like in the occurrence with consequences (risk–likelihood);
- The 'equivocation' may be about **uncertainty** in reception as there maybe "ambiguity" which is the uncertainty in the transmission; and
- In a **two-way** communication system, there perhaps are plausible kinds of conditional entropies at different scales and multi dimensional.

Entanglements of Entities and Events?

'Informatics Medicine' as a Toolkit may involve entangled–single–multiple chains based (Tables 2 and 3) 'Informatics Medicine' Themes, such as:

- **Complexities** in the supporting knowledge—information—learning domains;
- **Multiple** operating dimensions as *Möbius strips*;
- '**Biorheology** causality logic gates' (B–CORG);
- **Biological–*machine*** 'knowledge—information—learning delivery engines' (KILDEE's);
- **Unique** trigger events, information—knowledge that has Biological—*machine* temporal meshing; and
- **Fusion** of human—machine via Biorheology interfaces evolution.

Semantics to Semiotic based Interfaces?

Compatible Interfaces are considered an enabler to achieving 'Informatics Medicine'. A plausible hybrid interface system may align to a Semantics, Semiotic, Information Theory (IT) and Theory of Conversation (ToC) nexus as:

- Domain (Entity) Ξ Human-machine (Biorheology informatics entanglements);
- Model (Representation) Ξ Theory's (IT+ToC); and
- Ontology (Ontological Construct [taxonomy]) Ξ Themes (Conjectures) Ξ styles? (Learning, Thinking, Intelligence).

This then suggests that developed “Informatics Medicine’ Semantic and Semiotic Interfaces needs a hybrid approach of:

- Effective means of communication which has proven itself in practice;
- Pre-established human-defined ontology that is adaptable and reliable;
- Communication can only proceed by the interpretations of behaviours, common behaviours;
- Metaphor such as a new idea is created from the fusion of the two original ideas, or our understanding of the first idea;
- Mental models that are cognitive based for human-computer interaction; and
- User System usage comprises *D1*: concepts the user knows and uses; *D2*: concepts used only occasionally and not initially known; *D3*: the user's model of the system; *D4*: the actual system [16].

6 Conclusions

“Informatics Medicine’ conceptualization and paradigm is considered plausible. The challenge and next step relates to ‘Informatics Medicine’ being used within a real world test environment (Biorheology interfaces). A Patient and Doctor–medical practitioner framework construct is an enabler that could exist within an IDT and Internet technologies and applications. This context may have a preference for soft technologies that minimise miss understanding by using Semantics and Semiotics based human—machine Biorheology interfaces. They may plausibly adjust to the Patient (host’s) as personalised ‘Informatics medicine’. Future work will focus on developing a functioning ‘Informatics Medicine’ interface.

References

[1] Hiroshi, S.: What is Occam’s Razor?, p. 1. Original by Phil Gibbs 1996. University of California (1997), http://www.math.ucr.edu/home/baez/physics/General/occam.html
[2] NICM, “Definition of complementary medicine”, National Institute of Complementary Medicine (NICM) (January 2009), http://www.nicm.edu.au/content/view/31/35/
[3] Merriam-Webster, “Informatics”, Cited a Reference, Merriam-Webster’s Medical Dictionary, Merriam-Webster, Inc., p. 1 (2007), http://dictionary.reference.com/browse/informatics
[4] Lorenzi, N.M.: The Cornerstones of Medical Informatics. American Medical Informatics Association 7(2), 204, 1 (2000), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC61474/
[5] Staggers, N., Thompson, C.B.: The Evolution of Definitions for Nursing Informatics - A Critical Analysis and Revised Definition. American Medical Informatics Association, 1 (2002), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC344585

[6] Collen, M.F.: Medical Informatics, MEDINFO 1980, Tokyo (1980), http://clinton.cs.depaul.edu/upe/MedBioWorkshop2006/slides/Raicu1.pdf

[7] Boobess, Y.A.: Informatics & Nephrology, p. 1. Tawam Hospital, Al Ain, UAE (2001, 2002), http://www.uninet.edu/cin2001/cin2001/conf/boobess.html

[8] FSU, Medical Informatics, p. 1. FSU College of Medicine, Florida State University, College of Medicine (2011), http://med.fsu.edu/index.cfm?page=medicalInformatics.home

[9] IASWWW, Informatics Medicine. Internet Advertising Solutions, Inc. (2011), http://www.iaswww.com/apr/Health/Medicine/Informatics

[10] Ronczka, J.P.: Wisdom Open-system Semantic Identification (WOSSI) Mapping of Causality Logic Gates. In: The First International Workshop on Web & Semantic Technology (WeST 2009) and Third International Conference on Internet Technologies and Applications (ITA 2009), pp. 1–10. Glyndŵr University. North East Wales, UK (2009), http://www.ita09.org

[11] Fu-berlin: Introduction to Dahlem Research School (DRS), p. 1. Freie Universität, Berlin (2010), http://www.fu-berlin.de/sites/en/promovieren/drs/index.html

[12] Ni, J.: Ontology-based Medical Imaging and Radiology Informatics— Domain Mapping and Data Modelling, p. 1. Carver College of Medicine, The University of Iowa, Shanghai Jiao Tong University Medical School, Iowa City, Iowa, USA, 21 Century Forum (December 14, 2009), http://www.uiowa.edu/~hri/

[13] Ronczka, J.P.: Information Theory Hypotheses for Internet Interfaces. In: International Programme Committee for the Fourth International Conference on Internet Technologies and Applications (ITA 2011), pp. 1–10. Glyndŵr University, North East Wales (2011), http://www.ita11.org

[14] Carter, S.: The Scientific Method, p. 1. Clermont College Biology, University of Cincinnati (1996 & 2011), http://www.http//biology.clc.uc.edu/courses/bio104/sci_meth.html

[15] Downing, K.L., Widmant, L.E.: Extending Alodel-Based Diagnosis to Medicine. Department of Computer and Information Science, S-58183, p. 1. Linlcoping University, Linkoping (1991), http://www.qrg.northwestern.edu/papers/Files/qr-workshops/QR91/Downing_1991_Extending_Diagnosis_Medicine.pdf

[16] Ronczka, J.P.: Coalescence Theory application to rheology of entangled, single and multiple chains to suggest biorheology logic gates. In: 5th Pacific Rim Conference on Rheology (PRCR-5), Sapporo Japan, pp. 1–2 (2010), http://prcr2010.com

[17] Hobbs, M.L.: The Value of Urban Design, Ministry for the Environment NZ, p. 1 (2005), http://www.mfe.govt.nz/publications/urban/value-urban-design-summary-jun05/html/page12.html

[18] Bellinger, G., Castro, G., Mills, A.: Data, Information, Knowledge, and Wisdom, p. 1. The International Society for the Systems Sciences (ISSS) (2004), http://www.systems-thinking.org/dikw/dikw.html

[19] Malykhina, E.: 8 Mobile Apps to Help Manage Your Health. Scientific American, 1 (November 2, 2011), http://www.scientificamerican.com/article.cfm?id=iphone-android-mobile-health-apps&WT.mc_id=SA_DD_20111103
[20] Aldhous, P.: App for sore eyes. Technology, New Scientist, 21 (October 1, 2011), http://www.newscientist.com
[21] Aron, L.: Smart app scans brain on the fly. Technology, New Scientist, 24 (September 17, 2011), http://www.newscientist.com
[22] Ledford, H.: Computerized Pathologist Grades Breast Tumors and Predicts Outcomes. The Magazine Nature, 1 (November 9, 2011), http://www.scientificamerican.com/article.cfm?id=computerized-pathologist-grades&WT.mc_id=SA_DD_20111110
[23] Greenemeier, L.: Easy to Beat Next-Gen Cardiac Care Includes Wireless. Scientific America, 1 (November 22, 2011), http://www.scientificamerican.com/article.cfm?id=wireless-cardiac-pacemaker&WT.mc_id=SA_CAT_BS_20111123
[24] Jones, W.D.: New Biosensor Chip Picks Up Heart Signals Remotely. IEEE Spectrum, Institute of Electrical and Electronic Engineers, 1 (November 2011), http://spectrum.ieee.org/biomedical/bionics/new-biosensor-chip-picks-up-heart-signals-remotely/?utm_source=techalert&utm_medium=email&utm_campaign=110311
[25] Geddes, J.: A cyborg is born. Technology, New Scientist, 24 (September 24, 2011), http://www.newscientist.com
[26] Stix, G.: Photographic Memory Wearable Cam Could Help Patients Stave Off Effects of Impaired Recall. Scientific America, 1 (October 31, 2011), http://www.scientificamerican.com/article.cfm?id=photographic-memory-wearable&WT.mc_id=SA_DD_20111101

Intelligent Community Embedded Speech Recognition System Research

Zhanhong Mo, Jiangyan Qi, and Cunmi Song

Abstract. After analyzing the intelligent community on the main functions of the foundation, in this paper we introduce a low power, miniaturization and intelligent embedded speech recognition system to the intelligent community. The Intelligent Community Embedded Speech Recognition system is based on the Samsung S3C2410 ARM microprocessor and operating system Linux. And cooperate with other resources in the Intelligent Community we can achieve community speech in the smart remote control and management.

Keywords: Embedded system, intelligence community, speech recognition, ARM9, Linux.

1 Introduction

The modern information technologies give the building with soul and the construction of "intelligence", realize the harmony between human and nature. Intelligent community has become an international focus of the frontier and the hotspot of this research fields. Intelligent community consists of computer, information, communication, the science and technology management, the intelligent control system of family, community information platform, security system, property management system and comprehensive information service system. Intelligent community provides a safe, environmental, high efficiency, comfortable, convenient living space for the residents by high-tech integrated service platform. [1]

With the development of economy and the improvement of living standards, the requirements of intelligent community are growing. Some developed countries, such as the United States, Japan and Europe, etc., are trying to use modern hi-tech construct intelligent ecological residence and intelligent home environment, to realize "the sustainable development technology" of human settlements. Cisco formally put forward "Internet family" plan in 2000 to adapt to

Zhanhong Mo
Hengshui University, Hengshui City, Hebei Province, China
e-mail: mozhan@126.com

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 385–390.
springerlink.com

the development of the Internet. Intel is to promote the electronic house, which emphasizes the computer as the center, links all sorts of home appliance with various kinds of wired and wireless way, and builds the electronic automation house. Panasonic Company is introducing the "digital home" concept to the elderly housing, which can realize the remote medical care in the family. [2]

But now, most of domestic and overseas intelligent communities of speech recognition system take the PC as a system application platform, and some take the server. Processing and recognizing the speech signal in the PC or server, this kind of speech recognition system is importable, and restricts the serviceable range and, to some extent, impedes the development of intelligent community. [3]

Therefore, based on the Samsung S3C2410 ARM microprocessor and operating system Linux, this paper will introduce an embedded speech recognition system which is low power, miniaturization and intelligent to the intelligent community. With the resources in the intelligence community, such as hardware resources and speech recognition algorithms software, this system realizes voice remote control, intelligent security, language sound management, voice gateway parking management, and provides safe, comfortable, convenient, energy conservation and sustainable development of the living environment.

The intelligent speech recognition system mentioned in this paper provides different simulation equipment of digital experience, with high stability, the lower cost, high performance, low power consumption, high expansibility of the features. The function of the system can meet the needs of the intelligent community, can be convenient for the home life, and enhance residents live in comfort and safety. Through the little change of single application in other terminal equipment, each function module of the system can add to other software system. For example, through the intelligent household online management system, as the 4G network developed, it can be added to 4G wireless network technology based on the intelligent household modules, we will have a better experience.

2 System Design and Implementation

The system uses a Samsung's S3C2410 ARM microprocessor with A/D, D/A circuits, intelligent community related equipment and control circuits, interface circuits and storage circuits, speakers, microphones and other hardware resources and embedded Linux operating system, the necessary drivers as well as the DTW algorithm, such as speech recognition software resources to achieve speech recognition function, so that we can achieve community speech in the smart remote control, intelligent security guard, speech control of import and export management, speech parking management.

Part I：Intelligent community speech recognition system working principle

Speech recognition system is divided into speech training and speech recognition process. Speech training at the course, from the microphone input voice signal by the digital signal chip UDA1341TS audio codes for voice signal acquisition and encoding, to maintain a certain accuracy of the sampling and frequency of sampling to ensure the quality of the sound reduction, the requirements of each household within the community voice must be collected and trained, after training, each model of community residents voice templates are stored in the system of storage devices, when a user reenter, the system will automatically collect the speech signal with each voice template matching and give recognition results, S3C2410 identification in accordance with the results of sending a signal corresponding to implementation of Intelligent speech community entrance management. [5]

The system uses embedded Linux operating system for intelligent community visual intercom, GPRS wireless and other features to provide an open platform. The system by the DTW algorithm template training, and the results of training stored in the embedded Linux operating system appropriate directory in order to identify the stages of the read template. Speech recognition stage is the input voice signal by the audio digital signal chip UDA1341TS codes processed notice embedded Linux operating system, and the existence of the reference template samples for comparison, to identify the most similarity of the effect of the voice as a means of identification, and then by embedded Linux operating system call control S3C2410 function of I/O port, and thus realize the import and export management of intelligent speech community. Linux device drivers for specific hardware are provided to the user program for a group of standardized interfaces, which hide the details of the equipment [4]. User program through the standardization of system calls, these calls and specific hardware are unrelated, and then by the Linux kernel to call specific device driver to operate and control a specific hardware device the actual. [6] Achieve at least the driver should be open, close, read and write system calls. System schematic diagram is shown in Figure 1.

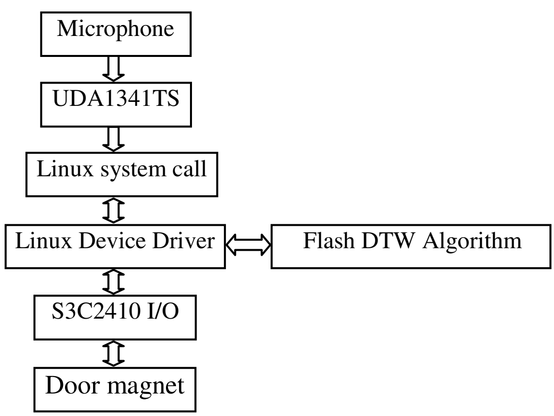


Fig. 1 System schematic diagram

Part II：System hardware design and implementation

System from SAMSUNG (Samsung)'s S3C2410 ARM chip as the core, with the A/D, D/A circuits, control circuits, interface circuits and storage circuits achieve the speech recognition training and recognition process. The external power supplies 5V voltage. SAMSUNG Company S3C2410 is a chip based on the ARM920T core. S3C2410 chip ROM, RAM, A/D, and various peripherals, such as general I/O port, timers, and serial port and so on. Because the system handle analog voice signal, it must first change these analog signals to digital signals, which must be used for A/D converter.

The system adopts PHILIPS company chip UDA1341TS audio codes digital signal. UDA1341TS can convert stereo analog signal to digital signals, digital signals can also be converted analog signals by UDA1341TS, and can be used PGA (programmable gain control), AGC for analog signal processing; for digital signals, the chip also provides a DSP (digital signal processing) feature. Input and output in part by the microphone, speakers, LCD (liquid crystal display), door magnet of intelligent community control equipment (import and export management of speech as an example) component. Analog voice signals through the microphone input and then proceed to pre-treatment, including the A/D conversion and AGC. A/D sampling frequency of 8 kHz is the speech signal sampling frequency. Storage part chosen 4M × 16bit Flash AM29LV160DB chips, used programmer procedures, two chips of 32M × 16bit SDRAM high-capacity model HY57V561620, 4Kbit EEPROM AT24C04. And the intelligent community door management control equipment (import and export management of speech as an example), and 20-pin JTAG port and so on. System hardware block diagram shown in Figure 2.

Part III：system software design and implementation

The system uses dynamic time warping DTW algorithm. DTW algorithm compares time series of speech feature vector with the reference template library

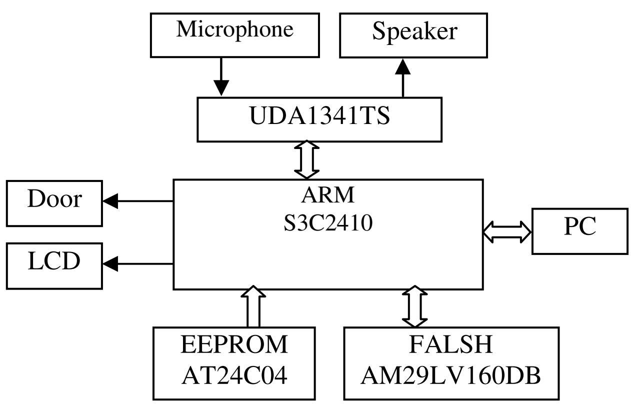


Fig. 2 Diagram of the system hardware

for each template, the highest similarity as a means of identification results. Because at S3C2410ARM board running embedded Linux operating system, written procedures should be based on the procedure under Linux. Recording the whole process is divided into program modules, DTW speech recognition program modules, ARM board hardware control program modules, and Linux makes file management documents. This article only give DTW algorithm in the embedded Linux platform C language implementation of the core code is as follows:

```
for(f=0;f<2;f++)
{wave_read(t,filename_test[f],f); // Read the test template parameters
for(d=1;d<=M;d++)
{ wave_read(r,filename_ref[d],d); // Read the reference template parameters
for (i=0;i<n;i++) {
for (j=0;j<n;j++) {
float x,y; sum1=0.0;
for(g=0;g<M;g++) {
x=t[i][g];
y=r[j][g];
sum1+=(x-y)*(x-y); }
D1[i][j]=sum1;
//Matching distance matrixD1[n][n] } }
D2[0][0]=D1[0][0];
for(i=1;i<n;i++)
for(j=0;j<n;j++){
D3=D2[i-1][j];
if(j>0) D4=D2[i-1][j-1];
else D4=REALMAX;
if(j>1) D5=D2[i-1][j-2];
elseD5=REALMAX; D2[i][j]=D1[i][j]+zmin(D3,D4,D5);
/* zmin() for D3, D4, D5 minimum*/}
strdata[d].data=D2[i-1][j-1];
aa=strdata[1].data; g=0;
for(k=1;k<=10;k++){
if(aa>strdata[k].data)
{ aa=strdata[k].data; g=k; } }
strdata[g].data=aa;
printf( "\nThe %s and %s is min and its value " "dist[%d] is %e\n\n",
filename_test[f], strdata[g-1].chstr[g-1], g, strdata[g].data);
/* f are the number of test templates, i is the distance at least corresponding to
the document, dist [i] are the corresponding minimum distance*/ } }
```

3 Concluding Remarks

In this paper, intelligent community speech recognition system design and implementation, based on ARM + Linux platform, are described. Followed by the development of intelligent community, the working principle, the system design and implementation of hardware, system software design and implementation are described. The system uses advanced micro-processor as a master ARM chip embedded Linux platform realize intelligent community speech recognition system. The system provides a natural, convenient and cost-effective speech control function, saving many human and material resources for community residents.

References

[1] Zhang, Y.: Building intelligent systems. Tsinghua University Press, Beijing (2010)

[2] Lingling, O.: Design and Implementation of Terminal System for Intelligent Building Based on Embedded Linux. Dissertation Xian University of Electronic science and Technology (2010)

[3] Wei, G.: The Study of Embedded Speech Recognition System Based on ARM. Dissertation Yanshan University (2010)

[4] Hu, H.: Voice signal digital processing. Harbin Institute of Technology Press, Harbin (2005)

[5] Zhang, S.: Embedded system design and application. Tsinghua University Press, Beijing (2008)

[6] Wang, S., Tong, Y.: Living community of intelligent system construction and technical guidelines for elements. China Building Industry Press, Beijing (2003)

Interactive Grammars: Toward Perception Based Computing

Andrzej Skowron and Piotr Wasilewski

Abstract. We present a general scheme of interaction and we discuss the role of interactions in modeling of perception processes. We use information systems as a starting point for perception modeling, i.e., modeling of the process of understanding of sensory measurements. The novelty of the paper is an attempt to present the perception process by means of interactive grammars.

Keywords: interactive computing, interactive information systems, interactive tables, rough sets, granular computing, wisdom technology.

1 Introduction

Perception Based Computing (PBC) methods are needed for solving problems of data mining (DM) and knowledge discovery in databases (KDD) with dynamically evolving complex data (e.g., stream data sources, sensory data). Another challenge, making PBC methods indispensable, is a growth of the size and complexity of data sources (e.g., Web sources, neuro-imaging data, data from network interactions). These challenges, in particular, discovery of complex concepts such as behavioral patterns, hardly can be met by classical methods [15]. They can be met by KDD systems which dialogue with experts or users during the discovery process [24] or

Andrzej Skowron
Institute of Mathematics
University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
e-mail: skowron@mimuw.edu.pl

Piotr Wasilewski
Institute of Informatics
University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
e-mail: piotr@mimuw.edu.pl

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 391–402.
springerlink.com

by adaptive learning systems changing themselves during the learning process as the response to evolving data.

Another area where PBC methods are needed is the multi-agent systems field. Behavior steering and coordination of multi-agent coalitions acting and cooperating in open, unpredictable environments call for interactive algorithms [4], i.e. algorithms interacting with the environment during performing particular steps of computations or changing themselves during the process of computation. Next challenge of this type comes from human -Erobot interaction. The problem of human control over autonomously coordinating swarms of robots is the central challenge in this field which should be solved before human E robot teams can be taken out of laboratories and put to practical use.

Coordination and control are essentially perception based thus PBC methods are indispensable for designing and behavior description of cognitive systems and for understanding interactions in layered granular networks [14] where granules can be interpreted both as data patterns and agents (e.g., robots or mobile sensors). Granules in such networks which are additionally self-organizing can be also understood as cores in pertinent multi-core computing engines in structurally and run-time reconfigurable hardware, what makes PBCs useful in computer engineering as well as an essential part of cognitive informatics.

The approach to perception considered in this paper is wider than studied in the visual perception domain, however it follows Marrs's ideas [6]. Perception here is treated as action-oriented perception [1, 9] and driven by actions. Thus, our approach is consistent with the intuition presented in [9] (p. 1):

> *... perceiving is a way of acting. Perception is not something that happens to us, or in us. It is something we do. Think of blind person tap-tapping his or her way around a cluttered space, perceiving the space by touch, not all at once, but through time, by skillful probing and movement. This is, or at least ought to be, our paradigm of what perceiving is. The world makes itself available to the perceiver through physical movement and interaction.*

We proposed to build foundations for Perception based Computing (PBC) on the basis of Interactive Granular Computing (IRGC), in particular on Interactive Rough Granular Computing. A step toward this goal is presented in [19, 20]. PBC can be considered in a more general framework of Wisdom Technology (Wistech) [5] based on a metaequation

$$wisdom = knowledge + adaptive\ judgment + interactions. \tag{1}$$

In general, adaptive judgment is a mixture of deductive and inductive reasoning methods for reasoning about interactive granular computations and on controlling such computations by adaptive strategies for. The mentioned mixture of deductive and inductive reasoning creates many challenges [22].

In this paper, we consider, perception as the process of understanding of sensory information. Perceiving units will be called agents. Agents are preforming computations on objects called (information) granules (see, e.g., [25, 26, 14, 5, 20]).

We discuss the role of information systems in modeling of perception processes. Using a general scheme of interactions, we distinguish basic information on interactions in time between agent and its environment which should be represented for proper interaction modeling. Next, we define a special class of decision tables in which such information can be represented. These decision tables create a starting point for modeling of perception processes. We outline an approach for modeling this process by means of interactive grammars.

The paper has the following organization.Section 2 presents interactive information systems together with some concepts from hierarchical modeling in rough set theory. Section 3 is devoted to elements of interactive computing and rough set analysis of interactive computing. Section 4 presents and introduction to modeling of perception process by means of interactive grammars.

2 Interactive Information Systems

Rough set theory, created by Zdzisaw Pawlak [10, 11, 12], is an approach to granular computing based on ability to discern between objects. Objects and granules are represented by means of information systems called also information tables. *An information system* is a triple $\mathscr{A} = (U, At, \{V_a\}_{a\in At})$, where U is a set of *objects*, At is a set of *attributes*, and each V_a is a *value domain* of an attribute $a \in At$, where $a : U \longrightarrow \mathscr{P}(V_a)$ ($\mathscr{P}(V_a)$ is a the power set of V_a). If $a(x) \neq \emptyset$ for all $x \in U$ and $a \in At$, then $\mathscr{A}$ is *total*. If $card(a(x)) = 1$ for every $x \in U$ and $a \in At$, then $\mathscr{A}$ is *deterministic*, otherwise $\mathscr{A}$ is *nondeterministic*. Let $(U, At, \{V_a\}_{a\in At})$ be an information system, partial information about a given object $x \in U$ is represented by the A-signature of x, $InfA(x) = \{(a, a(x)) : a \in A\}$ where $A \subseteq At$.

Another assumption about original information systems was previously accepted, namely that, as mappings, attributes are surjections, i.e. every attribute value is possessed by at least one object, what reflects a closed-world data base point of view. This assumption was implicitly broken in information systems with real-valued attributes and explicitly by introducing *sensory* and *perception* attributes [17, 19]. Attributes which are injections can be viewed as open for interactions in the sense that it is admitted that a new value, not assigned previously to objects, can appear. Thus attributes which are surjections or injections are also called *closed* or *open attributes*, respectively [19].

In the case of perception attributes another partition of attributes is essential. One can differentiate between *atomic* and *constructible* attributes. *Atomic attributes* are basic in the sense that their values depend only on some external factors, with respect to a given information system and are independent from the values of other attributes of this system. Atomic attributes can be closed as well as open. *Constructible attributes* are complex attributes which are inductively defined from atomic attributes of a given information system: if b is a constructible attribute, then for any some object x and already defined atomic attributes $a_1, a_2, \ldots, a_k$:

$$b(x) = F(a_1(x),\ a_2(x), \ldots,\ a_k(x)), \tag{2}$$

where $F : V_{a_1} \times V_{a_2} \times \ldots \times V_{a_K} \longrightarrow V_b$ and values form V_b are constructed on the basis of values from V_i for $i = 1, \ldots, k$.

Sensory attributes represent sensor measurements. They are open atomic attributes which values are results of measurements conducted by sensors thus they depend only on the environment and are independent from values of other attributes. *Perception attributes* are sensory attributes or constructible attributes defined on the basis of sensory ones. The latter are also called *complex perception attributes*. Complex perception attributes represent higher order result of perception, e.g. some identified patterns or created perceptual granules. For formal descriptions of sensory and perception attributes the reader is referred to [19].

We also distinguish *action attributes: atomic or complex* representing basic actions or plans, respectively. Any action attribute value corresponds to an action and is linked with some sensory attributes used for predicting and recording the sensory measurements before or after the action execution. For more details on sensory and action attributes in information systems the reader is referred to [19].

Constructible attributes can take as their values relational structures defined over its value domains. Hence, we consider a generalization of traditionally used information systems [10, 11, 12] by considering together with value set V_a a relational structure over V_a. Note that also objects in such information systems may have complex structure. For detailed explanation one can consult [18, 19].

3 Interactive Computing

Interactive algorithms are adaptive in the sense that they can change themselves during computation performing, the next state of the algorithm can be only foreseen with some probability, since every step can be changed as a response to the environment influence during performing of that step [4].

In the process of interactive computation both an agent and an environment are involved. A system performing interactive computing consists of an agent and the agent's environment, more exactly, a part of the environment that is perceived by an agent. The global states of such system are defined as pairs $(s_{ag}(t), s_e(t))$, where $s_{ag}(t)$ and $s_e(t)$ are states of a given agent ag and the environment e at time t, respectively. Figure 1 illustrates how, in the case of interactive computations, the transition relation $\longrightarrow$ between global states is performed, i.e., when $(s_{ag}(t), s_e(t)) \longrightarrow (s_{ag}(t+\Delta), s_e(t+\Delta))$ holds, where Δ is a time necessary for performing the transition. $A(t)$, $E(t)$ denote the set of attributes available by agent ag at the moment of time t and the set of attributes (sensors) influenced by environment e at time t, respectively. $Inf_{A(t)}(s_{ag}(t), s_e(t))$ is the signature [19] of $(s_{ag}(t), s_e(t))$ relative to the set of attributes $A(t)$ and $Inf_{E(t)}(s_{ag}(t), s_e(t))$ is the signature of $(s_{ag}(t), s_e(t))$ relative to the set of attributes $E(t)$, i.e. signature $Inf_{A(t)}(s_{ag}(t), s_e(t))$ describe a state of agent ag at the moment of time t, while signature $Inf_{E(t)}(s_{ag}(t), s_e(t))$ describe a part of environment e that is perceived by agent ag at time t. These signatures are arguments of strategies Sel_Int_{ag}, Sel_Int_e selecting interactions I_{ag} and I_e of agent ag with environment e and environment e with agent ag, respectively. I_{ag} represents the planned influence of agent ag on environment e (and on the agent ag itself), i.e. an action of agent ag while I_e represents an influence of environment e on agent ag (and on the environment) which results will be perceived by ag. This includes also predicted results of agent action on environment e (and on the agent ag itself) as well as a perception of change of e caused by the previous global state. $I_{ag} \otimes I_e$ denotes the result of the interaction product $\otimes$ on I_{ag} and I_e. Since set $E(t)$ can be insufficient for describing environment e therefore agent ag can have very incomplete information about I_e as well as the result $(I_{ag} \otimes I_e)(s_{ag}(t+\delta), s_e(t+\delta))$ only, where δ denotes the delay necessary for computing the signatures and selection of interactions (for reasoning simplicity we assume that these delays for ag and e are the same). Thus, information about $s_{ag}(t+\Delta)$ and $s_e(t+\Delta)$ perceived by ag can be very incomplete too. Usually, agent ag can only estimate of $s_{ag}(t+\Delta)$ and $s_e(t+\Delta)$ during planning selection of interaction I_{ag}. These predictions then can be compared with the perception of global state $(s_{ag}(t+\Delta), s_e(t+\Delta))$ by means of attributes $A(t+\Delta)$. Interaction $I_{ag} \otimes I_e$ can change the content of both the agent state and the environment state. The current set of attributes $A(t)$ is a part of the agent state $s_{ag}(t)$ and can be changed, for example, by adding new attributes discovered using I_{ag}. As we mention at the beginning of this section, interactiveness of agent ag is formally reflected by the fact that the description of strategy Sel_Int_{ag} is stored in the current state of agent $s_{ag}(t)$. This strategy itself can be modified as the result of interaction, and generally, sets of attributes as well as strategies for selecting interactions can be adopted in time.

Interactive computing is not performed solely by agents in the sense that also environments are essentially involved in the computation process. Therefore, more exactly is to say that a given agent ag observes computation than that ag perform it, and that an interactive computation is performed commonly by the agent and its environment, namely, this part of the environment which is perceived by the agent and

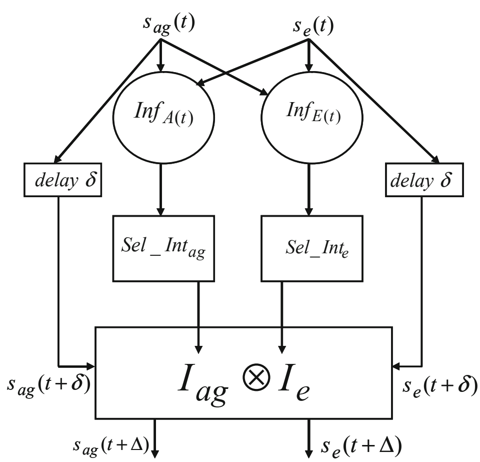


Fig. 1 Transition from global state $(s_{ag}(t), s_e(t))$ to global state $(s_{ag}(t+\Delta), s_e(t+\Delta))$

affects the agent. However, agent *ag* affected by its environment does not passively respond to the environment, but it applies strategy Sel_Int_{ag} selecting interactions with its environment. This has another consequence, namely that agent *ag* does not observes the whole environment, only part of it determined by the activated sensors of *ag*. Hence, in particular, values of not all attributes in a given row of information systems describing *ag* are known. An agent posses only partial information about the environment. It should be noted that another interaction also can take place, namely internal interaction inside an agent between its components: an agent can observe this interaction using its sensors, in this case internal sensors, like in animals signals of a somatosensory system. And, as in the case of the environment, an agent usually posses only partial information about its internal states.

Interactive computations in our approach are sequences of signatures of global states connected by transition relation. More formally, *a computation observed by an agent ag in interaction with its environment e* is any sequence $sig_1, \ldots sig_n, \ldots$ fulfilling the following conditions: for some t, Δ and for any i, sig_i is the signature of global state $(s_{ag}(t+i), s_e(t+i))$ relative to the attribute set $A(t+i)$ available by *ag* at a moment of time $t+i\Delta$ and

$$(s_{ag}(t+i\Delta), s_e(t+i\Delta)) \longrightarrow (s_{ag}(t+(i+1)\Delta), s_e(t+(i+1)\Delta)). \qquad (3)$$

Length of a computation Comp observed by *ag* is a number of elements of sequence minus one, when *Comp* is a finite or is a cardinality of *Comp*, when it is infinite sequence.

4 Towards Interactive Grammars

Research on perception based computing is made in analogy to investigation of perception in cognitive science. It is a commonly accepted view in the contemporary cognitive science that human perception is not passive but constructive: in the process of interpretation of sensory data, some non-sensual elements are involved possibly including knowledge. Therefore, human perception as information processing has a hierarchical nature: it consists of few levels starting from sensory one, where higher levels becoming less similar to sensual one being results of applications of non-sensual elements preexisting in a cognitive system (preexisting with respect to sensations). By this hierarchical nature, perception leads from sensations (which are always particular and singular) to creation or interpretation of knowledge (which always is less or more general) and makes applications of this knowledge in real actions successful. Thus developing PBC we need to develop means for representing sensory and perception data, ways of transforming them and put them to work in hierarchical setting. This gives a motivation for introducing *interactive grammars*.

In Figure 2, it is illustrated the main idea of perception function in interactive grammars. On the lower plane (level I in Figure 2), the black rectangles represent perceived by sensors states of the environment. Theses perception words are from the training sample. The dotted arrows link perceptions of states with their successors after performing an action named as *ac*. The induced perception function c_f

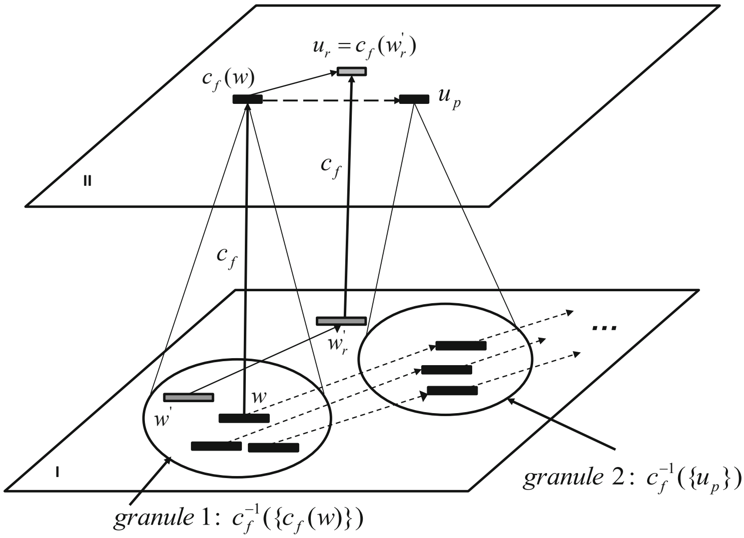


Fig. 2 Perception function in interactive grammars

should granulate the perceived states, i.e., the states from one granule should be transformed into states from another granule. These granules are defined as counter images of the perception function c_f on perception words representing the vector of sensory measurements. In Figure 2, the granule 1 corresponding to the sensory word w is pictured by ellipse denoting the set $c_f^{-1}\{c_f(w)\}$. If the perception function was learned correctly than, on the higher level (level II in Figure 2), it should be possible to predict the description of granule 2 obtained after performing the action *ac* (pointed by dotted arrow in Figure 2) on perception words from granule 1. This means that all (or almost all) perception words from the first granule belonging to the training sample will be transformed by *ac* into the granule 2 equal to $c_f^{-1}(\{u_p\})$ defined by the c_f-counter image of the perception word u_p (on the level II) pointed by the solid arrow. However, we require, as usual in learning, that this property will be valid also for perception words from the first granule which are unknown in the training sample (are not belonging to the training sample). We denote such a perception word w' in Figure 2 by a gray rectangle. However, it may happen, that the perception word w'_r (denoted by a gray rectangle on level I) corresponding to the environment state after performing *ac* will be outside of the granule 2 (defined by $c_f^{-1}\{u_p\}$) and the defined by c_f perception word $u_r = c_f(w')$ on the second level II (marked by gray rectangle and pointed by the solid arrow) will be different than the predicted one u_p (pointed by the dotted arrow). The difference between predicted perception word u_r (on level II) and real perception word u_r (on level II) should be taken into consideration in planning of the next action or in the reconfiguration of the current plan. The perception function may be treated as a compression function or generalization function.

These ideas are encoded in the interactive grammars. Learning of perception functions is a challenge. This requires hierarchical learning and domain knowledge (see, e.g., [15, 3]).

Interactive grammars share some aspects with classical formal grammars and differ with respect to others. As classical formal grammars, interactive grammars consist of two finite and disjoint sets: N of nonterminal symbols and Σ of terminal symbols. however, in the case of interactive grammars we have two types of *terminal symbols* (in short *terminals*): *perception terminals* and *action terminals*, as well as we have two types of *nonterminal symbols* (in short *nonterminals*): *perception nonterminals* and *action nonterminals*. Perception terminals correspond to results of sensory measurements made by particular sensors and can be called also *sensory terminals* while sensory nonterminals correspond to higher order results of perception, e.g. some identified sensory patterns or created perceptual granules. Action terminals correspond to atomic actions while action nonterminals to compound actions inductively constructed from atomic ones. Therefore, the set Σ of terminal symbols of an interactive grammar consists of two disjoint sets Σ_p, Σ_{ac} of, respectively, perception terminals and action terminals: $\Sigma = \Sigma_p \cup \Sigma_{ac}$. Similarly, the set N of nonterminal symbols consist of two disjoint sets N_p, N_{ac} of, respectively, perception terminals and action terminals: $N = N_p \cup N_{ac}$. Instead of Σ_p we can use also symbol Σ_s in order to highlight the fact that perception terminals correspond only to values returned by sensors.

Instead of one distinguished start symbol, like in classical formal grammars, in interactive grammars we have a set of start words, SW, these are *sensory words*, they are vectors over a set of sensory terminals i.e.: $SW := \Sigma_s^n = \Sigma_s \times \ldots \times \Sigma_s$. Sensory words correspond to results of measurements made by coalitions of sensors. We assume also that there is a special sensory terminal symbol $\perp \in \Sigma_s$. The symbol $\perp$ can be interpreted as 'unknown value' and when it appears on some position in a vector $v \in \Sigma_s^n$ it correspond to the fact that a particular sensor was deactivated during measurement process.

Let us note that set of sensory terminals is infinite, possibly non-denumerable. Contrary to this, sets $N_p \cup N_a$ and Σ_a are finite. However, formal grammars consist only of finite many rules. Therefore we need a more general procedure than rules, for translating sensory words over infinite alphabet to words over the finite alphabet of perception nonterminals: $P := N_p^*$, where * is the Kleene star operator. Elements of the set $\Sigma_s^n \cup N_p^*$ are called *perception words* while elements of the set $(\Sigma_a \cup N_a)^*$ - *action words*,

This translation is done by a *compression function* which maps sensory words to perception words of higher orders, usually shorter than sensory words but possibly containing sensory terminal symbols. Thus a compression function has the following form:

$$c_f : \Sigma_s^n \longrightarrow (N_p \cup \Sigma_s)^*.$$

It is adaptatively learned from examples of perception-action processes. At every stage of learning process, compression function c_f consists of finitely many examples of the form:

$$w \longmapsto c_f(w).$$

This examples are in fact *compression rules* of the form (in order to differentiate formal grammars and functional notations we use short arrows for grammar rules[1]):

$$w \to w', \text{where } w \in \Sigma_s^n \text{ and } w' \in N_p^*.$$

Thus using rule notation we describe the form of compression rules in the following way:

$$\Sigma_s^n \to (N_p \cup \Sigma_s)^*.$$

Perception nonterminal symbols correspond to higher order stages of perception processes as sensory patterns, perceptual granules or complex vague concepts. Thus compression rules reflect an idea that perception processes can be represented by compression functions, possibly lossy compression. Therefore, contrary to classical formal grammars, interactive grammars in addition to *production rules* have also *compression rules* given by compression function.

Production rules are of the form:

$$(\Sigma_s^n \cup \Sigma_a \cup N_p \cup N_a)^+ \to (N_a \cup \Sigma_a)(\Sigma_s^n \cup \Sigma_a \cup N_p \cup N_a)^*,$$

i.e. they map words of any kind to words starting from action words. Production rules are rules of two distinguished kinds:

- *decision rules* are of the form:

$$(\Sigma_s^n \cup \Sigma_a \cup N_p \cup N_a)^+ \to (N_a \cup \Sigma_a)^*,$$

while

- *prediction rules* are of the form:

$$p_0(\Sigma_a \cup N_a)^+(\Sigma_s^n \cup N_p)^* \to p_0(\Sigma_a \cup N_a)^+ p_0 \Sigma_s^n,$$

where $p_0 \in \Sigma_a$ is distinguished symbol called *prediction symbol*

Decision rules map perception words to action ones and they correspond to planning processes or to activation processes. More exactly, *planning decision rules*, i.e. decision rules corresponding to planning processes, where on the basis of perceptual conditions plans are selected, have the form:

$$(\Sigma_s^n \cup N_p)^+ \to (\Sigma_s^n \cup N_p)^+ p_0(N_a \cup \Sigma_a)^*,$$

while *activation decision rules*, i.e. decision rules corresponding to action activation processes, where perceptual guards, usually complex vague concepts activates some actions, have the form:

[1] For describing forms of grammar rules we use also a generalized notation: $X \to Y$ refers to a form of rules whose conditions belongs to set X and productions to set Y, e.g. $w \to w'$, where $w \in X$ and $w' \in Y$. In addition, wX refers to a form of words consisting of word w at the beginning and at the end some word form set X, e.g. wx where $x \in X$, and XY refers to a form of words such that the starting part belongs to X and the ending part belongs to Y, e.g. xy, where $x \in X$ and $y \in Y$.

$$p_0(\Sigma_a \cup N_a)^+(\Sigma_s^n \cup N_p)^* \to p_0(\Sigma_a \cup N_a)^+(\Sigma_s^n \cup N_p)^*(N_a \cup \Sigma_a)^*\mathbf{?},$$

where $\mathbf{?} \in \Sigma_a$ and denotes query to the environment, i.e. activating an action of measurement.

Words produced by interactive grammars represents histories of particular computations. For example word:

$$x\perp yz\alpha p_0Abp_0sw_2Ab\mathbf{?}Absw_3. \tag{4}$$

represents the history of the following computation: a sensory measurement gave result $x\perp yz$ which was processed to a higher order perception granule α by compression function c_f, i.e. $x\perp yz \longmapsto c_f(x\perp yz) = x\perp yz\alpha$, on the basis of α a plan Ab was selected with a predicted sensory result sw_2 what led to executione of plan Ab, i.e. activation of actions A and b what led to the observed sensory result sw_3. The represented computation reached the phase when a decision should be made whether the goal was reached and the system should terminate of doing actions Ab (since goal is reached), continue of doing actions Ab (since goal is not reached yet but execution of the plan is likely leading to it) or change the plan, abandon doing actions Ab and start to do other action. In order to make such decisions a new function of selecting strategies should be learned from data, i.e. from examples of interaction histories represented by words derived by interactive grammars. Developing of interactive grammars is a key step to learning selecting strategies functions.

5 Conclusions

We restricted our considerations to rather intuitive presentation of the main ideas of interactive grammars as a tool for modeling perception. In particular, we presented several illustrative examples of granules interactions (e.g., information systems interactions) and we explained the basic intuitions behind interactive grammars. Deeper analysis of perception based computing will require considering construction of hierarchical information systems. For example, the (semi)optimal selected action at a given moment of time may be predicted on the basis of high level features of histories recorded in discussed information systems. This directly refers to the main idea of perception discussed in [8]. One of the main challenge of perception is related to discovery of relevant features of such histories. This problem will be discussed in our next paper.

Acknowledgements. The research has been supported by the individual research project realized within Homing Plus programme, edition 3/2011, of Foundation for Polish Science, co-financed from European Union, Regional Development Fund, and by the grants 2011/01/D/ST6/06981 from the Polish National Science Centre and NN516 077837 from the Ministry of Science and Higher Education of the Republic of Poland.

References

1. Arbib, M.A.: The Metaphorical Brain 2: Neural Networks and Beyond. Willey & Sons (1989)
2. Barwise, J., Seligman, J.: Information Flow: The Logic of Distributed Systems. Cambridge University Press (1997)
3. Bazan, J.G.: Hierarchical Classifiers for Complex Spatio-Temporal Concepts. In: Peters, J.F., Skowron, A., Rybiński, H. (eds.) Transactions on Rough Sets IX. LNCS, vol. 5390, pp. 474–750. Springer, Heidelberg (2008)
4. Goldin, D., Smolka, S., Wegner, P. (eds.): Interactive Computation: The New Paradigm. Springer, Heidelberg (2006)
5. Jankowski, A., Skowron, A.: Wisdom Technology: A Rough-Granular Approach. In: Marciniak, M., Mykowiecka, A. (eds.) Bolc Festschrift. LNCS, vol. 5070, pp. 3–41. Springer, Heidelberg (2009)
6. Marr, D.: Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. Freeman (1982)
7. Nguyen, S.H., Bazan, J.G., Skowron, A., Nguyen, H.S.: Layered Learning for Concept Synthesis. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B., Świniarski, R.W., Szczuka, M.S. (eds.) Transactions on Rough Sets I. LNCS, vol. 3100, pp. 187–208. Springer, Heidelberg (2004)
8. Nguyen, T.T., Skowron, A.: Rough-Granular Computing in Human-Centric Information Processing. In: Bargiela, A., Pedrycz, W. (eds.) Human-Centric Information Processing Through Granular Modelling. SCI, vol. 182, pp. 1–30. Springer, Heidelberg (2009)
9. Noë, A.: Action in Perception. MIT Press (2004)
10. Pawlak, Z.: Information Systems – theoretical foundation. Information Systems 6, 205–218 (1981)
11. Pawlak, Z.: Rough sets. International Journal of Computing and Information Sciences 18, 341–356 (1982)
12. Pawlak, Z.: Rough sets. In: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers (1991)
13. Pawlak, Z., Skowron, A.: Rudiments of rough sets. Information Science 177, 3–27 (2007)
14. Pedrycz, W., Skowron, A., Kreinovich, V. (eds.): Handbook of Granular Computing. John Wiley & Sons (2008)
15. Poggio, T., Smale, S.: The mathematics of learning: Dealing with data. Notices of the AMS 50(5), 537–544 (2003)
16. Skowron, A., Stepaniuk, J.: Hierarchical modelling in searching for complex patterns: constrained sums of information systems. Journal of Experimentaland Theoretical Artificial Intelligence 17(1-2), 83–102 (2005)
17. Skowron, A., Wasilewski, P.: Information Systems in Modeling Interactive Computations on Granules. In: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (eds.) RSCTC 2010. LNCS, vol. 6086, pp. 730–739. Springer, Heidelberg (2010)
18. Skowron, A., Wasilewski, P.: An Introduction to Perception Based Computing. In: Kim, T.-H., Lee, Y.-H., Kang, B.-H., Ślęzak, D. (eds.) FGIT 2010. LNCS, vol. 6485, pp. 12–25. Springer, Heidelberg (2010)
19. Skowron, A., Wasilewski, P.: Information systems in modeling interactive computations on granules. Theoretical Computer Science 412(42), 5939–5959 (2011)
20. Skowron, A., Wasilewski, P.: Toward interactive rough–granular computing. Control & Cybernetics 40(2), 1–23 (2011)
21. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press (1998)

22. Valiant, L.:
http://people.seas.harvard.edu/~valiant/researchinterests.html
23. Vapnik, V.: Statisctical Learning Theory. John Wiley & Sons (1998)
24. Vapnik, V., Gilad-Bachrach, R.: Learning Has Just Started: an interview with Prof. Vladimir Vapnik (2008), http://learningtheory.org
25. Zadeh, L.A.: From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. IEEE Transactions on Circuits and Systems 45, 105–119 (1999)
26. Zadeh, L.A.: A new direction in AI-toward a computational theory of perceptions. AI Magazine 22(1), 73–84 (2001)

Line Graph for Weighted Networks toward Overlapping Community Discovery

Tetsuya Yoshida

Abstract. We propose generalized line graph for weighted networks toward overlapping community discovery from the networks. Community discovery from a network has often been conducted by assigning each node in a network only to one community. However, in real world networks, a node (e.g., user) might belong to several communities. For undirected networks without self-loops, we propose to generalize line graph by defining the weights in the line graph based on the weights in the original network. Based on the line graph representation, a node can be assigned to more than one community by assigning the links adjacent to the node to the corresponding communities. Various properties of the proposed generalized line graph are clarified, and the properties indicate that our proposal is a natural extension of the conventional line graph. Preliminary experiments are conducted over several real-world networks, and the results indicate that the proposed generalized line graph can improve the quality of the discovered overlapping communities.

1 Introduction

Networks (e.g.,the Internet) have been widely utilized as the daily communication media these days. Besides sending and receiving business messages over the network, social networks (e.g., Facebook, Twitter, etc.) play the role of social infrastructure and have been widely utilized [5]. In social networks, users are treated as nodes in a network, and they are connected to each other via links depending on their social relations. Various research efforts have been conducted for analyzing networks and finding out structures in the networks [11, 5].

When users interact with each other over a network, it is important to decide to which community a user would likely to belong to. For instance, some specific

Tetsuya Yoshida
Graduate School of Information Science and Technology, Hokkaido University,
N-14 W-9, Sapporo 060-0814, Japan
e-mail: yoshida@meme.hokudai.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 403–413.
springerlink.com

information to a community (e.g., coupon) can be sent to the users in the community. Community discovery from networks is usually defined as a task of detecting (dense) subgraphs as communities with respect to the observed links in a network [9, 8, 6, 7, 13]. In most previous approaches, community discovery from a network has been conducted by assigning each node in a network only to one community. However, in real world networks, a node might belong to several communities in the network.

We propose generalized line graph for weighted networks toward overlapping community discovery from the networks. For undirected networks without self-loops, we propose to generalize line graph by defining the weights in the line graph based on the weights in the original network. For an undirected network, their links in the original network are mapped to nodes in the corresponding line graph, and the mapped nodes are connected based on the connectivity in the original network. Community discovery over a line graph corresponds to link partitioning of the original network [3]. By assigning the links adjacent to a node to the corresponding communities, the node in the original network can be assigned to more than one community. Various properties of the proposed generalized line graph are clarified, and the properties indicate that the proposed line graph representation is a natural extension of the conventional line graph.

Preliminary experiments are conducted over several real-world networks, and the results indicate that the proposed generalized line graph can improve the quality of the discovered overlapping communities. The results are encouraging, especially with respect to the quality of the discovered communities over the proposed line graph representation.

Section 2 explains the overview of community discovery from networks. Section 3 describes the details of the proposed line graph representation and its properties. Section 4 reports the experiments and discusses the results. Section 5 summarizes our contributions and suggests future directions.

2 Community Discovery from Social Networks

2.1 Preliminaries

We use a bold italic lowercase letter to denote a vector, and a bold normal uppercase letter to denote a matrix. $\mathbf{X}_{ij}$ stands for the element in a matrix $\mathbf{X}$. and $\mathbf{X}^T$ stands for the transposition of $\mathbf{X}$. $1_n \in \mathbb{R}^n$ stands for a column vector where each element is 1, and 1_n^T represents the transposed row vector.

Let n stands for the number of nodes in a network G, and m stands for the number of links [1]. Since most social networks are represented as undirected graph without self-loops [5], we focus on this type of networks in this paper.

The connectivity of a network is usually represented as a square matrix $\mathbf{A} \in \{0,1\}^{n \times n}$ which is called an adjacency matrix. $\mathbf{A}_{ij} = 1$ when the pair of vertices

[1] We also call a network as a graph, a node as a vertex, and a link as an edge.

(v_i, v_j) is connected; otherwise, 0. For an undirected graph without self-loops, the corresponding adjacency matrix $\mathbf{A}$ is symmetric and its diagonal elements are set to 0. The vector $k = \mathbf{A}1_n$ denotes the degree vector, where k_i represents the degree (number of links) of vertex i.

For a network G, an incidence matrix $\mathbf{B} \in \{0,1\}^{n\times m}$ is defined, where $\mathbf{B}_{i\alpha}$ =1 if a link α is related to a node i in G, and 0 otherwise [2]. Following properties hold for the incidence matrix $\mathbf{B}$:

$$\mathbf{B}1_m = k \tag{1}$$
$$1_n^T \mathbf{B} = 2 1_m^T \tag{2}$$

2.2 Modularity

A quality measure called modularity has been widely utilized in community discovery [6]. Under so-called *null model*, the modularity Q of a network G is defined in terms of its adjacency matrix $\mathbf{A}$ as:

$$Q = \frac{1}{2m} \sum_{C\in\mathscr{P}} \sum_{i,j\in C} (\mathbf{A}_{ij} - \frac{k_i k_j}{2m}) \tag{3}$$

where $\mathscr{P}$ stands for the (node) partition of the network, and C runs over the communities in $\mathscr{P}$. Communities with larger modularity are considered as better partitioning of nodes in a network.

2.3 Community Discovery Methods

Various researches have been conducted on community discovery from social networks. In most previous approaches, hard assignment of each node into the corresponding community is conducted.

LabelPropagation[9] conducts community discovery based on majority voting of the community labels of the neighboring nodes. Initially, every node is given a unique community label. Then, it repeats the relabelling process where each node adopts the label that most of its neighbors currently have.

LeadingEigenvector [6] utilizes the leading eigenvector (with the largest eigenvalue) of the so-called modularity matrix, and recursively separates the vertices into two communities based on the sign of the corresponding element in the eigenvector.

Walktrap [8] utilizes short random walks over a network to measure the distance (metric) between nodes as well as the distance between communities in the network. This metric is utilized to conduct the agglomerative clustering (partition) of a network into communities.

3 Line Graph for Weighted Networks

We propose a generalized line graph representation for weighted networks with which a node can be assigned to more than one communities.

3.1 Overlapping Communities Discovery over Line Graph

In most previous approaches, hard clustering was conducted for community discovery in the sense that each node is assigned only to one community. However, in real world networks, a node might belong to several communities. Generally speaking, two kinds of approaches can be considered for enabling the assignment of a single node into multiple communities:

1) invent (or, utilize) soft-clustering methods (e.g., EM algorithm [1]) which would be applied to the original representation of a network,
2) transform the representation of a newtork so that off-the-shelf hard-clustering methods can be applied

In this work, we take the above approach **2)** to realize overlapping community discovery from neworks.

The paper [3] proposed to utilize the line graph of a network based on the 0-1 adjancy matrix of the network. A line graph $L(G)$ of an undirected simple graph G (without self-loops) is defined as [2]:

- an edge (link) in G is mappted to a vertex (node) in $L(G)$
- two vertices in $L(G)$ are connected if the corresponding edges in G share a vertex in G

Among various properties of line graphs, Whitney's uniqueness theorem [12] states that the structure of G can be recovered from its corresponding line graph $L(G)$ except G is neigher a triangle nor a star of four nodes.

Since links adjacent to a node in the original network G are represented as different nodes in the corresponding line graph $L(G)$, by applying some off-the-shelf partitioning based community discovery method to $L(G)$, the links in the original network G can be assigned to communities. By utilizing the communities of the links to which a node in G are adjacent to, each node in the original network G can be assigned to more than one communities.

3.2 Previous Line Graph Representation

For an undirected graph G which consists of n nodes and m edges, let $\mathbf{B}$ stands for its 0-1 incidence matrix in Section 2.1. It is known that the 0-1 adjacency matrix $\mathbf{A}$ of G can be represented in terms of $\mathbf{B}$. On the other hand, the 0-1 adjacency matrix $\mathbf{C}$ of the corresponding line graph $L(G)$ can be represented as:

$$\mathbf{C} = \mathbf{B}^T\mathbf{B} - 2\mathbf{I}_m \tag{4}$$

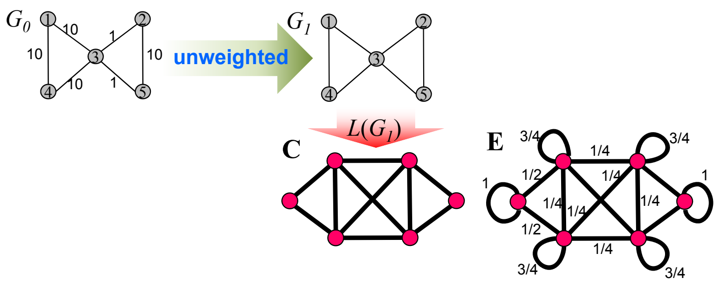


Fig. 1 previous Line Graph representations

where $\mathbf{I}_m$ stands for the square identity matrix of size m. Note that the line graph $L(G)$ defined in eq.(4) is unweighted in the sense that all the links in $L(G)$ share the common weight. Since nodes in $L(G)$ correspond to links in the original graph G, this means that the line graph with the adjacency matrix $\mathbf{C}$ in eq.(4) cannot differentiate to what extent the links in the original network are related to each other.

Based on the random walk interpretation of the modularity in eq.(3), three types of "weighted" line graphs were proposed in [3] in terms of the 0-1 adjacency matrix $\mathbf{A}$ of G. Among them, the following matrices are recommended for the line graph representation:

$$\mathbf{E} = \mathbf{B}^T\mathbf{D}^{-1}\mathbf{B} \tag{5}$$

$$\mathbf{E}_1 = \mathbf{B}^T\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B} \tag{6}$$

where $\mathbf{D}$ represents a diagonal matrix with $\mathbf{D}_{ii} = k_i$ [2]. The diagonal matrix $\mathbf{D}$ is also represented as $diag(k)$.

3.2.1 Issues in Line Graph Representation

The weighted adjacency matrices $\mathbf{E}$ in eq.(5) and $\mathbf{E}_1$ in eq.(6) enable to differentiate to what extent the links in the original network are related to each other over the corresponding line graph. However, these matrices (and other adjacency matrices of $L(G)$ proposed in [3]) are defined solely in terms of connectivity (degree) in the original network G.

Let's consider community discovery from a weighted graph G_0 in Fig. 1. The number over the edges in G_0 represents the weights, which corresponds to the similarities between nodes. According to the weights, the node 1, 3, 4 form a community, and the node 2, 5 form another community (since these nodes are connected with links with larger weights). However, in previous approach, the weighted graph G is treated (or, converted) as an unweighted graph G_1, and the adjacency matrix of the corresponding $L(G)$ is defined based on G_1.

As illustrated in the above example, although weights can be defined over a line graph in previous approach, the defined weights are irrelevant to the weights on

[2] The i-th diagonal element is set to the i-th element of k in $\mathbf{D}$.

the original network G. Thus, even when the original network G is defined as a weighted graph, the weights in G cannot be reflected on the weights on $L(G)$ in previous approach.

3.3 Line Graph Representation for Weighted Networks

To alleviate the above problems, we propose to generalize the line graph representation based on the weights in the original network. Especially, we propose to utilize the weights in the original network and define several matrices in order to define a generalized line graph representation.

Suppose an undirected graph G (with n nodes and m links) contains non-negative weights over links, and w_{ij} stands for the weight on the link between node i and node j. We assume that the weight w_{ij} can be interpreted as the similarity between the nodes.

Let $\tilde{\mathbf{A}}$ stands for a generalized (or, weighted) adjacency matrix of a network G where $\tilde{\mathbf{A}}_{ij} = w_{ij}$. When the weights are interpreted as similarities, the representation of $\tilde{\mathbf{A}}$ is also utilized in spectral clustering [10] in machine learning research. Note that the modularity in eq.(3) can also be defined for $\tilde{\mathbf{A}}$, by replacing $2m$ in eq.(3) with $\sum_{i,j} \tilde{\mathbf{A}}_{ij}$.

Based on $\tilde{\mathbf{A}}$, we define several matrices. First, we define the following vector and matrix:

$$\tilde{k} = \tilde{\mathbf{A}} 1_n \tag{7}$$

$$\tilde{\mathbf{D}} = diag(\tilde{k}) \tag{8}$$

The vector $\tilde{k}$ in eq.(7) represents the sum of weights on links adjacent to each node in G. The matrix $\tilde{\mathbf{D}}$ in eq.(8) corresponds to the generalization of the diagonal matrix $\mathbf{D}$ in eq.(5) and eq.(6).

Finally, we define a generalized incidence matrix $\tilde{\mathbf{B}}$ based on the weights in the original network G (which are represented as $\tilde{\mathbf{A}}$). For a link α=(i, j) which is adjacent to a node i and a node j in G, $\tilde{\mathbf{B}}_{i\alpha}$ and $\tilde{\mathbf{B}}_{j\alpha}$ are set to the weight w_{ij} in G; other elements in the α-th column of $\tilde{\mathbf{B}}$ are set to 0. Thus, as in the generalization from $\mathbf{A}$ to $\tilde{\mathbf{A}}$, the standard binary 0-1 incidence matrix $\mathbf{B}$ is generalized based on the edge weight in $\tilde{\mathbf{B}}$.

3.3.1 Properties of Generalized Matrices

Following properties hold for the above matrices:

$$\tilde{\mathbf{B}} 1_m = \tilde{k} \tag{9}$$

$$1_n^T \tilde{\mathbf{B}} = 2w^T \tag{10}$$

where $w \in \mathbb{R}^m$ in eq.(10) is the weight vector in G. By comparing the above properties with eq.(1) and eq.(2). we can see that the proposed weighted incidence matrix

$\tilde{\mathbf{B}}$ is a natural generalization of the conventional 0-1 incidence matrix $\mathbf{B}$ in Section 2.1. Especially, $1_m \in \mathbb{R}^m$ in eq.(2) is generalized to the weight vector $w \in \mathbb{R}^m$ in eq.(10).

3.3.2 Generalized Line Graph

For an undirected graph G, previous line graph representations are often defined in terms of its incidence matrix $\mathbf{B}$ in in Section 2.1. As long as the line graph representation is defined in terms of $\mathbf{B}$, its representation can be generalized based on the proposed generalized matrices. For instance, the line graph representations in Section 3.3 can be generalized based on the above matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{D}}$ as:

$$\tilde{\mathbf{C}} = \tilde{\mathbf{B}}^T\tilde{\mathbf{B}} - 2diag(w') \tag{11}$$

$$\tilde{\mathbf{E}} = \tilde{\mathbf{B}}^T\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{B}} \tag{12}$$

$$\tilde{\mathbf{E}}_1 = \tilde{\mathbf{B}}^T\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1}\tilde{\mathbf{B}} \tag{13}$$

where $w' = w \odot w$. Here, $\odot$ stands for the Hadamard product (element-wise product) [4].

3.3.3 Properties of Generalized Line Graph Representation

Theorem 1. *Following properties hold for the previous line graph representations and the proposed representations:*

$$1_m^T\mathbf{C} = (k - 1_n)^T\mathbf{B} \tag{14}$$

$$1_m^T\tilde{\mathbf{C}} = (\tilde{k} - 1_n)^T\tilde{\mathbf{B}} \tag{15}$$

$$1_m^T\mathbf{E} = 1_m^T\mathbf{E}_1 = 21_m^T \tag{16}$$

$$1_m^T\tilde{\mathbf{E}} = 1_m^T\tilde{\mathbf{E}}_1 = 2w^w \tag{17}$$

Proof

$$1_m^T\mathbf{C} = 1_m^T(\mathbf{B}^T\mathbf{B} - 2\mathbf{I}_m) = k^T\mathbf{B} - 21_m^T = k^T\mathbf{B} - 1_n^T\mathbf{B} = (k - 1_n)^T\mathbf{B}$$

The second rewriting follows based on eq.(1) and eq.(2). Similarly, by utilizing eq.(9) and eq.(10), we can prove eq.(15). On the other hand, the same properties can be utilized to show eq.(17):

$$\begin{aligned} 1_m^T\mathbf{E} &= 1_m^T\mathbf{B}^T\mathbf{D}^{-1}\mathbf{B} = k^T\mathbf{D}^{-1}\mathbf{B} = 1_n^T\mathbf{B} = 21_m \\ 1_m^T\mathbf{E}_1 &= 1_m^T\mathbf{B}^T\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B} = k^T\mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}\mathbf{B} = 1_n^T\mathbf{A}\mathbf{D}^{-1}\mathbf{B} \\ &= k^T\mathbf{D}^{-1}\mathbf{B} = 1_n^T\mathbf{B} = 21_m \end{aligned}$$

Similarly, by utilizing eq.(9) and eq.(10), we can prove eq.(17). □

Again, from Theorem 1, we can see that the proposed generalized line graph representation is a natural extension of the previous weighted line graph representation.

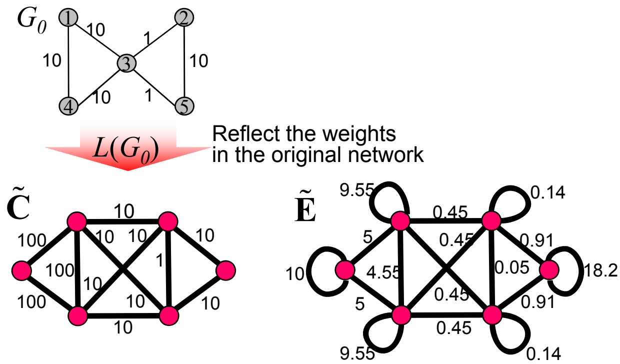


Fig. 2 proposed Line Graph representations over a weighted network

3.3.4 Illustrating Examples

For the same weighted graph G_0 in Fig. 1, Fig. 2) illustrates that our approach directly defines the weights over the line graph by reflecting the weights on G_0. In Fig. 2, since the weights on the original network G_0 differs depending on the links, weights on $\tilde{\mathbf{E}}_1$ does not become symmetric in the corresponding line graph (since the original information (weights) on G_0 are not discarded in our approach).

4 Preliminary Experiments

4.1 Experimental Settings

4.1.1 Datasets

Preliminary experiments were conducted over real-world networks. We collected several benchmark networks with weighted links. Utilized networks are shown in Table 1 and Table 2. Networks in Table 1 are available as GML (graph markup language) format[3]. Networks in Table 2 are co-authorship networks among researchers[4]. The co-authorship relations is represented as an adjacency matrix $\tilde{\mathbf{A}}$, where $\tilde{\mathbf{A}}_{ij}$ stands for the number of papers which are co-authored by authors (nodes) i and j.

4.1.2 Community Discovery Methods

We utilized the community discovery methods described in Section 2.3, namely: 1) labelPropagation[9], 2) leadingEigenvector [6], 3) walktrap [8]. Since all of

[3] http://www-personal.umich.edu/~mejn/netdata/ (celegansneural was converted into undirected network in the experiment)

[4] Pascal: http://analytics.ijs.si/~blazf/pvc/data.html
IV'04: http://iv.slis.indiana.edu/ref/iv04contest/

Table 1 Weighted Networks

dataset	#nodes	#links
lesmis	77	254
celegansneural	297	2345
netscience	1589	2742

Table 2 Co-authorship Networks

dataset	# nodes	#links
Pascal	313	381
IV'04	735	1360

them conduct hard clustering, these were utilized to conduct link partitioning of the constructed line graphs.

4.1.3 Quality Measures

In each method, we evaluated the modularity in eq.(3) with respect to the link partitioning of the constructed line graphs. Communities with larger modularity are considered as better in terms of modularity measure. Note that the modularity is calculated with respect to the weighted adjacency matrix in the line graph representation ($2m$ in eq.(3) becomes the sum of the weights).

In addition, we counted the number of discovered communities, since the link partitioning generally leads to larger number of communities. Furthermore, since the utilization of line graph representation enables overlapping community discovery, we recorded the maximum number of communities per node (as an indirect measure of diversity of overlapping communities).

4.1.4 Type of Line Graph

One good property of $\mathbf{E}_1$ in eq.(6) is that, the modularity with respect to node partitioning of a network G is equal to the modularity with respect to the link partitioning of $L(G)$ [3] . Thus, since community discovery in terms of the node partitioning of G can be pursued in terms of link partitioning of G over the line graph representation with $\mathbf{E}_1$, we used $\mathbf{E}_1$ for 0-1 adjacency matrix $\mathbf{A}$ and $\tilde{\mathbf{E}}_1$ for the weighted $\tilde{\mathbf{A}}$ in the experiments, and compared their performance in terms of the modularity.

4.2 Results

The results are summarized in Table 3. The column for "with $\mathbf{E}_1$" and that for "with $\tilde{\mathbf{E}}_1$" show the modularity of the discovered communities over the corresponding line graphs (the larger ones are shown in bold). The column for "#comm." shows the number of discovered communities, and that for "#max." shows the maximum number of communities per node. Since walktrap does not work over disconnected networks, the results for netscience, Pascal and IV'04 (these are disconnected networks) are not shown in Table 3.

The results in Table 3 indicate that, by utilizing the proposed line graph representation in Section 3.3, it was possible to discover better communities by applying standard node partitioning based methods in terms of modularity in eq.(3). The

Table 3 Results

dataset	method	with $\mathbf{E}_1$	#comm.	#max.	with $\tilde{\mathbf{E}}_1$	#comm.	#max.
lesmis	labelPropagation	**0.473**	6	4	0.448	7	4
	leadingEigenvector	0.424	9	6	**0.451**	8	6
	walktrap	0.435	4	2	**0.515**	6	4
celegansneural	labelPropagation	0.000	1	1	**0.297**	2	2
	leadingEigenvector	0.198	17	10	**0.271**	17	10
	walktrap	0.291	2	2	**0.319**	3	3
netscience	labelPropagation	0.837	510	10	**0.857**	470	14
	leadingEigenvector	**0.894**	153	3	0.826	270	4
Pascal	labelPropagation	0.784	106	8	**0.812**	96	4
	leadingEigenvector	0.857	41	3	**0.875**	41	3
IV'04	labelPropagation	**0.861**	259	9	0.858	252	8
	leadingEigenvector	**0.862**	128	3	0.852	119	3

proposed approach contributed to larger modularity over the datasets in Table 1, especially over celegansneural dataset (the second row in Table 3). On the other hand, although it outperformed the previous approach (with $\mathbf{E}_1$) over Pascal dataset in Table 2, but not for IV'04.

As for the number of discovered communities and the maximum number of communities per node, the results differed depending on the datasets. Still, the number of discovered communities was similar with both methods (except for netscience). On the other hand, the maximum number of communities per node slightly increased with the proposed approach. We plan to conduct more detailed investigation of the discovered communities.

4.3 Discussions

The proposed line graph representation in Section 3.3 can reflect the weights on the original network, and enables overlapping community discovery via link partitioning over the weighted line graph. The difference with the previous line graph representation in Section 3.2 will become prominent as the diversity in the weights on the original network increases. Although reflecting the original weights over the line graph representation does not necessarily guarantee the discovery of "better" overlapping communities (e.g., in terms of modularity), we believe that reflecting the information in the original network leads to better understanding (and, discovery) of overlapping communities.

5 Concluding Remarks

We proposed generalized line graph for weighted networks toward overlapping community discovery from the networks. For undirected networks without self-loops, we proposed to generalize line graph by defining the weights in the line graph based on the weights in the original network. By assigning the links adjacent to a node to the corresponding communities, the node can be assigned to more than one community. Various properties of the proposed generalized line graph are proved, and the properties indicate that the proposed line graph representation is a natural extension of the conventional line graph.

Preliminary experiments are conducted over several real-world networks, and the results indicate that the proposed generalized line graph can improve the quality of the discovered overlapping communities. The results are encouraging, especially with respect to the quality of discovered communities. We plan to conduct more in-depth analysis of the discovered communities and extend the proposed method based on the analysis in near future.

Acknowledgements. This work is partially supported by Casio Science Promotion Foundation.

References

1. Dempster, A., Laird, N., Rubin, D.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39(2), 1–38 (1977)
2. Diestel, R.: Graph Theory. Springer (2006)
3. Evans, T., Lambiotte, R.: Line graphs, link partitions, and overlapping communities. Physical Review E 80(1), 016105, 1–8 (2009)
4. Harville, D.A.: Matrix Algebra From a Statistican's Perspective. Springer (2008)
5. Mika, P.: Social Networks and the Semantic Web. Springer (2007)
6. Newman, M.: Finding community structure using the eigenvectors of matrices. Physical Review E 76(3), 036104 (2006)
7. Newman, M.: Networks: An Introduction. Oxford University Press (2010)
8. Pons, P., Latapy, M.: Computing communities in large networks using random walks. Journal of Graph Algorithms 10(2), 191–218 (2006)
9. Raghavan, U., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Physical Review E 76, 036106 (2007)
10. von Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing 17(4), 395–416 (2007)
11. Watts, D.J.: Six Degrees: The Science of a Connected Age. W W Norton & Co., Inc. (2004)
12. Whitney, H.: Congruent graphs and the connectivity of graphs. American Journal of Mathematics 54, 150–168 (1932)
13. Yoshida, T.: Toward finding hidden communities based on user profile. Journal of Intelligent Information Systems (2011) (accepted)

Modeling of Climate Data in Terms of AR(1) Process Contaminated with Noise

Yuta Arai and Hiroshi Iyetomi

Abstract. It is widely recognized that teleconnections (correlations between climates at remote places) such as El Niño play a crucial role in understanding abnormal weather phenomena. To extract such correlations in multivariate climate data, the random matrix theory (RMT) combined with the principal component analysis (PCA) can be successfully used; the RMT has power to distinguish between statistically meaningful correlations and noises. Here we demonstrate that sea level pressure (SLP), which is one of basic meteorological measurements for teleconnections, have characteristic autocorrelations. Unfortunately the standard RMT is not able to distinguish between autocorrelations and cross-correlations. We show that an AR(1) process contaminated with noise reproduces the autocorrelations of the SLP quite well. Then we estimate autocorrelation effects on the eigenvalue distribution of the SLP correlation matrix, which makes the extraction procedure of genuine cross-correlations more reliable.

1 Introduction

We often use the principal component analysis to extract the significant correlations in the multivariate time series data. Then the RMT was used to estimate how many principal components should be retained as being statistically significant[1, 2, 3, 4, 5, 6, 7]. In this case, the RMT works as a null hypothesis.

We consider a random matrix H which is an $N \times T$ matrix; its elements $\{h_{ij}\}$ $(1 \leq i \leq N, 1 \leq j \leq T)$ are random variables with zero mean and unit variance, and totally independent. The random correlation matrix is calculated as

Yuta Arai
Graduate School of Science and Technology, Niigata University
Ikarashi, Niigata 950-2181, Japan
e-mail: y.arai@phys.sc.niigata-u.ac.jp

Hiroshi Iyetomi
Faculty of Science, Niigata University, Ikarashi, Niigata 950-2181, Japan

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 415–422.
springerlink.com

$$C = \frac{1}{T} HH^T. \tag{1}$$

In the limit $N, T \to \infty$ with fixed $Q \equiv T/N$, the probability density function $\rho(\lambda)$ of eigenvalue λ of the random correlation matrix C is analytically obtained as

$$\rho(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}, \tag{2}$$

$$\lambda_\pm = 1 + 1/Q \pm 2\sqrt{1/Q}, \tag{3}$$

for $\lambda \in [\lambda_-, \lambda_+]$, where λ_- and λ_+ are the minimum and maximum eigenvalues of C, respectively. Therefore, the RMT serves as a powerful null hypothesis for the PCA. So the eigenvalue which is larger than λ_+ is regarded as the appearance of the significant correlations. However the analytic result Eq. (2) is valid for $\{h_{ij}\}$ are mutually independent. If the time series data contains the autocorrelations, we can not distinguish between the autocorrelations and the cross-correlations.

2 SLP Data

Here we use the monthly data [8] for sea level pressure (SLP) which are given on the 2-degree grid for 80-year period of 1910-1990 ($T = 960$). Since the whole data is too large, we reduced the number N of observation points from 16200 to 948. To remove seasonal variation involved in the original data, we then defined a normal value by averaging the data in each month over 30 years ranging from 1960 through 1989. We analyze deviation of the SLP from its normal value thus calculated, which is referred to as anomaly.

Figure 1(a) shows the autocorrelations of the SLP anomalies for time lag 1, 2 and 3 months. And the corresponding autocorrelations of monthly difference of the same data are given in Fig. 1(b). From these figures, we see that the autocorrelations of the SLP data and their monthly difference have rather universal behavior in all observation points. Therefore we calculate average of the autocorrelations

$$\overline{\phi}(\tau) = \frac{1}{N} \sum_{i=1}^{N} \phi_i(\tau), \tag{4}$$

$$\overline{\phi}'(\tau) = \frac{1}{N} \sum_{i=1}^{N} \phi_i'(\tau), \tag{5}$$

where $\phi_i(\tau)$ denotes the autocorrelation of the SLP data and $\phi_i'(\tau)$ is that of the monthly differences at τ month time lag. The two results are compared in Fig. 2.

Especially we note

$$\overline{\phi}(1) \simeq 0.17, \tag{6}$$

$$\overline{\phi}'(1) \simeq -0.46. \tag{7}$$

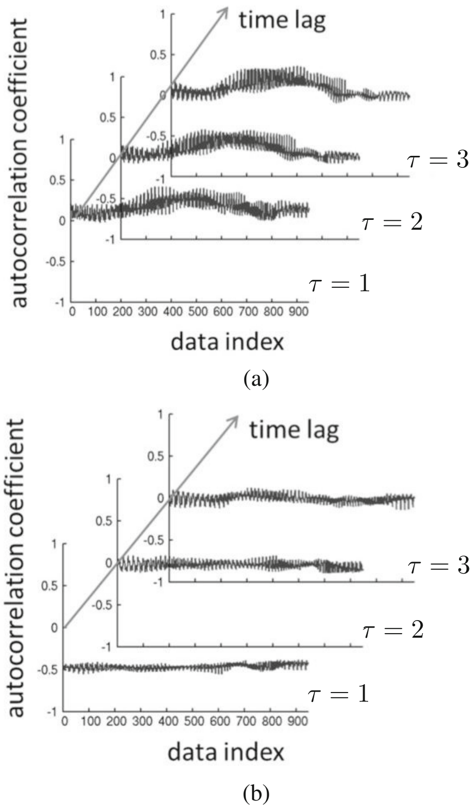


Fig. 1 Autocorrelation coefficients at one-, two- and three-month time lags for the SLP data (a) and for their monthly differences (b)

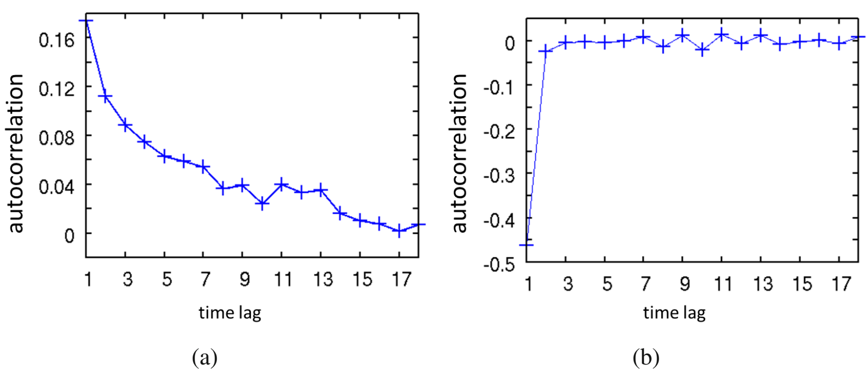


Fig. 2 (a) The autocorrelation function for the SLP data themselves averaged over the observation points; (b) the corresponding function for monthly differences of the SLP data

Figure 2 demonstrates that the autocorrelation of the SLP has a long-range tail, in contrast, the autocorrelation of the corresponding monthly differences is very short-lived (nearly for 1 month).

These results force us to study autocorrelation effects on the eigenvalue distribution of the correlation matrix of the SLP data.

3 Stochastic Processes for the SLP Data

Here we work out a stochastic process which explains the characteristics of autocorrelation in the SLP data as has been shown in Figs. 1 and 2.

3.1 AR(1) Process

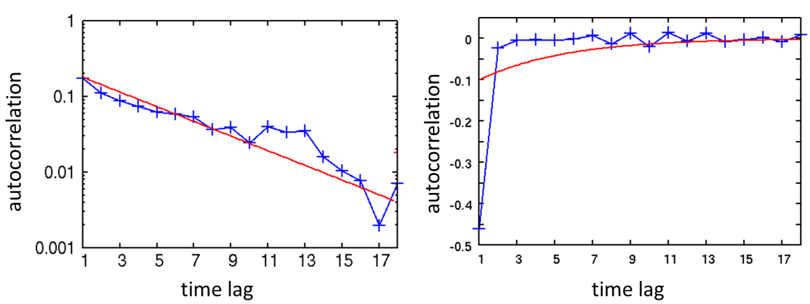
First, we adopt an autoregressive process of order 1, referred to as AR(1), to model the autocorrelations of the SLP data. The AR(1) process is given by

$$x_t = \rho x_{t-1} + \varepsilon_t, \tag{8}$$

where the parameter ρ takes a value of $|\rho| < 1$, and ε_t is a random variable with zero mean and unit variance. In this model the autocorrelation at τ time lag, exactly calculated as

$$\phi_{\mathrm{AR}(1)}(\tau) = \rho^{\tau}, \tag{9}$$

has a long-range tail. Figure 3(a) indicates the autocorrelation of the SLP data for the long-range nature of decays in the form (9) with $\rho \simeq 0.8$. We see that the AR(1) process is able to account the autocorrelation of the SLP data quite well.



(a) (b)

Fig. 3 Comparison of the autocorrelation function of the AR(1) process with that of the SLP data and their monthly differences. The cross symbols depicts the empirical data and the solid line is the AR(1) process.

On the other hand, the autocorrelation of differences of the AR(1) process is given by

$$\phi'_{\mathrm{AR}(1)}(\tau) = -\frac{1-\rho}{2}\rho^{\tau-1}. \tag{10}$$

In this case, the AR(1) process does no reproduce the autocorrelations of the SLP data. The autocorrelation of monthly difference of the SLP data exists only 1 month time lag. However, the autocorrelation of this model is long range period.

3.2 Random Number Sequence

Alternatively, we consider a series of random numbers to describe the autocorrelation of the SLP data. The random number sequence model is just a special case of the AR(1) process (8) with $\rho = 0$;

$$x_t = \varepsilon_t, \tag{11}$$

The autocorrelation of difference of the random number sequence is given by

$$\phi'_{\mathrm{rand}}(\tau) = \begin{cases} -\frac{1}{2} & (\tau = 1) \\ 0 & (\tau \neq 1) \end{cases} \tag{12}$$

Therefore, this model reproduces the autocorrelation of monthly differences of the SLP data quite well as shown in Fig. 2(b).

However, the random number sequence has no autocorrelations as opposed to the empirical fact as shown in Fig. 2(a).

3.3 AR(1) Process Contaminated with Noise

As previously explained, the AR(1) process and the random number sequence satisfy only either autocorrelations. Then we propose the AR(1) process contaminated with noise. This model is described by

$$\tilde{x}_t = x_t + \eta_t, \tag{13}$$

where x_t is the AR(1) process and η_t is random variable with zero mean and variance σ^2. In the limit $\langle \eta^2 \rangle \to \infty$, Eq. (13) equals the random number sequence. On the other hands, in the limit $\langle \eta^2 \rangle \to 0$, Eq. (13) equals the AR(1) process.

Autocorrelation of this process is calculated as

$$\phi_{\tilde{x}}(\tau) = \frac{\rho^\tau}{1 + \frac{\sigma^2}{\langle x^2 \rangle}}. \tag{14}$$

And the autocorrelation of difference of the AR(1) process contaminated with noise is given by

$$\phi'_{\tilde{x}}(\tau) = \begin{cases} -\frac{1}{2}\frac{(1-2\rho+\rho^2)+\frac{\sigma^2}{\langle x^2\rangle}}{1-\rho+\frac{\sigma^2}{\langle x^2\rangle}} & \text{for } \tau = 1 \\ -\frac{1}{2}\frac{(1-2\rho+\rho^2)\rho^{\tau-1}}{1-\rho+\frac{\sigma^2}{\langle x^2\rangle}} & \text{for } \tau \geq 2 \end{cases}. \tag{15}$$

Here we assume the value of ρ is $\rho = 0.8$ because the autocorrelation of the SLP data decays in the form 0.8^{τ} as shown in Fig. 3(a). Then We calculate the mean square error to estimate the best value of $\frac{\sigma^2}{\langle x^2\rangle}$ in reproducing the autocorrelations of the SLP data. The mean square error is given by

$$\Delta\left(\frac{\sigma^2}{\langle x^2\rangle}\right) = \frac{1}{T}\sum_{\tau=0}^{T-1}\left(\phi_{\sigma^2/\langle x^2\rangle}(\tau) - \phi_{\mathrm{SLP}}(\tau)\right)^2, \tag{16}$$

where $\phi_{\sigma^2/\langle x^2\rangle}(\tau)$ stands for the autocorrelation of the time series $\{\tilde{x}_t\}$ at τ time lag with given each $\frac{\sigma^2}{\langle x^2\rangle}$ and $\phi_{\mathrm{SLP}}(\tau)$, for that of the SLP data at the same time lag.

Table 1 Comparison of the mean square error of the autocorrelations with given $\frac{\sigma^2}{\langle x^2\rangle}$

$\Delta(3.4)$	$\Delta(3.6)$	$\Delta(3.8)$	$\Delta(4.0)$
6.0×10^{-5}	5.8×10^{-5}	5.9×10^{-5}	6.0×10^{-5}

Table 1 shows the result of the mean square errors for each $\frac{\sigma^2}{\langle x^2\rangle}$. From this result, we regard that SLP data follows the AR(1) process contaminated with noise which satisfies under the condition

$$\rho = 0.8, \frac{\sigma^2}{\langle x^2\rangle} = 3.6. \tag{17}$$

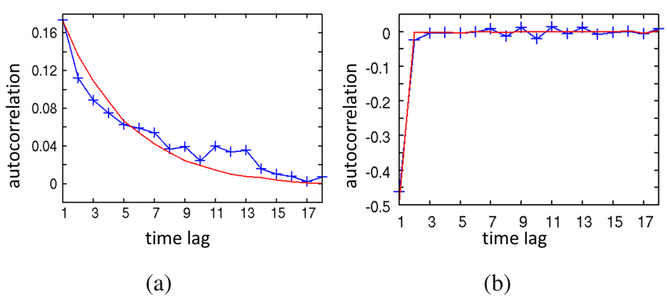


Fig. 4 Comparison of the autocorrelation of the AR(1) process contaminated with noise and that of the SLP data. The right panel corresponds to the difference of the time series data. In each figures, the cross symbol corresponds to the SLP data and the solid line is the AR(1) process contaminated with noise.

Figures 4(a) and (b) compare the autocorrelation of the AR(1) process contaminated with noise and that of the SLP data. In Fig. 4(a), we can see that the AR(1) process contaminated with noise has the autocorrelation for a long time. On the other hands, in Fig. 4(b), the autocorrelation of difference of the AR(1) process contaminated with noise has the autocorrelation at only 1 month time lag. So this model reproduces the autocorrelation of the SLP data.

4 Eigenvalue Distribution

Figure 5 shows the eigenvalue distribution of the AR(1) process contaminated with noise (10^4 samples were generated) and that of the SLP data. Through the influence of the autocorrelation effects, the eigenvalues which larger than λ_+ appear clearly visible. Therefore, the eigenvalues of the SLP data which near the λ_+ are regarded as the appearance of the autocorrelation effects. If we adopt the criterion $\lambda > \lambda_+$, the number of the principal components are 42. However, we adopt the criterion considered the autocorrelation effects, the number of the principal component is 38. Then using this criterion, we can distinguish between the autocorrelations and the cross-correlations.

On the other hands, Fig. 6 shows $\langle \eta_t^2 \rangle$ dependence of the maximal eigenvalue of the correlation matrices which consist of the time series $\{\tilde{x}_t\}$. To neglect the statistical error, 10^4 samples were generated. In the limit $\sigma^2/\langle x^2 \rangle \to \infty$, the maximal eigenvalue equals that of the time series $\{x_{t+1} - x_t\}$ given $\rho = 0$. For small $\sigma^2/\langle x^2 \rangle$, the maximal eigenvalue equals that of the time series $\{x_{t+1} - x_t\}$ given $\rho = 0.8$. Since the time series $\{\tilde{x}_{i,t+1} - \tilde{x}_{i,t}\}$ contains only autocorrelations, the eigenvalues which are larger than λ_+ are regarded as appearance of cross-correlation structures.

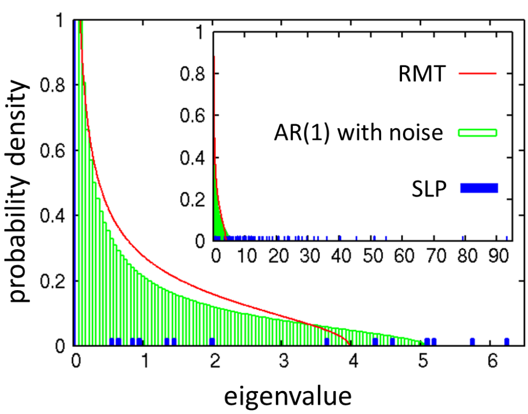


Fig. 5 Comparison of the eigenvalue distribution of the AR(1) process contaminated with noise and that of the SLP data

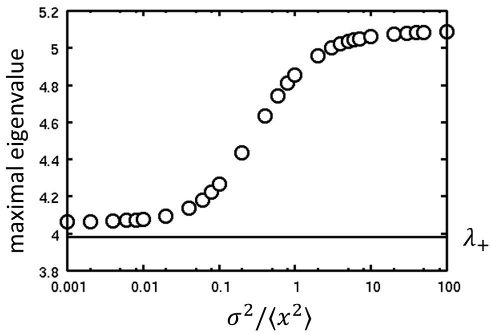


Fig. 6 $\langle \eta_t^2 \rangle$ dependence of the maximal eigenvalue of the correlation matrices which consist of the time series $\{\tilde{y}_t\}$. The block line denotes λ_+.

5 Summary

In this paper, we modeled the SLP data. As a result, the SLP data was modeled by the AR(1) process contaminated with noise. This process reproduce the autocorrelation of the SLP data quite well. Moreover we calculate the eigenvalue distribution of the correlation matrices which consist of the time series $\{\tilde{x}_t\}$. The behavior of its distribution is almost the same as that of the monthly differences of the SLP data. Moreover the view of application for the principal component analysis, we investigate the maximal eigenvalue. Such an extensive study enables us to distinguish between the autocorrelations and the cross-correlations in various data. Detailed analysis of the meaningful correlations of the SLP data is in progress.

Acknowledgements. This work was partially supported by the Program for Promoting Methodological Innovation in Humanities and Social Sciences by Cross-Disciplinary Fusing of the Japan Society for the Promotion of Science and by the Ministry of Education, Science, Sports, and Culture, Grants-in-Aid for Scientific Research (B), Grant No. 22300080 (2010-12).

References

1. Laloux, L., Cizeau, P., Bouchaud, J.P., Potters, M.: Phys. Rev. Lett. 83, 1467 (1999)
2. Santhanam, M.S., Patra, P.K.: Phys. Rev. E 64, 016102 (2002)
3. Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., Guhr, T., Stanley, H.E.: Phys. Rev. E 65, 066126 (2002)
4. Utsugi, A., Ino, K., Oshikawa, M.: Phys. Rev. E 70, 026110 (2004)
5. Kim, D.H., Jeong, H.: Phys. Rev. E 72, 046133 (2005)
6. Kulkarni, V., Deo, N.: Eur. Phys. J. B 60, 101 (2007)
7. Pan, R.K., Sinha, S.: Phys. Rev. E 76, 046116 (2007)
8. http://www.ncdc.noaa.gov/oa/ncdc.html/

Moment Approach for Quantitative Evaluation of Randomness Based on RMT Formula

Mieko Tanaka-Yamawaki, Xin Yang, and Ryota Itoi

Abstract. We develop in this article a quantitative formulation of the randomness-test based on the random matrix theory (RMT-test), in order to compare a subtle difference of randomness between given random sequences. Namely, we compare the moments of the actual eigenvalue distribution to the corresponding theoretical expression that we derive from the formula theoretically derived by the random matrix theory. We employ the moment analysis in order to compare the eigenvalue distribution of the cross correlation matrix between pairs of sequences. Using this method, we compare the randomness of five kinds of random data generated by two pseudo-random generators (LCG and MT) and three physical generators. Although the randomness of the individual sequence can be quantified in a precise manner using this method, we found that the measured values of randomness fluctuate significantly. Taking the average over 100 independent samples each, we conclude that the randomness of the random data generated by the five generators are indistinguishable by the proposed method, while the same method can detect the randomness of the derivatives of the sequences, or the initial part of LCG, which are distinctly lower.

Keywords: Randomness measure, RMT-test, Moment analysis, Eigenvalues of cross correlation matrix, Marcenko-Pastur distribution.

1 Introduction

The random matrix theory (RMT, hereafter) [1,2] has a wide variety of application from the nuclear energy levels [3] to the principal component analysis of stock markets [4-8]. Especially, the latter example has attracted much attention of the community of econo-physics in which many researchers have been working to discover possible rules or structures behind the motion of stock prices. As a byproduct of this activity, that we call the RMT-PCA [7,8] to study correlations of

Mieko Tanaka-Yamawaki · Xin Yang · Ryota Itoi
Department of Information and Knowledge Engineering, Graduate School of Engineering, Tottori University, Tottori, 680-8552 Japan
e-mail: mieko@ike.tottori-u.ac.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 423–432.
springerlink.com

stock prices, we have realized to apply RMT to measure the degree of randomness of given time series.

In the previous meeting of IDT [9], we proposed the first version of a new algorithm of testing the randomness of marginally random sequences, the qualitative version of the RMT-test, and applied that method on two popular pseudo-random numbers, the Linear Congruential Generator (LCG) [10] and the Mersenne Twister (MT) [11] and showed that both pass the test.

In this article we develop a quantitative formulation of the randomness-test based on the random matrix theory (the quantitative version of the RMT-test), in order to compare a subtle difference of randomness between given random sequences. For this purpose we employ the moment method in order to compare the shape of the eigenvalue distribution of the cross correlation matrix obtained from the data to the theoretically derived eigenvalue distribution from the random matrix theory, which is so-called Marcenko-Pastur distribution [12,13]. We test this new method on the same pseudo-random numbers as before, and three physical random numbers which can be downloaded from the web-site of the Institute of Statistical Mathematics [14].

2 Synopsis of the RMT-Test

The method of the RMT-test is outlined as follows [6,7]. We aim to test the randomness of a long 1-dimensional sequence of numerical data, S.

At the first step, we cut S into N pieces of equal length T by discarding the remainder if the length of S is not divisible by T. Each piece, $S_i = (S_{i,1}, S_{i,2}, \ldots, S_{i,T})$, is converted to a normalized vector $x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,T})$ of zero mean and unit variance. Since the original sequence S is random, in general all those vectors are independent. We construct a cross correlation matrix C by taking the inner product of two vectors x_i and x_j,

$$C_{i,j} = \frac{1}{T}\sum_{t=1}^{T} x_{i,t}\, x_{j,t} \tag{1}$$

which is symmetric under the interchange of *i* and *j* by definition. We solve an eigenvalue problem of C

$$C v_k = \lambda_k v_k \quad (k=1,\ldots,N) \tag{2}$$

to obtain λ_k, the k-th eigenvalue, and its eigenvector v_k

The distribution of the eigenvalues is compared [4-9] to the theoretical formula derived in the RMT at the limit of $N \to \infty,\ T \to \infty,\ Q = T/N = \text{const}$.

$$P_{RMT}(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \tag{3}$$

where the upper bound and the lower bound are given by the following formula

$$\lambda_{\pm} = (1 \pm Q^{-1/2})^2 \tag{4}$$

Unlike the qualitative version of the RMT-test [9], we utilize the moment analysis to compare the k-th moments of the eigenvalues, obtained from the Eq. (2) for the data

$$m_k^{EXP} = \frac{1}{N}\sum_{i=1}^{N}(\lambda_i)^k \tag{5}$$

compare to the theoretical moment

$$m_k^{RMT} = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{RMT}(\lambda)d\lambda \tag{6}$$

The explicit formula of moments as a function of Q up to k=6 can be obtained, by substituting Eq.(3) and Eq.(4), into Eq.(6), as follows [15]:

$$m_1^{RMT} = 1 \tag{7}$$

$$m_2^{RMT} = 1 + \frac{1}{Q} \tag{8}$$

$$m_3^{RMT} = 1 + \frac{3}{Q} + \frac{1}{Q^2} \tag{9}$$

$$m_4^{RMT} = 1 + \frac{6}{Q} + \frac{6}{Q^2} + \frac{1}{Q^3} \tag{10}$$

$$m_5^{RMT} = 1 + \frac{10}{Q} + \frac{20}{Q^2} + \frac{10}{Q^3} + \frac{1}{Q^4} \tag{11}$$

$$m_6^{RMT} = 1 + \frac{15}{Q} + \frac{50}{Q^2} + \frac{50}{Q^3} + \frac{15}{Q^4} + \frac{1}{Q^5} \tag{12}$$

The quantified criterion of randomness is identified as the deviation of m_k^{EXP} from m_k^{RMT} :

$$\text{Error} = \frac{m_k^{EXP}}{m_k^{RMT}} - 1 \tag{13}$$

By setting the upper limit of the Error in Eq.(13), we accept the data to be random if

$$|\text{Error}| < \varepsilon \tag{14}$$

and reject it if otherwise.

The quantitative version of the RMT-test can be summarized in the following five steps in Table 1.

Table 1 Algorithm of the quantitative version of the RMT-test

1. Prepare a string to be tested and cut it into N pieces of length T.
2. Each piece, $S_i = (S_{i,1}, S_{i,2}, \ldots, S_{i,T})$, is converted to a normalized vector $x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,T})$ of zero mean, unit variance. Taking the inner product of two vectors x_i and x_j and divide it by T, we make the correlation matrix, C in Eq. (1).
3. Obtain the k-th moment m_k^{EXP} for k=1,…,6 from the trace of C^k.
4. Compute the Error in Eq. (14) from the ratio of m_k^{EXP} and m_k^{RMT}.
5. The string passes the RMT- test if $|\mathrm{Error}| < \varepsilon$ and fails if otherwise.

Note that we do not need to solve the eigenvalue problem in Eq.(2) directly but simply obtain the trace (sum of all the diagonal elements)

$$\sum_{i=1}^{N} (C^k)_{i,i} = \sum_{i=1}^{N} (\lambda_i)^k \tag{15}$$

of the matrix of the k-th power C^k of the cross correlation matrix to obtain the k-th moment in Step 3.

3 Applications of the RMT-Test on Random Sequences

3.1 Discussion on the Reasonable Range of N and T

We need to choose large enough N and T in order to justify the test, since the theoretical formula of the eigenvalue distribution P_{RMT} is derived at the limit of N and T being infinity. In order to determine the practically reasonable value for those parameters, we apply the test on the data taken from the two pseudo random number generators, LCG and MT, for N=200, 300, 400, 500 at Q=3, and compare the moments up to the 6-th order to the corresponding theoretical formula.

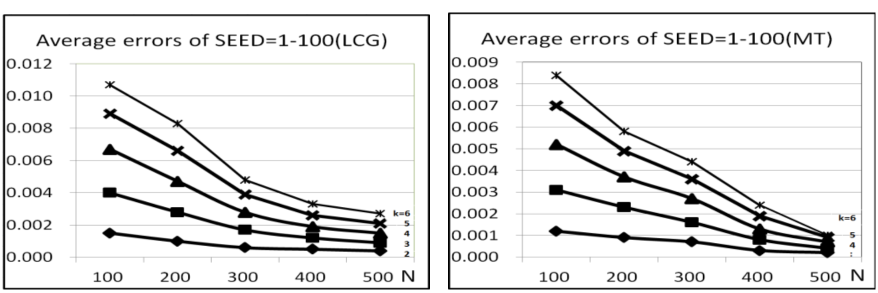


Fig. 1 Errors as a function of N obtained for two pseudo-random data (LCG,MT)

We show the average errors over 100 different seeds, of SEED=1,.., 100, by six lines corresponding to the 1st to the 6th moments from the bottom to the top, for LCG and MT. In both figures, the errors go down gradually as N increases from 200 to 400, and then become stable after N reaches 500. At N=500, the errors become smaller than 0.3%.

3.2 *Quantitative Evaluation of Randomness by Means of Moments*

By using the value Q=3 and N=500 in the last section, we apply the algorithm of Table 1 together with Eq. (15). At this moment, we need to set the constant ε, the upper limit of the Error for the randomness test. This limit is supposed to work for widely-used pseudo-random numbers, and physical random numbers as well. Thus we investigate the level of errors by using the two pseudo-random generators, LCG and MT, that we have used to determine N and T in the last section, and the physically generated random numbers downloaded from a web page. There are three different physical generators, Toshiba, Hitachi, and Tokyo-Electron, offered by the Institute of Statistical Mathematics [14]. We have obtained the results of using those five data for k=2-6, Q=3, T=1500, and showed in Table 2, the average (and the standard deviation in the parentheses) over the 100 independent tests.

Table 2 Errors for the 5 data: the average (SD) over 100 independent tests for Q=3(T=1500)

K	LCG	MT	Toshiba	Hitachi	Tokyo
2	-.0004(.0010)	-.0004(.0009)	-.0004(.0010)	-.0004(.0010)	-.0004(.0009)
3	-.0010(.0026)	-.0009(.0024)	-.0011(.0026)	-.0011(.0025)	-.0009(.0025)
4	-.0018(.0047)	-.0014(.0041)	-.0019(.0046)	-.0019(.0044)	-.0015(.0044)
5	-.0027(.0072)	-.0019(.0062)	-.0026(.0070)	-.0028(.0066)	-.0019(.0067)
6	-.0036(.0100)	-.0022(.0085)	-.0033(.0096)	-.0037(.0092)	-.0021(.0093)

Table 3 Errors for the 5 data: the average (SD) over 100 independent tests for Q=6(T=3000)

K	LCG	MT	Toshiba	Hitachi	Tokyo
2	-.0003(.0006)	-.0001(.0006)	-.0002(.0006)	-.0002(.0006)	-.0002(.0005)
3	-.0008(.0016)	-.0004(.0015)	-.0005(.0015)	-.0006(.0016)	-.0004(.0014)
4	-.0014(.0028)	-.0006(.0027)	-.0008(.0027)	-.0011(.0028)	-.0007(.0026)
5	-.0020(.0043)	-.0009(.0042)	-.0012(.0042)	-.0016(.0043)	-.0009(.0040)
6	-.0026(.0060)	-.0012(.0058)	-.0015(.0058)	-.0020(.0060)	-.0010(.0056)

4 Detecting Off-Randomness by Means of the RMT-Test

Using the moment method that we have developed so far, we now deal with some off-random examples.

4.1 Testing the Initial Part of LCG

The initial part of LCG is expected to have low randomness. In order to quantity the degree of randomness, we apply the quantitative RMT-test on the collection of initial parts of LCG, and compare them with the corresponding data of MT.

Table 4 Randomness of the initial parts of pseudo random numbers

Errors in RMT test for the initial 500 elements		
k	LCG	MT
2	.0046	-.0018
3	.0102	-.0042
4	.0198	-.0064
5	.0356	-.0083
6	.0592	-.0099

From this, we learn that the errors of moment ratio for the initial 500 elements of LCG are considerably large compared to the corresponding elements of MT. As we observed in the previous work [9], the off-randomness of initial 500 elements of LCG can be distinguishable by eyes, by the fact that some eigenvalues exceed the theoretical limit of the RMT.

4.2 Testing the Randomness of Log-Return Sequences

It is customary to convert the price time series $p_1, p_2, ..., p_T$ to the log-return time series $r_1, r_2, ..., r_T$ by means of Eq.(16) in the financial analysis, in order to eliminate the unit/size dependence of different stock prices.

$$r_i = \log(P_i/P_{i-1}) \tag{16}$$

However, this process involves the same P_i for r_i and for r_{i+1}. Because of this, the time series of log-returns loose the randomness that existed in the original price time series and a certain pattern specific to the log-return time series emerges.

We test the randomness of such log-return series by using the two pseudo-random generators, LCG and MT, and examine how the randomness is retained in the process of converting the original time series to the log-return sequence. By generating the series to make N=500 and T=1500 (Q=3) and execute the process of the RMT-test in Section 2, we show the result of moment analysis in Table 5

Table 5 Errors in RMT-test for the overlapping(left) and non-overlapping(right) log-return sequences

	Overlapping		Non-overlapping	
k	LCG	MT	LCG	MT
2	.1268	.1230	-.0013	-.0016
3	.3211	.3089	-.0039	-.0042
4	.5692	.5434	-.0072	-.0076
5	.8732	.8282	-.0112	-.0115
6	1.2416	1.1702	-.0156	-.0159

(left). We also point out that this effect can be eliminated if we take the non-overlapping log-return by giving up the half of the total elements of r_i (i=even or odd), in exchange of the length of data T to one half of the original. The result of moment analysis is given in Table 5 (right).

5 Comparing with NIST Randomness Test

Randomness test of NIST [16] (National Institute of Standards and Technology) is usually used to judge the randomness of sequences. The NIST Test Suite is a statistical package consisting of 15 tests that were developed to test the randomness of (arbitrarily long) binary sequences produced by either hardware or software based cryptographic random or pseudorandom number generators. Some tests are decomposable into a variety of subtests. The 15 tests are: Frequency

Table 6 Result of log-return sequences test by NIST Randomness test

1	Frequency (Monobit) Test	○
2	Frequency Test within a Block	○
3	Runs Test	×
4	Test for the Longest Run of Ones in a Block	×
5	Binary Matrix Rank Test	○
6	Discrete Fourier Transform (Spectral) Test	×
7	Non-overlapping Template Matching Test	×
8	Overlapping Template Matching Test	×
2	Maurer’s “Universal Statistical” Test	×
3	Linear Complexity Test	○
4	Serial Test	×
5	Approximate Entropy Test	×
6	Cumulative Sums (Cusum) Test	○
7	Random Excursions Test	×
8	Random Excursions Variant Test	○

(Monobit) Test, Frequency Test within a Block, Runs Test, Tests for the Longest-Run-of-Ones in a Block, Binary Matrix Rank Test, Discrete Fourier Transform (Spectral) Test, Non-overlapping Template Matching Test, Overlapping Template Matching Test, Maurer's "Universal Statistical" Test, Linear Complexity Test, Serial Test, Approximate Entropy Test, Cumulative Sums Test, Random Excursions Test, and Random Excursions Variant Test.

Here we check the result with the NIST randomness test, that the two pseudo-random numbers and three physical random numbers, have passed the NIST random test (15/15), but as shown in Table 6 less than 50% tests has been passed (6/15), the log-return sequences failed the NIST randomness test. However the initial part of LCG passed the NIST randomness test, we can say the criterion of the RMT-test is stricter than NIST randomness test.

6 Criterion of the Quantitative RMT-Test

Since all the five examples, LCG, MT, and the three physical random generators, that we used in this article pass the randomness test of NIST, we confirm that they are good random generators. From this fact, we determine the upper bound of the Error for the quantitative RMT test. As shown in Table 2 and Table 3 for N=500, the average values of errors of the five examples with 100 independent sequences are all smaller than 0.4%. However, because of large standard deviations, the maximum value of the Error for the 6-th moment reaches as large as 2.65% for N=500. Thus we are forced to allow the Error of the size 3% for the randomness test. By choosing an optimum upper bound ε=0.05 as the criterion, we can make all the five samples pass the RMT-test. On the other hand, the off-random example such as the initial part of LCG with the error slightly over 5% as in Table 4(left) can be immediately observed visually, as in Fig.4(left) in Ref. [9], the size of the Error to discriminate randomness must be between 3% and 5%. By choosing the optimum value for the upper bound of the Error to be 5%, i.e., ε =0.05 for N=500, we can set the criterion of the quantitative RMT-test. In order to see whether the criterion of ε = 0.05 is valid for much lower N, we examine the error levels of the 6-th moments at N=200 and N=100. As shown in Fig. 2 and Fig. 3, the ε=0.05 is not valid for N< 200. Thus we conclude the minimum value of the parameter N to be N$\geq$200 for our quantitative test of RMT.

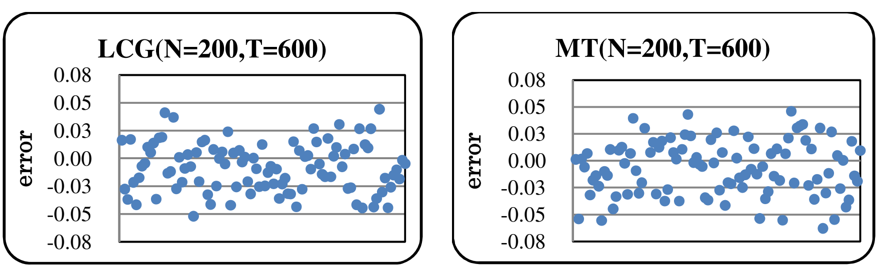


Fig. 2 More than 90% samples pass the RMT-test if ε=0.05 for N=200

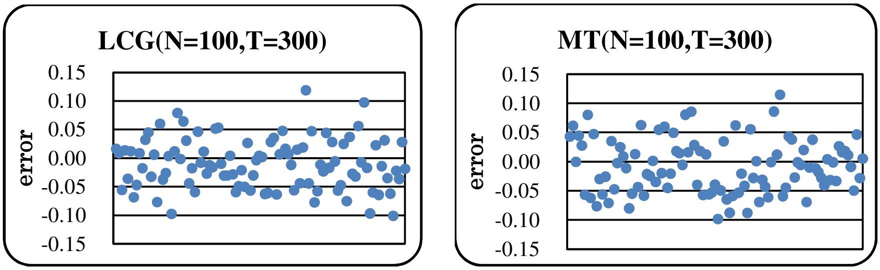


Fig. 3 More than 90% samples pass the RMT-test if ε=0.10 for N=100

7 Summary

In this paper, we have improved the RMT-test by using moment analysis. In order to examine its effectiveness, we tested it on two pseudo-random number generators, LCG and MT, and three physical random numbers made by Toshiba, Hitachi and Tokyo-Electron. Based on those experiments, we determine the criterion of the quantitative RMT-test, i.e., the upper limit of the Error ε to be 5% for N>200. As a result, the sequences indistinguishable by the qualitative RMT-test [9], are also indistinguishable in quantitative RMT-test.

Finally, we mention advantages and disadvantages of our RMT test. Compared to other conventional methods of testing randomness, the RMT-test can be applied on wide range of numerical data, independent of its data format or types. Moreover, the result is visually presented in a graph that can be grasped intuitively. It is particularly suitable to test the randomness of very long, massive data sequences. No null hypothesis, or other complicated process is required. On the other hand, the method requires a very long data sequence.

References

1. Mehta, M.: Random Matrices, 3rd edn. Academic Press (2004)
2. Edelman, A., Rao, N.R.: Random Matrix Theory. Acta Numerica, 1–65 (2005)
3. Wigner, E.P.: Ann. Math. 67, 325–327 (1958)
4. Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., Stanley, H.E.: Physical Review Letters 83, 1471–1474 (1999)
5. Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L., Stanley, H.E.: Random Matrix Approach to Cross Correlation in Financial Data. Physical Review E 65(066126) (2002)
6. Laloux, L., Cizeaux, P., Bouchaud, J., Potters, M.: Noise Dressing of Financial Correlation Matrices. Physical Review Letters 83, 1467–1470 (1998)
7. Tanaka-Yamawaki, M.: Cross Correlation of Intra-day Stock Prices in Comparison to Random Matrix Theory. Intelligent Information Management 3, 65–70 (2011)

8. Tanaka-Yamawaki, M., Kido, T., Itoi, R.: Trend-Extraction of Stock Prices in the American Market by Means of RMT-PCA. In: Watada, J., Phillips-Wren, G., Jain, L.C., Howlett, R.J. (eds.) Intelligent Decision Technologies. SIST, vol. 10, pp. 637–646. Springer, Heidelberg (2011), doi:10.1007/973-642-22194-1
9. Yang, X., Itoi, R., Tanaka-Yamawaki, M.: Testing Randomness by Means of RMT Formula. In: Watada, J., Phillips-Wren, G., Jain, L.C., Howlett, R.J., et al. (eds.) Intelligent Decision Technologies. SIST, vol. 10, pp. 589–596. Springer, Heidelberg (2011), doi:10.1007/973-642-22194-1
10. Knuth, D.E.: The Art of Computer Programming. Seminumerical Algorithms, vol. 2. Addison-Wesley (1980)
11. Matsumoto, M., Nishimura, T.: Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudorandom Number Generator. ACM Trans. Modeling and Computer Simulation 8, 3–30 (1998)
12. Marcenko, V., Pastur, L.: Distribution of Eigenvalues for Some Sets of Random Matrices. Mathematics of the USSR-Sbornik 1, 457–483 (1994)
13. Sengupta, A., Mitra, P.: Distribution of Singular Values for Some Random Matrices. Physical Review E 60, 3389–3392 (1999)
14. Random Number Library: `http://random.ism.ac.jp/random`
15. Yang, X., Itoi, R., Tanaka-Yamawaki, M.: Testing Randomness by Means of Random Matrix Theory. Accepted for Progress of Theoretical Physics (2012) (supplement)
16. NIST: `http://csrc.nist.gov/groups/ST/toolkit/rng/documentation_software.html`

Network Security Evaluation Method via Attack Graphs and Fuzzy Cognitive Maps

Aodah Diamah, Masoud Mohammadian, and Bala M. Balachandran

Abstract. When presented with an attack graph, network administrator may raise question on how to harden the network. To defend his network, network administrator should be supplied with list of all attack paths that can compromise the network. With this list, he can decide which paths are worth paying attention to and defending against. In the event of limited resources, network administrator may only be interested in certain critical paths which cause worst network attack. Attack graph alone is not always helpful on its own and needs additional work for this purpose. In this paper we present the use of a Fuzzy Cognitive Map which is converted from attack graph with genetic algorithm to find attack scenarios causing worst impact on network security. The identified scenarios can then help network administrator to mitigate risks associated with the attack scenarios and improve his network security.

Keywords: Network security, attack graph, FCM, Genetic Algorithm.

1 Introduction

Vulnerabilities are inevitable in any computer network. There are a few factors contributing to this: First, networks generally use commercial software which comes with their vulnerabilities. In most cases, there is a window of time between the discovery of vulnerabilities and the release of patches to fix them, allowing room for an attacker to take action on the vulnerabilities [1]. Second, a network may seem to be secure with individual vulnerabilities being managed by security services. However the effect of these isolated vulnerabilities exploited together by an attacker to penetrate into a network is often overlooked [2]. In most cases rarely network attacks are done with one single step. They are generally composed of multi-step attack with each step exploiting an individual vulnerability to advance

Aodah Diamah · Masoud Mohammadian · Bala M. Balachandran
University of Canberra
Faculty of Information Sciences and Engineering
Canberra, ACT 2606, Australia
* Business Planning Associates Pty Ltd.
Canberra, ACT 2606, Australia

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 433–440.
springerlink.com

deeper into the network [1][3]. Hence, no network is free from attacker. While passive measures such as intrusion detection system and firewalls have been used widely for network defence, researchers are in the opinion that passive mechanism alone is not adequate and incorporating proactive defence in network security management can provide better and effective security governance [1]. Such proactive defence is manifested in finding possible attacks in computer network before they become incident [4] and hardening the network [5][6].

Attack graphs have been used as a tool to analyse network vulnerabilities since they can show paths that might be taken by an attacker to penetrate into a network. However attack graphs do not always useful on their own. Although attack graphs shows paths of exploits, network administrator care more of how to harden their networks [7]. When a network grows, the attack graph may become too complex for network administrator to understand how it helps him to prevent attackers from reaching their goals [5].

This paper presents the use of attack graph to help network administrator identify worst case scenarios and which vulnerabilities of his network that urgently need to be resolved. Worst case scenario happens when an attacker has minimum effort in launching attack but obtain maximum access to restricted privileges within the network. In our approach, we employ an attack graph that has been converted into a Fuzzy Cognitive Map (FCM) and Genetic Algorithm (GA) to search for minimum set of attacks that allow achieving maximum goals.

The paper is organised as follows: First we discuss related past works then we discuss a case study on analysing different attack scenario where we would like to find minimum set of attack that result in greatest probability of accessing restricted privileges. Finally, result, conclusion and future work are presented.

2 Related Work

Attack graphs are used for network security analysis since they can help determine if an attacker can reach his goal of accessing restricted network resources from a starting location. Attack graphs consist of nodes and edges which represent attacker's actions (normally exploiting vulnerabilities) and consequences of the actions, respectively [8]. In [9], minimization analysis was done using attack graph to quantify the likelihood of an attacker to succeed by evaluating every subset of attack paths. If a set of exploits are available to attacker, checking every possible subset of attacks is exponential in the number of attack. The proposed method is able to find minimum critical sets of attacks (MCSA) that must be removed to fail attacker reaching his goal in polynomial time. If the size of the attack graph is m and the number of attacks is n, maximum time to find MCSA is $O(mn)$. Noel, et al in [7] employed attack graphs to compute attack paths that guarantee network security with minimum cost to fix. They started by building graph of dependencies among exploits and conditions. Once this graph is done, they did breadth-first backward tracing from final attack goals to find minimum hardening options of exploits. Our approach is different since we will demonstrate finding minimal sets of attacks by using genetic algorithm. The closest method to this approach is by Danforth [6]. Danforth demonstrated a method to find minimal set of critical patches that maximizes network security and minimize cost of deploying the patches. Danforth used two multi-objective genetic algorithms, NSGA-II and priority algorithm to test their approach. The search starts from initial nodes with available patches and proceed into the network to see if the patches being applied affect the final goal. Combinations of patches are treated as chromosomes and their fitness is evaluated based on the efficacy to disable final goal to be attained by an attacker and on the associated cost of applying the patches. Our search will focus on minimum combination of attacks that cause greatest impact to network (i.e. reaching critical resources). We will use attack graph that has been converted to a Fuzzy Cognitive Map. Fuzzy Cognitive Maps are fuzzy graph structures that represent causal reasoning [10]. Fuzzy Cognitive Maps comprise of nodes which represent concepts and directed edges with weight attached to it to represent causal influence between concepts as Figure 1 shows.

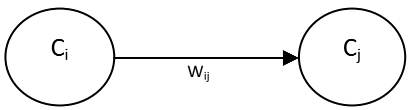


Fig. 1 Ci causes Cj with a degree of Wij

To calculate effect of concepts, an FCM is represented with adjacency matrix. An adjacency matrix indicates the effect of cause variables (row) on the effect variables (column) by taking the weight of FCM edges as the matrix entries [11]. An FCM with n concepts, $C_1, C_2, \ldots, C_n$ will have an nxn matrix. Entry (i,j) denotes the weight of the edge connecting C_h and $C_j(W_{ij})$. Assessing impact of each concept is done by multiplying initial vector with adjacency matrix. Initial vector represents which concepts are activated (their value is set to 1). The result of the multiplication, the new state vector, is tresholded within a defined range and after each multiplication and the activated concepts in new state vector are reset back to 1. This process is repeated to let FCM interact freely one of these conditions is met [12]: (1) equilibrium, a state where new state vector is equal to a previous vector; (2) limit cycle, where state vector value is repeating over the same values in a given time period and (3) chaotic state where the state vector values varies non deterministically.

3 Network Security Evaluation through FCM and Genetic Algorithm

This section demonstrate the use of FCM to evaluate the security of a network. Given a number of network vulnerabilities that directly exposed to attacker (attack surface), which ones when exploited will have the worst effect on network security?

To perform the evaluation, FCM is obtained from converting an attack graph. First, nodes and edges in attack graph become concept nodes and edges in FCM. Second, weight assignment to FCM edges is done by carefully consider impacts of one concepts to another [13]. Finally, an adjacency matrix storing all weight values is produced.

Figure 2 shows the FCM in [13] converted from an attack graph in [14] used in this paper. Attack surface concepts (C1, C2, C3, C4 and C5) are represented as chromosomes of binary 1, if activated and 0 if not activated. This is a simple example with only 2^5 combinations and the effect of all the combinations can easily be simulated. However, this example demonstrates assessing different concepts activation with FCM and search for which minimum combination of vulnerability concepts being obtained by the attacker that can result in worst impact on the network with genetic algorithm, in this case, access to global and local data. The fitness function is formulated as the probability of accessing global and local data (i.e., C11 and C12) divided by the number of concepts activated.

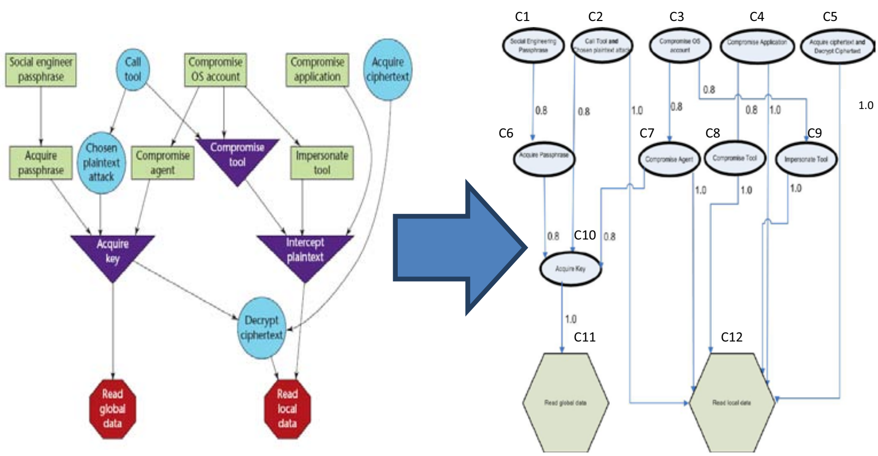


Fig. 2 An FCM [13] converted from an attack graph [14]

The adjacency matrix for the FCM in Figure 2 is as following:

$$E = \left[\begin{array}{ccccccccccccc}
0 & 0 & 0 & 0 & 0 & 0.8 & 0 & 0 & 0 & 0 & 0 & 0 & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0 & 1 & \\
0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0 & 0.8 & 0 & 0 & 0 & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0 & 0 & 0 & 1 & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]$$

4 Simulation

The simulation was run with following steps:

1. Generates initial chromosomes of C1C2C3C4C5 randomly, set crossover probability, mutation probability, crossover point and number of generations. Cross over and mutation probabilities are set to 0.8 and 0.1, respectively. Chromosomes are checked to make sure none of them are identical. If this occurs, another chromosome will be randomly generated.
2. Express the chromosomes as input vector I_{t0}=[C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12] and feed the vector into the FCM. The resulting output vector, that is the new state vector I_{t1}, is thresholded within the range of [0,1] since this represents minimum and maximum probability of an event happening. Continue FCM interaction until output vector reaches limiting value. Evaluate the output vector fitness.
3. Generate random number, if it's smaller than crossover probability; create two new chromosomes by crossing over chromosomes with two best fitness values. Generate random number, if the number is smaller than mutation probability then mutate one bit of the chromosomes
4. Replace the worst two chromosomes with the newly generated one (if none generated they remain in the pool).
5. Repeat step 2-4 until the maximum number of generations is reached.

Worst case scenario that we searched is the ones with maximum fitness value that (theoretically) is 2. This value is derived from maximum number of obtained goals (2) divided by minimum number of activated nodes (1). The program was run each time for 50 generations and the best chromosome yielded fitness value of 1.8.

We take an example of our case study. The chromosomes of first generation randomly generated are: 00010, 11000, 00100, 10011, 00101, and 10101. Each of these chromosomes is first passed to the FCM to calculate their impact on the system. The chromosomes are expressed in a vector I_{to}=[C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12] and then multiplied with FCM adjacency matrix to produce the next state vector I_{t1}. As an example, impact of chromosomes 10011 is expressed as vector I_{t_0}=[100110000000] can be calculated as follows:

$$I_{t_1} = I_{t_0} * E = [100110000000]\begin{bmatrix} 0\ 0\ 0\ 0\ 0\ \ 0.8\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 0\ 0\ 0\ 0\ \ 0\ \ 0\ 0\ 0\ 0.8\ \ 0\ \ 1 \\ 0\ 0\ 0\ 0\ 0\ 0\ 0.8\ 0\ 0.8\ 0\ 0\ 0 \\ 0\ \ 0\ \ 0\ 0\ 0\ 0\ 0\ 0.8\ 0\ 0\ \ 0\ 1 \\ 0\ \ 0\ \ 0\ 0\ \ 0\ \ 0\ 0\ 0\ 0\ 0\ 0\ 0.1 \\ 0\ \ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0.8\ 0\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0.8\ 0\ 1 \\ 0\ \ 0\ \ 0\ \ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1 \\ 0\ \ 0\ \ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1 \\ 0\ \ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0 \\ 0\ \ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ \ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \end{bmatrix} = [0\ 0\ 0\ \ 0\ \ 0\ \ 0.8\ 0\ \ 0.8\ \ 0\ \ \ 0\ \ 0\ \ \ 2]$$

The new vector I_{t_1} is tresholded within the range [0, 1], hence it becomes [0 0 0 0 0 0.8 0 0.8 0 0 0 1]. I_{t_1} tells us activation of C1, C4 and C5 causes 100% probability of C12 to happen. For the next iteration, I_{t_1} now becomes the input to the FCM by keeping C1, C4 and C5 activated, i.e. changing C1, C4 and C5 to 1. This process continues until the output vector either reaches equilibrium or limit cycle. After 3rd iteration, new state vector remains the same as previous state vector which means equilibrium has been reached. Hence we stop the process with new state vector I_{t_3}=[0 0 0 0 0 0.8 0 0.8 0 0.64 0.64 2.8]. Again tresholding the value we will have I_{t_3} as [1 0 0 1 1 0.8 0 0.8 0 0.64 0.64 1]. I_{t_3} can be interpreted as: if C1, C4 and C5 happen, (social engineering pass-phrase, call tool compromise application and chosen plain text attack) there is 64% probability of C11 happening and 100% probability of C12 happening, which means attacker can read local data.

Once we have I_{t_3}, we evaluate its fitness value. We simplify fitness function by summing up the probability of C11 and C12 happening divided by the number of activated concepts. In this case fitness value is 1.64/3=0.5467. Fitness values for the rest of chromosomes are listed in Table 1.

Table 1 Chromosomes of first generation

Chromo-somes	Number of Concept activated	I_{ti}	Probability of goal reached		Fitness value (probability of goals reached/number of concept activated)
			C11	C12	
00010	1	0 0 0 1 0 0 0 0.8 0 0 0 1	0	1	1
11000	2	1 1 0 0 0 0.8 0 0 0 1 1 1	1	1	1
00100	1	0 0 1 0 0 0 0.8 0 0.8 0.64 0.64 1	0.64	1	1.64
10011	3	1 0 0 1 1 0.8 0 0.8 0 0.64 0.64 1	0.64	1	0.5467
00101	2	0 0 1 0 1 0 0.8 0 0.8 0.64 0.64 1	0.64	1	0.82
10101	3	1 0 1 0 1 0.8 0.8 0 0.8 1 1 1	1	1	0.667

We run several simulations and plot best fitness value of each generation as the simulation run, as seen in Figure 3.

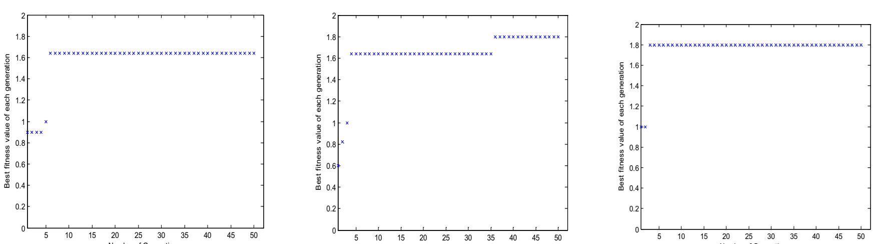


Fig. 3 Best fitness value of each generation of three simulations

From three simulations of 50 generations, two best fitness values were obtained, 1.8 and 1.64. Figure 3 (left) shows best fitness value is 1.64 where two other (middle and right) have best fitness value of 1.8. These values come from chromosomes as listed in Table 2.

Table 2 Chromosomes with best fitness values

Chromosomes	Number of Concepts activated	Probability of goals reached		Fitness value (probability of goals reached/number of concept activated)
		C11	C12	
01000	1	0.8	1	1.8
00100	1	0.64	1	1.64

The chromosomes suggest that C2 alone can result in 80% and 100% probability of C11 and C12 happening, respectively and C3 alone can result in 64% probability of C11 and 100% probability of C12 taking place. Hardening network can be done taking measures to mitigate or minimize risks associated with these chromosomes, i.e., risk C2: call tool and chosen plaintext attack, and C3: compromise OS tool.

5 Conclusion and Future Work

This paper has presented the use of attack graph together with FCM and GA to search for worst case scenario in a network. It demonstrated how different attack possibilities in an FCM can be presented with simple vectors and how their impacts can be evaluated by passing the vectors to an FCM matrix. GA helped creating different attack scenarios and filter which scenarios remain in the pool and which will be discarded. In this paper fitness function is simplified by assuming cost of launching different attacks is equal. Improvement can be done for future work by assigning different cost to different vulnerabilities when formulating fitness function. Different vulnerabilities may have different difficulties of exploiting them and final goals may as well have different importance. For example acquiring and decrypting cipher text has different difficulties to launching plaintext attack, or global data is far guarded against compared to local data.

References

[1] Franqueira, V.N.L., Lopes, R.H.C., van Pascal, E.C.K.: Multi-step Attack Modelling and Simulation (MsAMS) Framework based on Mobile Ambients. In: The 2009 ACM Symposium on Applied Computing, pp. 66–73. ACM, New York (2009)

[2] Ritchey, R.W., Ammann, P.: Using Model Checking to Analyze Network Vulnerabilities. In: IEEE Security and Privacy Proceedings, Berkeley, CA, USA, pp. 156–165 (2000)

[3] Liu, Z., Wang, C., Chen, S.: Correlating Multi-Step Attack and Constructing Attack Scenarios Based on Attack Pattern Modeling, pp. 214–219. IEEE Computer Society (2008)
[4] Franqueira, V.N.L.: Finding Multi-Step Attacks in Computer Network using Heuristic Search and Mobile Ambient. s.l.: CTIT Ph.D. Thesis Series No. 09-154 (2009)
[5] Danforth, M.: Models for Threat Assessment in Networks, PhD Thesis. s.l.: University of California-Davis, Computer Science Department (2006)
[6] Danforth, M.: Scalable Patch Management using Evolutionary Analysis of Attack Graphs. In: Seventh International Conference on Machine Learning and Applications, IEEE Computer Society (2008)
[7] Noel, S., et al.: Efficient Minimum-Cost Hardening via Exploit Dependency Graphs. s.l. In: IEEE 19th Annual Computer Security Applications Conference, pp. 86–95 (2003)
[8] Lippmann, R.P., Ingols, K.W.: An Annotated Review of Past Papers on Attack Graphs. s.l. MIT Lincoln Laboratory (2005)
[9] Oleg, S., et al.: Automated Generation and Analysis of Attack Graph. s.l. In: 2002 IEEE Symposium on Security and Privacy, IEEE (2002)
[10] Kosko, B.: Fuzzy Cognitive Maps. International Journal of Man-Machine Studies, 65–75 (1986)
[11] Ross, T.J.: Fuzzy Logic with Engineering Applications. s.l. John Wiley and Sons (2010)
[12] Andreou, A.S., Mateou, N.H.: Soft Computing for Crisis Management and Political Decision Making: The use of Genetically Evolved Fuzzy Cognitive Maps. Soft Computing 9, 194–210 (2004)
[13] Mohammadian, M.: Intelligent Risk Identification and Analysis in IT Network Systems. In: Iliadis, L., Maglogiannis, I., Papadopoulos, H. (eds.) Artificial Intelligence Applications and Innovations. IFIP Advances in Information and Communication Technology, vol. 364, pp. 370–377. Springer, Heidelberg (2011)
[14] Gupta, S., Winstead, J.: Using Attack Graph to Design Systems. IEEE Security and Privacy, 80–83 (2007)

On a Patent Analysis Method for Identifying Core Technologies

Chulhyun Kim and Hyeonju Seol

Abstract. This study proposes a new approach to identifying core technologies from the perspectives of co-occurrence, relatedness, and cross-impact based on patent co-classification information with consideration of the overall interrelationships among technologies. First, association rule mining is employed to derive the co-occurrence, relatedness, and cross-impact indexes and three technological matrixes are constructed. Second, the analytic network process is conducted to produce importance values of technologies for three perspectives with consideration of their direct and indirect impacts. Finally, data envelopment analysis is applied to identify priorities of technologies. The proposed approach can be utilized for technology monitoring for both technology planning of firms and innovation policy making of governments.

Keywords: Association rule mining (ARM), Analytic network process (ANP), Data Envelopment Analysis (DEA), Core technology, Patent co-classification.

1 Introduction

It has become more important to grasp technological trend and advances by analyzing the overall structure of technologies and interaction among them (Lee et al. 2009a). There have often been attempts to identify technological structure and relationship and they are mainly conducted through the patent analysis (Trajtenberg 1990). The most commonly used information of patents for analyzing technological relationship is citation information. However, they have some shortcomings

---

Chulhyun Kim
Induk University, Choansangil 14, Nowon-Gu, Seoul, 139-749, South Korea
Tel.: +82-2-950-7608; Fax: +82-2-950-7608
e-mail: stddevs@induk.ac.kr

Hyeonju Seol
Korea Air Force Academy, 335-1, Ssangsu-ri, Namil-myeon, Cheongwon-gun, Chungbuk, 363-849, South Korea
Tel: +82-43-290-6492; Fax: +82-43-297-0479
e-mail: hjseol@afa.ac.kr

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 441–448.
springerlink.com

reported in the literature. First, the average time-lag between citing-cited patents is over 10 years (Hall et al. 2001). Moreover, since citation analysis considers citing-cited relationship between individual patents, it is difficult to identify technological relatedness and characteristics from the perspective of technological fields (Yoon and Park 2004).

On the contrary, the patent analysis with co-classification information has some advantages over citation information. Co-classification analysis is to analyze technological relationship based on the fact that patents are classified to some technological classes considering their technological characteristics (OECD 1994). In contrast to citation analysis, it is based on the hierarchical technological classification system so that technological relationship can be analyzed not on the level of individual patents but on the various technological levels according to the purpose of studies. Furthermore, errors from time-lag are relatively less since the time of classification information of a patent is equal - patent registration time.

Among the various index using the information of patent co-classification, the number of co-occurrences (Englesman and Van Rann 1992), cosine index (Salton and MacGill 1983) and technological cross-impact index (Choi et al. 2007) has been used to identify core technologies and interrelationships between technologies. Most of researches on patent co-classification have, however, only focused on one of the three patent co-classification indexes, not on combining them.

In response, this paper proposes a new approach to identifying core technologies based on patent co-classification information with consideration of the overall interrelationships among technologies considering three perspectives mentioned above. The proposed approach is comprised of three methods: association rule mining (ARM), analytic network process (ANP) and data envelopment analysis (DEA). At first, association rule mining (ARM) is employed to calculate three indexes. Since the support measure in ARM is defined as a probability of the co-occurrence of two technologies, it is adopted as the index of the number of co-occurrences. The lift measure in ARM has the same formula with cosine index and it is used as the index of evaluating technological relatedness. Also, the confidence measure in ARM is of the same formula with cross-impact index, and it is employed as the index of evaluating technological cross-impacts. Then, the technological co-occurrence matrix, the technological relatedness matrix and the technological cross-impact matrix are constructed with all calculated indexes. Secondly, the ANP, which is a generalization of the analytic hierarchy process (AHP), is employed to calculate the importance of technologies based on the three matrixes derived previous step. Since the ANP is capable of measuring the relative importance that captures all the indirect interactions in a network, the derived limit centrality indicate the importance of a technology in terms of co-occurrence with, relatedness with and impact on other technologies, taking all the direct and indirect influences into account (Lee et al. 2009a). Finally, DEA is applied to identify the priority of technologies based on three criteria. Technologies (classes) correspond to DMUs (decision making units) and the three calculated importance values match to outputs. Also, efficiency scores mean the priority scores of each technology. Therefore, core technologies considering co-occurrence, relatedness, and cross-impact together can be identified by applying DEA.

The remainder of the paper is organized as follows. Section 2 deals with methodological background including ARM, ANP and DEA. The proposed approach is explained in Section 3. The paper ends with conclusions in Section 4.

2 Methodological Background

2.1 Association Rule Mining (ARM)

ARM is one of the data mining techniques to search for interesting relationships among items in large database. Association rule stands for the co-occurrence of two items and indicates that if two items occurs together frequently they have strong association relationship (Han and Kamber 2001). The three measures of evaluating the rule interestingness are support, confidence, and lift. Their descriptions brief are shown in Table 1. The typical procedure of ARM consists of two steps (Agrawal et al. 1993): (1) Search for frequent itemsets – To create all item combinations over the threshold value of support (2) Generate association rules – To Select itemsets over the threshold value of confidence or lift among the frequent itemsets found in (1). The step (1) is a very time consuming job and the most representative technique for this is Apriori algorithm (Agrawal and Srikant 1994).

Table 1 Measures of interestingness

Measure	Description	Formula
Support	The usefulness of discovered rule A→B	$P(A \cap B)$
Confidence	The certainty of discovered rule A→B	$P(B \mid A)$
Lift	The correlation between the occurrence of items in discovered rule A→B.	$\frac{P(B \mid A)}{P(B)}$

2.2 Analytic Network Process (ANP)

The ANP is a generalization of the AHP (Saaty 1996). The ANP extends the AHP to problems with dependence and feedback. It allows for more complex interrelationships among decision elements by replacing the hierarchy in the AHP with a network. Thus, the ANP produces priorities or relative importance of elements in a complex network model with consideration of interdependency among elements. The process of ANP is composed of four steps (Saaty 1996; Lee et al. 2009): (1) Network model construction, (2) Pairwise comparison and local priority vectors, (3) Supermatrix formation and transformation, and (4) Final priorities.

2.3 Data Envelopment Analysis (DEA)

DEA is a non-parametric approach for evaluating relative efficiency of DMUs with multiple inputs and outputs with others (Charnes et al. 1978). Efficiency in DEA is defined to a ratio of the weighted sum of inputs to the weighted sum of outputs and the value of a DMU's efficiency is restricted to be less or equal than 1. Under this restriction, efficiency is measured based on the weight of each element maximizing the efficiency of target DMUs. Furthermore, DEA is a useful tool for benchmarking by providing the information of improving efficiency through reference set for inefficient DMUs.

The first DEA model proposed by Charnes et al. (1978) is the CCR model that assumes that production exhibits constant returns to scale. Banker et al. (1984) extended it to the BCC model for the case of variable returns to scale. DEA models are also distinguished by the objective of a model: maximize outputs (output-oriented) or minimize inputs (input-oriented) (Lee et al. 2009b).

3 Proposed Approach

The overall procedure of the proposed approach is as follows. First, patent data of interesting technological area is collected. Second, technological co-occurrence matrix, technological relatedness matrix, and technological cross-impact matrix are constructed with the support, lift, and confidence values calculated by applying ARM to the co-classification information of gathered patent data. Third, importance of technologies from the perspectives of each criterion is calculated by applying ANP to three technological matrixes. Finally, core technologies are identified through employing DEA to the extracted technological importance. Figure 1 depicts the overall process of the proposed approach. More detailed explanations are provided below.

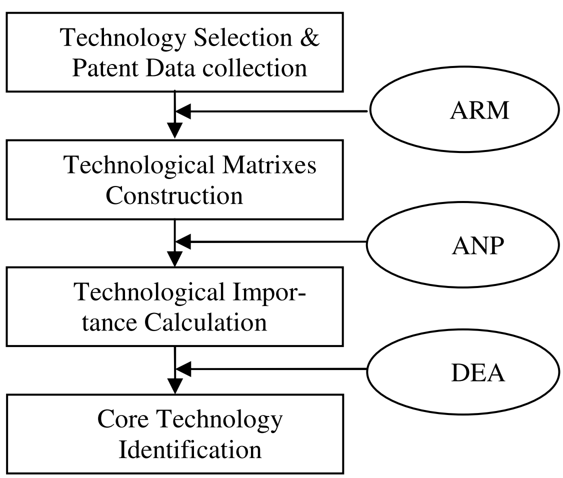


Fig. 1 Overall research framework

3.1 Technology Selection and Patent Data Collection

First of all, the technological area to be analyzed should be decided. This research adopts the patent classification system for classifying technologies. Patent classification system stands for the hierarchical system to classify and manage patents considering their technological characteristics. Patents have more than two claims for diversifying subjects and they are affiliated to more than two classes based on the patent classification system (USPTO 2006). Class, therefore, indicates which technological areas the patents (individual technologies) are affiliated in technological classification systems.

The data source used in this study is patents registered in the United States Patent and Trademark Office (USPTO). The USPTO has classified granted patents into corresponding technology classes defined by the USPC (United States Patent Classification). Each subject matter division in the USPC includes a major component called a class and a minor component called a subclass (USPTO 2006). A class generally distinguishes one technology from another and consists of subclasses that delineate processes, structural features, and functional features of the subject matter encompassed within the scope of a class.

3.2 Construction of Technological Matrixes with ARM

The co-occurrence index is defined as P(A∩B), which is the same to the support of an association rule A→B in ARM. The cosine index which explains the relatedness between two technologies is defined as a form of correlation, P(A∩B)/P(A)P(B), which is the formula of the life of an association rule A→B in ARM. The cross-impact index is defined as the conditional probability P(B|A), which is of the same formula with the confidence of an association rule A→B in ARM. Accordingly, this study applies ARM to the co-classification information of gathered patents to construct the co-occurrence, the relatedness, and the cross-impact matrix. Table 2 represents the form of three matrixes with the support, lift, and confidence value between two technological areas. T_i means the ith technological area (class). Supp($T_i \rightarrow T_j$), Lift($T_i \rightarrow T_j$), and Conf($T_i \rightarrow T_j$) indicate the support, lift, and confidence values of the association rule $T_i \rightarrow T_j$, respectively.

3.3 Calculation of Technological Importance with the ANP

This study tries to calculate the technological importance from three perspectives using the ANP. The first two steps in the ANP, (1) Network model construction, (2) Pairwise comparison and local priority vectors, are not actually required in the proposed approach. The network in the proposed approach is made on the basis of relationships represented in each technological matrix. A cluster in the ANP network corresponds to a class and each cluster has no elements. In the ANP context, then, the resulting network model only includes alternative clusters. Thus, the importance of alternatives (classes) is only evaluated with respective to relationships with other alternatives. It is assumed that the three indexes between two classes is a proxy of intensity of influences between them. Thus, pairwise comparisons are

Table 2 Form of technological matrixes

(a) Technological co-occurrence matrix

	T_1	T_2	…	T_n
T_1	1	Supp($T_1 \rightarrow T_2$)		Supp($T_1 \rightarrow T_n$)
T_2	Supp($T_2 \rightarrow T_1$)	1		Supp($T_2 \rightarrow T_n$)
…			1	…
T_n	Supp($T_n \rightarrow T_1$)	Supp($T_n \rightarrow T_2$)	…	1

(b) Technological relatedness matrix

	T_1	T_2	…	T_n
T_1	1	Lift($T_1 \rightarrow T_2$)		Lift($T_1 \rightarrow T_n$)
T_2	Lift($T_2 \rightarrow T_1$)	1		Lift($T_2 \rightarrow T_n$)
…			1	…
T_n	Lift($T_n \rightarrow T_1$)	Lift($T_n \rightarrow T_2$)	…	1

(c) Technological cross-impact matrix

	T_1	T_2	…	T_n
T_1	1	Conf($T_1 \rightarrow T_2$)		Conf($T_1 \rightarrow T_n$)
T_2	Conf($T_2 \rightarrow T_1$)	1		Conf ($T_2 \rightarrow T_n$)
…			1	…
T_n	Conf($T_n \rightarrow T_1$)	Conf($T_n \rightarrow T_2$)	…	1

not required, and instead, a set of indexes between a class and the other classes form a local priority vector on the class.

The supermatrix in the ANP is a partitioned matrix composed of all local priority vectors. Since each technological matrix is a set of all local priority vectors, it is equivalent to the supermatrix. The supermatrix is then transformed into the weighted supermatrix each of whose column sums to one. Finally, the weighted supermatrix is transformed into the limit supermatrix by raising itself to powers. The columns of the limit supermatrix converge to the same, which is called limit priorities capturing all of the direct and indirect influences among technologies. Then, each technological importance can be identified based on the limit priorities of technologies.

3.4 Identification of Core Technologies with the DEA

To evaluate the individual technologies (classes) based on three technological importance calculated at previous stage is a problem of MCDM. That is, each technology can be considered as an alternative and three aspects as criteria for evaluation. If DEA is applied for this aim, technologies correspond to DMUs and the three importance values from the perspectives of co-occurrence, relatedness, and cross-impact to outputs. Also, efficiency scores derived by applying DEA mean the priority scores of each technology. Figure 2 depicts the relationship among technological priorities, MCDM, and DEA.

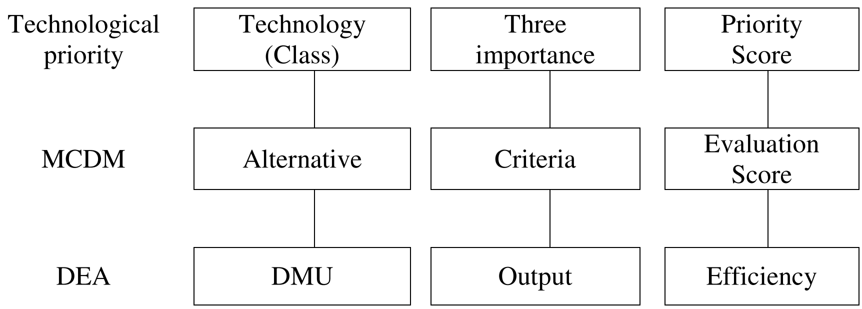


Fig. 2 The relationship among technological priority, MCDM, and DEA

Since only three outputs exist and there is no input, output-oriented BCC model without inputs is adopted for applying DEA. Though R&D investments and labor for R&D may be considered for inputs, this study does not aim to measure the efficiency of investment or labor for R&D. This study only tries to judge the priority scores of technologies and it is not appropriate to reflect on these for inputs. The core technologies considering co-occurrence, relatedness, and cross-impact together can be identified with this process.

5 Conclusions

This study has proposed a systemic approach to identification of core technologies from the perspectives of the technological co-occurrence, relatedness, and cross-impact. ARM was applied to the patent co-classification data, and the three technological matrixes were constructed with the derived support, lift, and confidence value of each technology. Each technological importance value of three perspectives was calculated through the ANP with overall interrelationship among technologies. Then, DEA was employed to prioritize technologies considering three perspectives together.

This research contributes to the field by proposing a new approach to identification of core technologies based on the overall technological relationships. The proposed approach can be utilized for technology monitoring for both technology planning of firms and innovation policy making of governments.

However, since this study only proposes a framework for identifying core technologies, an extension to carrying out detailed case study could be considered as future research issues.

Acknowledgement. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2010-0024709).

References

Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proc. of the ACM SIGMOD Int. Superence on Manag. of Data, pp. 207–216 (1993)

Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proc. of the 20th VLDB Superence, pp. 478–499 (1994)

Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiency in data envelopment analysis. Manag. Sci. 30(9), 1078–1092 (1984)

Charnes, A., Cooper, W.W., Rhodes, E.: Measuring efficiency of decision making units. Eur. J. Oper. Res. 2(6), 429–444 (1978)

Choi, C., Kim, S., Park, Y.: A patent-based cross-impact analysis for quantitative estimation of technology impact: The case of information and communication technology. Technol. Forecast. Soc. Chang. 74(8), 1296–1314 (2007)

Engelsman, E.C., van Rann, A.F.J.: A patent-based cartography of technology. Res. Policy 23(1), 1–26 (1992)

Hall, B., Jaffe, A., Trajtenberg, M.: The NBER patent citations data file: Lessons, insights and methodological tools. National Bureau of Economic Research Working Paper No. 8498 (2001)

Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Academic Press, San Diego (2001)

Lee, H., Kim, C., Cho, H., Park, Y.: An ANP-based technology network for identification of core technologies: A case of telecommunication technologies. Expert Syst. Appl. 36(1), 894–908 (2009a)

Lee, H., Park, Y., Choi, H.: Comparative evaluation of performance of national R&D programs with heterogeneous objectives: A DEA approach. Eur. J. Oper. Res. 196(3), 847–855 (2009b)

OECD, Using patent data as science and technology indicators - patent manual. OECD, Paris (1994)

Saaty, T.: Decision making with dependence and feedback: The analytic network process. RWS Publications, Pittsburgh (1996)

Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill, New York (1983)

Trajtenberg, M.: A penny for your quotes: patent citations and the value of inventions. RAND J. Econ. 21(1), 172–187 (1990)

USPTO, Overview of the US patent classification system (2006), `http://www.uspto.gov`

Yoon, B., Park, Y.: A text-mining based patent network: Analytical tool for high-technology trend. J. High Technol. Manag. Res. 15(1), 37–50 (2004)

Ontology-Based CRIS Design Using Three Stage Development Approach

Nan-Chen Hsieh, Huan-Chao Keh, Chien-Hui Chan, and Szu-Ting Wang

Abstract. In healthcare area, plenty of existing expert knowledge is used to support decision making. Representation of the knowledge becomes an important problem. Especially, how it can be effectively used for reasoning as part of the decision making systems. Ontologies are commonly used for describing information system design or generating and validating system components. In this study we demonstrate a three stage approach to developing cardiovascular surgery ontology and then to making use of use that ontology to design an ontology-based cardiovascular clinical research information system. By using ontology approaches to designing and implementing the system we can convert 'pure data' back to 'reality'.

1 Introduction

Cardiovascular disease is a complex medical problem. Surgeons-made decisions about patients' care are majorly determined by their past experience with similar patients. The interventional methods, procedures, and treatments should be evident for the cardiovascular diseases, and modern medical practice should be recorded as electronic medical records (EMRs). A surgical workflow is defined as a partial to full automation process in the management of patients in different surgical stages: preoperative, intraoperative, postoperative, and perioperative. A large amount of information is transferred from one stage to the next, including patient characteristics, patient history, surgical documents, laboratory tests, surgical information, and surgical procedures[1].

Nan-Chen Hsieh · Szu-Ting Wang
Department of Information Management, National Taipei University of Nursing and Health Sciences, Taiwan, ROC
e-mail: {nchsieh,emily21714}@gmail.com

Huan-Chao Keh · Chien-Hui Chan
Department of Computer Science and Information Engineering, Tamkang University, Taiwan, ROC
e-mail: keh@cs.tku.edu.tw, emmacc@gmail.com

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 449–458.
springerlink.com

A cardiovascular CIS (clinical information system) is helpful prior to and during cardiac surgery[2]. The information collected can be used by surgeons and patients to evaluate whether or not the surgical procedures are likely to be successful. The development of a robust CIS can therefore both assist vascular surgeons in evaluating the expected outcome for a given patient as well as facilitate counseling and operative decision-making[3]. CRIS with EMRs can be used as a clinical passive tool to store patient information, and can include intelligent data analysis functions. With these functions, CRIS can make substantial contributions to supporting the delivery of patient care, effective storage, results reporting, care planning, analysis, and visualization of complex correlations between data pieces [4, 5]. Additionally, CRIS' procedure-specific modules contain intelligent data analysis tool which assists surgeons making decisions and facilitates detail reports with reference value.

In this study, we will demonstrate how an ontology development method applies to build a cardiovascular surgery clinical research information system (CRIS). Figure 1 shows the proposed CIS framework. The framework is described as follows:

- Preoperative stage: all patients start the process of evaluation of their health conditions, such as their family's health history, risk factor, and physiology examination. Certain statistically significant intra-operative parameters which have been estimated using the information from previous patients will be used to predict the surgery outcome.
- Operative stage: the major task of this stage is to perform surgery. At the end of the stage, surgery procedure information and intra-operative findings will be collected.
- Postoperative stage: besides the routine examination for checking patients' postoperative physical condition, in-hospital mortality and 30-day readmission rates will be tracked to evaluate the healthcare quality. This stage could provide additional information to the risk prediction of a future cardiovascular surgery patient. It takes into account past events' outcome and follow-up information of the current patient
- Ontology-based data modeling: It is used in order to describe the basic general clinical concept, the whole processes, and the relationship between semantics of cardiovascular surgery information.

2 The Architecture of Clinical Research Information Systems

The core of a clinical information system(CIS) is the computerized registry[2]. Each patient in the system has his/her individual file folder, and each file folder contains many forms, including a demographic datasheet, surgical datasheets, clinic datasheets and laboratory datasheets. By incorporating the hospital

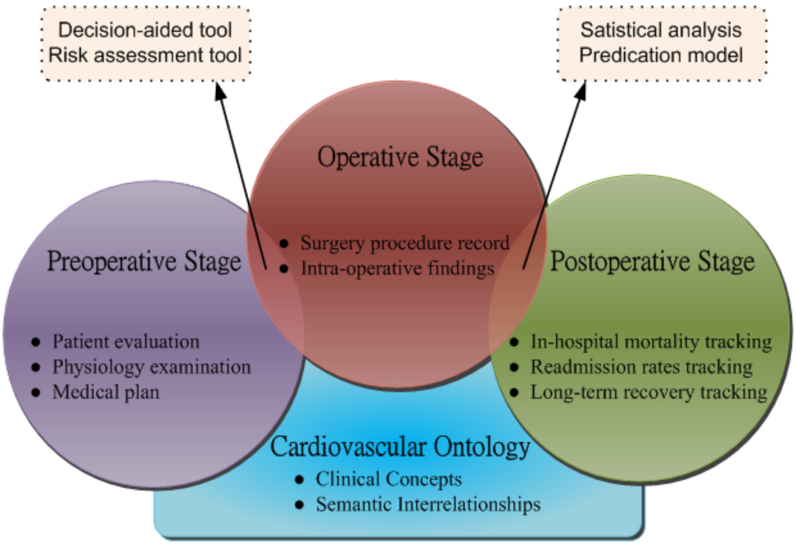


Fig. 1 A conceptual framework for CRIS

guidelines into the registry, all of the follow-up visits are uniform regardless of the specialist. These high-risk cardiovascular patients require close monitoring and careful therapeutic decision-making to avoid complications, which may lengthen their hospital stay, increase the cost of care, and decrease the quality of life. Moreover, the CIS should provide a search function for the registries, extract the most relevant data for the patient, and accumulate and summarize the data using disease-specific or population-based information.

In order to provide a more efficient research environment for clinicians and increase knowledge retention, it is necessary to improve a CIS into CRIS which can acquire the current data accumulated in HIS, LIS, PACS and NIS. Through data analysis, the clinical research information system works providing both statistics and data mining approach to analyze data and retrieve knowledge. For storing and reusing knowledge, a knowledge base needed to be built. The proposed system architecture is shown in Fig. 2.

3 The Ontology-Based System Development Approach

Representation of knowledge becomes an important, especially, when the question is how it can be effectively used for reasoning as part of the decision making systems. Kuziemsky and Lau[6] point out that if the system has been built based on wrong assumptions or poorly articulates user needs then the CIS project will fail, and they suggest ontologies can assist in health information system design by providing a comprehensive model of the information and process needs for healthcare delivery.

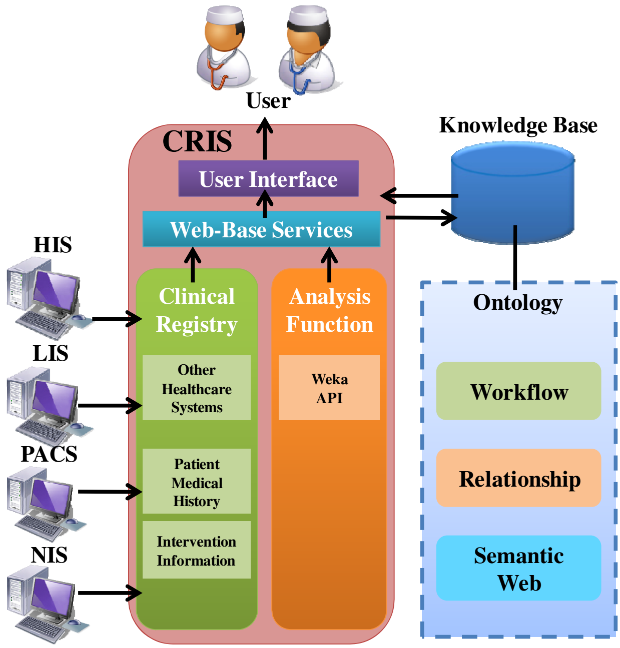
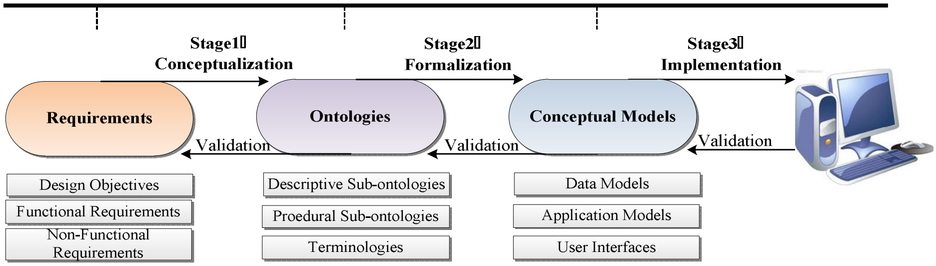


Fig. 2 System architecture for the CRIS

Our cardiovascular ontology is based on the surgical operation stages and other related sources of medical knowledge that include patient information, medical plan, and other medical ontologies. Developing ontology is usually an iterative process. In practical terms, developing ontology includes [7]:

- Defining classes in the ontology
- Arranging the classes in a subclass-superclass hierarchy
- Defining slots and describing allowed values for these slots
- Filling in the values for slots for instances

There are several ontology development methodologies, including On-To-Knowledge and METHONTOLOGY there are several ontology development methodologies existed [8]. Pinto and Martin [9] describe an ontological engineering process that includes: specification, conceptualization, formalization, implementation and maintenance. We modified the stages in order to support ontology based CRIS design. Our methodological approach is illustrated in Fig. 3 and has three stages: conceptualization, formalization and implementation. Each stage is defined by work content, method and output.



	Stage1:Conceptualization	Stage2:Formalization	Stage3:Implementation
Content	Data collection: Special chart Surgical workflow	Problem-solving approaches: provide specific solutions to domain problems Ontology relationship: data definitions and semantics, and find relationship between data Ontology modeling: design of ontology for cardiovascular surgery	Database design: design database conceptual model based on ontologies, and ensure entities are appropriate
Method	Literature review Conducting interview Requirement analysis	Ontology methodology UML methodology	Rapid prototyping WEKA data analysis API
Output	Research scope and design objectives Requirements	Ontology models	CRIS Intelligent data analysis tool

Fig. 3 Three stages approach to design cardiovascular surgery ontologies.

- Stage 1：Conceptualization

The first step in conceptualization requires analysis of the research scope and structure. By means of literature review, interview and field investigation, the medical domain knowledge was collected by professionals who are familiar with medical and information technology. There are two tasks done in this stage: data collection and analysis. Details are described as follows:

— Data collection: To identify ontology scope and purpose there are three goals the data sources need to accomplish. The first is to be able to validate concepts and processes. The second goal is to incorporate conceptual models and other relevant research literature as a means of linking research and practice. The third goal involves the use of historical data, such as retrospective patient cases, to understand current data collection practices and how that data can be formalized into information and knowledge to be returned to end users. We conducted interviews with surgeons and related system users to understand existing problems and requirements. During the data collection process to determine the scope some specific questions must be posed.

— Data analysis: The first goal is ensuring the collected data and clinical workflow must be the same as established models. The second goal is reviewing literature which is relevant to establish models. The third goal is how to represent and establish models of the knowledge base, and feedback to healthcare providers.

- Stage 2：Formalization

The first step in this stage is developing the domain ontology which is a formal model of the concepts and categories. First, the special charts were re-designed as the computerized registry. Each patient has his/her individual registry or file folder, and each registry contains many forms, the registry contains data on patients' preoperative characteristics, risk factors, details of the surgical procedure, physical characteristics of the heart and postoperative physiological and laboratory findings. The goal of this step was to develop clear and concise forms, and the forms must be conforming surgical workflow and acceptable to formal medical records.

- Stage 3：Implementation

There are two main tasks to be done in the implementation stage. The database development task involves the conversion of domain and ontologies into a physical database model, and ensures the appropriateness of the built model. Each workflow in CRIS is tested by key users to ensure no workflow problems have been found until the application module reaches a stable state. This stage implements the ontology as a CRIS and use WEKA API to visualize the result of clinical data analysis.

4 Practical Implementation for the Cardiovascular CRIS

Before the development of CRIS, detailed patient and procedural information is hand-gathered on multipage forms the cardiovascular fellows and then entered into the Excel spreadsheets after each care process is ended.

The establishment of the CRIS is primarily used to collect, store, analyze and interpret information from various sources [10]. Ontology is a specification of a conceptualization that could be separately identified by domain users, and used as a self-contained way to communicate domain information. Ontologies are intended to give details and to explain objects, while conceptual data modeling is used to represent objects and their relationships to some application purpose [11].

The patient care process of cardiovascular surgery is roughly composed of three stages: preoperative, perioperative, and postoperative. The information collected can be used by surgeons and patients to evaluate whether or not the surgical procedures are likely to be successful, and for operative and postoperative death and complications rates.

The cardiovascular special chart is a confidential document that contains detailed and comprehensive information on an individual and the care experience related to the patient. The specific information contained in the chart is intended to provide a record of a patient's clinical condition by detailing diagnoses, treatments, tests and responses to treatment, as well as any other factors that may affect the patient's health or clinical state.

The core task of ontology development is a categorization process. Good categorization can facilitate information retrieval from database system. The development of cardiovascular surgery ontology should include relevant concepts, attributes, constraints and instances. A schematic ontology class hierarchy structure for the cardiovascular surgery ontology is illustrated in Fig.4. This sub-ontology includes all concepts and risk factors that are relevant for patients in the preoperative stage. The intraoperative sub-ontology includes important information for cardiovascular surgical operation. The postoperative sub-ontology includes health conditions after surgery discharge related information for follow-up.

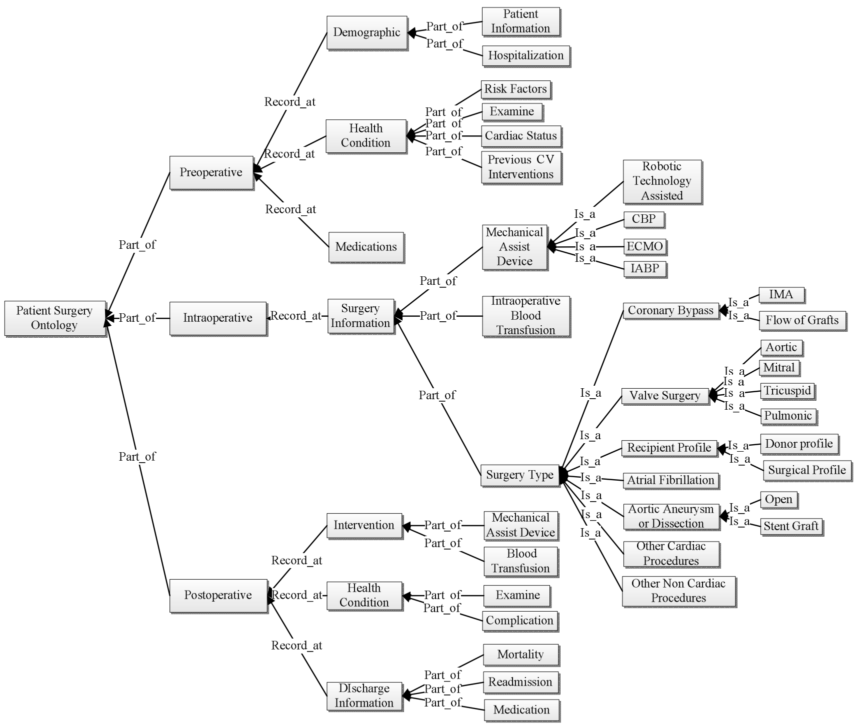


Fig. 4 The cardiovascular surgery ontology class hierarchy in supporting the CRIS

Using ontology approaches, we can easily describe the patient registry shown in Fig.5 as: A 73 year-old female patient with atrial fibrillation was being prepared for surgery. She had hypertension, peripheral vascular disease, and chronic renal insufficiency. Preoperative examination revealed height 149cm, weight 46.4kg. Cardiac sonography and pertinent lab data were as below: LVEDD 49mm, IVS 9mm, RVSP 90mm, creatinine 1.9mg/dl, FEV1 1.03L/sec, FEV1/FVC ratio 78%.

The gathering medical datasets are retrieved not only from the CRIS database but also obtained from other healthcare systems. The development of a robust prediction model in medical datasets can both assist surgeons in evaluating the

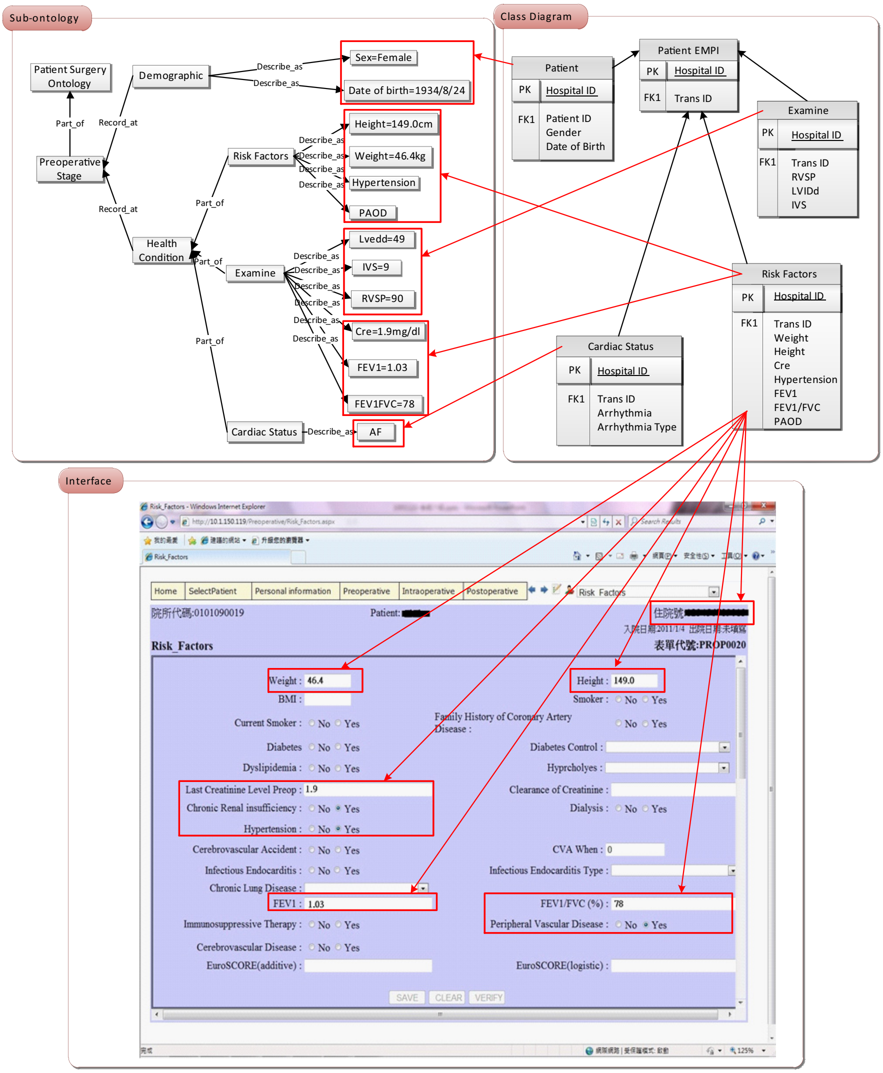


Fig. 5 Shows how a sub-ontology of current patient risk factors was used to develop a conceptual database schema, and to design the system screen

expected outcome for a given patient and facilitate counseling and preoperative decision-making. A prediction model is an intelligent data analysis function that is often used for clinical decision-making prior to and during surgery.

Machine learning techniques are gradually receiving more and more attention in research field of medical informatics. Weka is a well-known open-source data mining tool, which includes many impressive machine learning and data mining algorithms. Weka provides a menu-based interface, called Weka Explorer, for general users or visualize Weka KnowledgeFlow interface to accomplish intelligent data analysis tasks.

Consequently, medical practitioners could get intelligent data analysis tool through Weka. Relative to the regular desktop application, web service is easier to maintain, deliver and access through network. In this study, by combining the CRIS and asynchronous web service calls, it is available to access immediate data analysis services from WEKA library[12].

5 Results and Discussions

The system is still being tested and the developed database structure allows for in-depth analyses of stored data [13].We hope that field tests can begin in the near future, in the framework of a national cardiovascular infrastructure program. Information reliability in a cardiovascular system can be provided through statistical analysis.

This study illustrated a three-stage methodological approach for ontology based CRIS design. Following the three-stage approach, we can collect and organize knowledge in a cardiovascular surgery domain area. Besides, the user-centered design process should reflect a development framework for patient-centered system designs. With ontology models we can easily convert “data” into “reality”. Our ultimate goal is to expand our ontology approach to many different healthcare domains, helping developers to build healthcare systems which can satisfy end users.

Thus, given that the cost consideration of full implementation is negligible, and the potential benefits in terms of increased patients care’s quality are large, we plan to expand the system to other cooperating hospitals. We anticipate that in the future, inter-hospital systems will achieve consistent application and significantly improve patient’s care quality.

References

1. Hsieh, N.-C., Chang, C.-Y., Lee, K.-C., Chen, J.-C., Chan, C.-H.: Technological Innovations in the Development of Cardiovascular Clinical Information Systems. Journal of Medical Systems, 1–14 (2010)
2. Wright, J.G., Bieniewski, C.L., Pifarre, R., Gunnar, R.M., Scanlon, P.J.: A database management system for cardiovascular disease. Computer Methods and Programs in Biomedicine 20(1), 117–121 (1985)
3. Legarreta, J.H., Boto, F., Macía, I., Maiora, J., García, G., Paloc, C., Graña, M., de Blas, M.: Hybrid Decision Support System for Endovascular Aortic Aneurysm Repair Follow-Up. In: Graña Romay, M., Corchado, E., Garcia Sebastian, M.T. (eds.) HAIS 2010. LNCS, vol. 6076, pp. 500–507. Springer, Heidelberg (2010)
4. Kauert, R., Topel, T., Scholz, U., Hofestadt, R.: Information System for the Support of Research. In: Diagnosis and Therapy of Inborn Metabolic Diseases. In: Medinfo 2001, London, pp. 353–356 (2001)
5. Shortliffe, E.H., Perreault, L.E., Wiederhold, G., Fagan, L.M.: Medical Informatics: Computer applications in health care and biomedicine, 2nd edn. Springer, New York (2001)

6. Bellazzi, R., Zupan, B.: Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of Medical Informatics 77, 81–97 (2008)
7. Sachs, E.: Getting started with Protégé-frames. Bmj Clinical Research Ed 307(6919), 1619–1620 (2006)
8. Corcho, O., Fernández-López, M., Gómez-Pérez, A.: Methodologies, tools and languages for building ontologies: where is their meeting point? Data & Knowledge Engineering 46(1), 41–64 (2003)
9. Pinto, H.S., Martins, J.P.: Ontologies: How can They be Built? Knowledge and Information Systems 6(4), 441–464 (2004)
10. Wright, J.G., Bieniewski, C.L., Pifarre, R., Gunnar, R.M., Scanlon, P.J.: A database management system for cardiovascular disease. Computer Methods and Programs in Biomedicine 20(1), 117–121 (1985)
11. Siricharoen, W.V.: Ontology modeling and object modeling in software engineering. International Journal of Software Engineering and Its Applications 3(1), 43–59 (2009)
12. Shetty, S.D., Vadivel, S., Vaghella, S.: Weka based desktop data mining as web service. World Academy of Science, Engineering and Technology, 697–715 (2010)
13. Wright, J.G., Bieniewski, C.L., Pifarre, R., Gunnar, R.M., Scanlon, P.J.: A database management system for cardiovascular disease. Computer Methods and Programs in Biomedicine 20(1), 117–121 (1985)

Revenue Sharing Contract in Dual Channel Supply Chain in Case of Free Riding

Liu Yezheng and Ding Zhengping

Abstract. After the Internet channel emerged, consumers can go to the traditional retailer store and accept sales services, and then buy the product from the Internet store. Thus a free riding problem occurs. This paper analyzed the optimal prices of dual channel for the entire supply chain and competing equilibriums of various dual channel structures when free riding problem exists. This paper also studied the application of coordination strategy of revenue sharing contract to various dual channel structures when free riding problem exists. The result shows that revenue sharing contract can fully coordinate the entire supply chain of decentralized dual channel, and can't fully coordinate the entire supply chain of horizontally integrated dual channel, the revenue sharing contract and a coordination policy of fixed price difference can fully coordinate the entire supply chain of partially integrated dual channel.

1 Introduction

With the rapid growth and exciting potential of Internet channel, more and more firms utilize both the traditional channel and Internet channel to sell their products. Many articles researched the price equilibrium or service equilibrium of dual channel of manufacturer's direct channel and retailer's traditional channel. Chiang et al. showed that the manufacturer's direct marketing can indirectly increase the flow of profits through the retail channel, help the manufacturer improve overall profitability by reducing the degree of inefficiency caused by double marginalization [1]. Yao et al. studied the Stackelberg game and Bertrand game between the manufacturer and the retailer when the manufacturer introduced a direct channel, the conclusion showed that the retailer will increase the selling effort and both the manufacturer and the retailer can gain more profits [2]. Yan et al. pointed out the retailer can use high quality of service weakening the competitiveness of manufacturer's direct channel and making the entire supply chain better [3].

The introduction of Internet channel makes parts of consumers turn to shopping on Internet and brings channel conflict. Webb et al. proposed twelve methods of softening the channel conflict [4]. Cattani et al. compared manufacturer's three

Liu Yezheng · Ding Zhengping
School of Management, Hefei University of Technology, P.R. China, 230009

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 459–469.
springerlink.com

pricing strategies (keep wholesale price unchanged, keep retail price unchanged, select wholesale price and retail price that optimize profits for the manufacturer) for dual channel which are used to diminish the channel conflict. He found that the third strategy is preferred by both the manufacturer and the retailer [5]. Xiao considered the strategy of letting the retailer finish the manufacturer's direct channel's orders and found that the strategy is attributable to both the manufacturer and the retailer under certain conditions [6]. Theoretically, channel coordination can make the retailer gain more profits and eliminate the channel conflict effectively [7]. When there is only one traditional channel, many coordination mechanisms such as revenue sharing contract, quantity discount contract, buy back contract and so on are designed to eliminate double marginalization [8,9]. Some scholars researched the application of coordination strategy used in single traditional channel to dual channel. Boyaci found that the buy back contract can't fully coordinate the entire supply chain in dual channel [10]. Xie and Cai et al. found that the quantity discount contract can fully coordinate the entire supply chain in dual channel of manufacturer's direct channel and retailer's traditional channel [11, 12].

The introduction of Internet channel also brings the free riding problem. Many consumers firstly accept the retail services such as explaining the product in detail, touch and feel the real product in the traditional store, and then buy the product through the Internet store after making purchase decision. Such behavior is called free riding. The free riding problem occurs if the presales activities can be conducted separately from the actual sale of the product. Free riding always hurt the retailer which provides the service [13]. Xin studied the impact of free riding on retailer's pricing strategy in single traditional channel [14]. Wu studied the impact of free riding problem on the retailer who provides the costly service under the circumstance of dual channel where the searching cost of consumers is decreased [15]. Because of the difficulty of providing the services provided by the traditional retailer for the Internet retailer and the lower price for Internet channel, the free riding behavior becomes a popular phenomenon in dual channel. Baal et al. found that 20% of the consumers are free riders [16]. Carlton et al. showed that the manufacturer is apt to set a higher or equal price in Internet channel to weaken the free riding problem [17].

In summary, the research on the price equilibrium or coordination strategy in dual channel when free riding problem exists is rare. On the other hand, existing researches mostly concentrate on the dual channel structure of manufacturer's direct channel and retailer's traditional channel. However, among the top 10 U.S. Internet retail sites in 2008, only two were manufacturer's direct channels, whereas five were owned by traditional retailers, and the rest were pure Internet retailers [18]. The existing studies are lack of analysis of these diverse channel structure possibilities. This paper studied the pricing strategy for the optimization of the entire supply chain in dual channel and pricing equilibriums of three different dual channel structures in case of free riding when the supply chain members make decision separately. This paper also studied whether the coordination strategy of revenue sharing contract can fully coordinate the entire supply chain of various dual channel structures when free riding problem exists.

2 Model

Similar to Xiao and Guo et al. [6, 19], we assume a linear downward sloping demand function. The demand of traditional channel and Internet channel are given as below respectively:

$$d_r = a_r - bp_r - \theta(p_r - p_e), \tag{1}$$

$$d_e = a_e - bp_e + \theta(p_r - p_e). \tag{2}$$

Where $p_i, i \in \{r, e\}$ is the product's retail price in channel i ($i = r$ denotes the traditional channel, $i = e$ denotes the Internet channel). $a_i, i \in \{r, e\}$ is the market potential of channel i. b is the price elasticity factor. θ is the factor of demand change with respect to the price in the other channel. So the change of quantity from the traditional channel to Internet channel is $\theta(p_r - p_e)$. The changed quantity contains two parts. One is the quantity that consumers buying from Internet store directly, the other is the quantity related to free riding. We assume the free riding ratio is μ .That is to say the quantity bought from free riders is

$$d_f = \mu\theta(p_r - p_e). \tag{3}$$

Let $c_i, i \in \{r, e\}$ denote the retail cost for each customer who visits the store in channel i. Similar to Xin (2007), we assume that the traditional retailer will incur selling cost to serve a customer who may or may not purchase a product. For the quantity of d_f, both the retailers in each channel should pay the corresponding costs. There are two reasons for having such an assumption. One is that traditional retailer must expand time and effort to help customers identify the product that best fits their needs irrespective of whether they buy the product. The other is that traditional retailer gives up its opportunity to make another sale by serving the wrong customers (free riders).

In order to utilize the above model, we assume $a_r > a_e$, $c_r > c_e$ and $a_r - a_e + (2\theta + b)(c_r - c_e) - 2\mu\theta c_r > 0$. $a_r > a_e$ means that the traditional channel has a larger market potential. $c_r > c_e$ means that the traditional retailer should pay a higher unit cost. $a_r - a_e + (2\theta + b)(c_r - c_e) - 2\mu\theta c_r > 0$ means $p_r > p_e$ when the entire supply chain realize the maximization of profits and also ensures the existence of free riding problem under various dual channel structures. We also have the assumption of $a_i > bc_i, i \in \{r, e\}$, which can guarantee that prices exceed marginal costs and quantities are nonnegative.

Similar to Yoo et al. [18], we considered four dual channel arrangements. They are vertically integrated dual channel (VID), partially integrated dual channel (PID), decentralized dual channel (DD) and horizontally integrated dual channel

(HID). The corresponding channel structures are shown in figure 1. The real line in the figure means the two parts belong to one entity. The broken line in the figure means the two parts belong to two entities. In vertically integrated dual channel (e.g., Dell), the manufacturer called vertically integrated manufacturer (vim) owns both the Internet channel and the traditional channel. In partially integrated dual channel (e.g., Lenovo and Haier), the manufacturer called partially integrated manufacturer (pim) owns Internet channel. It competes with the traditional retailer (tr) in the market. In decentralized dual channel (e.g., Amazon.com), the manufacturer (m) doesn't owns any channel, and the traditional retailer (tr) and Internet retailer (er) compete with each other in the market. In horizontally integrated dual channel (e.g., Wal-mart), the retailer called horizontally integrated retailer (hir) owns both the Internet channel and the traditional channel.

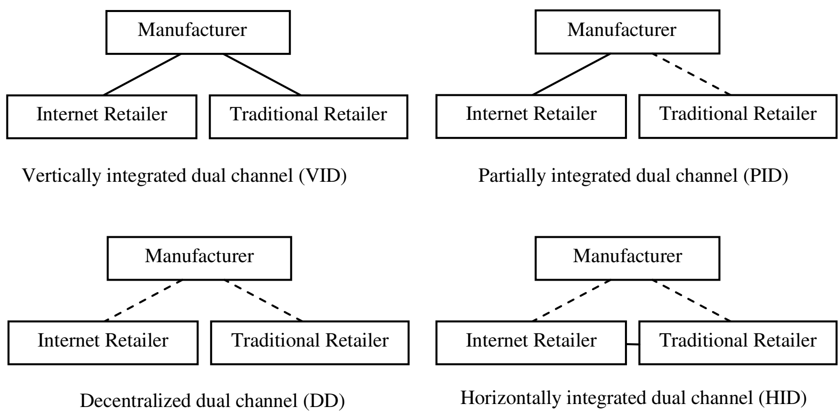


Fig. 1 Four Dual Channel Structures

Without coordination, the players optimize their own profits sequentially rather than maximize the entire supply chain profits globally. Similar to numerous previous studies, we assume the following: (1) The manufacturer, as the Stackelberg leader, sets the wholesale price to maximize its own profits with the foresight of the independent retailers' optimal responses; (2) As a Stackelberg follower, the independent retailer sets its retail price to maximize its own profits conditional on the wholesale price. The price competition between two independent retailers is assumed to be Bertrand–Nash; (3) If dual channel members are vertically or horizontally integrated, their pricing decisions are coordinated to maximize the joint profits. When the supply chain utilizes coordinated strategy of revenue sharing contract, the retailer shares ρ $(0<\rho<1)$ of its revenue with the manufacturer. The value of ρ generally reflects the negotiation power between the manufacturer and the retailer and is usually determined in a negotiation process which is not a focus of this paper.

3 Model Analysis

3.1 Vertically Integrated Dual Channel

The profits of the vertically integrated manufacturer is $\pi_{vim}^{VID} = (p_r - c_r)d_r + (p_e - c_e)d_e - c_r d_f$. The vertically integrated manufacturer decides p_r and p_e to maximize its profits. We can know that π_{vim}^{VID} is jointly concave with respect to p_r and p_e. Letting the differentials of π_{vim}^{VID} on p_r and p_e both be zero, we can get

$$p_r = \frac{\theta a_e + \theta a_r + b a_r + 2\theta b c_r + b^2 c_r - \mu\theta b c_r}{2b(b+2\theta)}, \tag{4}$$

$$p_e = \frac{\theta a_e + \theta a_r + b a_e + 2\theta b c_e + b^2 c_e + \mu\theta b c_r}{2b(b+2\theta)}. \tag{5}$$

The profits of the entire supply chain are π_{vim}^{VID} in every dual channel structure. It is to say that the entire supply chain will realize the maximization of profits if p_r and p_e satisfied equation (4) and equation (5) respectively. So we have the following result.

Result 1. The retail prices of traditional channel and Internet channel are equation (4) and equation (5) respectively in vertically integrated dual channel, and the entire supply chain realizes the maximization of profits.

3.2 Partially Integrated Dual Channel

Without coordination, the profits of the partially integrated manufacturer and the traditional retailer are $\pi_{pim}^{PID} = w_{tr} d_r + (p_e - c_e)d_e$ and $\pi_{tr}^{PID} = (p_r - w_{tr})d_r - -(d_r + d_f)c_r$. Where w_{tr} is the wholesale price for the traditional retailer. The sequence of the game is that the partially integrated manufacturer sets w_{tr} first and then the partially integrated manufacturer and the traditional retailer set p_r and p_e respectively at the same time. We can know that π_{pim}^{PID} is concave with respect to p_e and π_{tr}^{PID} is concave with respect to p_r. In accordance with backward induction, letting the differential of π_{pim}^{PID} on p_e and the differential of π_{tr}^{PID} on p_r both be zero, we obtain

$$p_r = \frac{\begin{array}{l} 4\theta b w_{tr} - 2\mu\theta^2 c_r + b\theta c_e + 2b^2 c_r + 4\theta b c_r + 2b a_r + \\ 2b^2 w_{tr} - 2\mu\theta b c_r + 2\theta a_r + 3\theta^2 w_{tr} + 2\theta^2 c_r + \theta^2 c_e + \theta a_e \end{array}}{(3\theta^2 + 4b^2 + 8b\theta)}, \tag{6}$$

$$p_e = \frac{\theta a_r + 3w_{tr}\theta^2 + \theta c_r b + \theta^2 c_r + 3\theta b w_{tr} + 2c_e b^2 - \mu c_r \theta^2 + 2ba_e + 2\theta^2 c_e + 2\theta a_e + 4b\theta c_e}{3\theta^2 + 4b^2 + 8b\theta}. \quad (7)$$

Substituting equation (6) and equation (7) into $\pi_{pim}{}^{PID}$ and letting the differential of $\pi_{pim}{}^{PID}$ on w_{tr} be zero, we obtain

$$w_{tr} = \frac{\begin{array}{l}25\theta^3 ba_e - 16\theta^4 bc_r - 8b^5 c_r + 9\theta^4 a_e + 32\mu\theta^2 b^3 c_r + 8\mu\theta b^4 c_r + 32\theta^3 ba_r - \\ 2\theta^4 bc_e + 48\theta^2 b^2 a_r - 56\theta^3 b^2 c_r - 72\theta^2 b^3 c_r + 32\theta b^3 a_r + 8\theta b^3 a_e + \\ 16\mu\theta^4 bc_r - \theta^3 b^2 c_e + 40\mu\theta^3 b^2 c_r + 24\theta^2 b^2 a_e + 8b^4 a_r + 9\theta^4 a_r - 40\theta b^4 c_r\end{array}}{2b(b+\theta)(b+2\theta)(8b^2+16\theta b+9\theta^2)}. \quad (8)$$

Substituting equation (8) into equation (6) and equation (7), we can get

$$p_r = \frac{\begin{array}{l}4b^5 c_r + 20\theta b^4 c_r + 4\theta b^4 c_e + 12b^4 a_r - 4\mu\theta b^4 c_r + 36\theta^2 b^3 c_r - 16\mu\theta^2 b^3 c_r + 16\theta^2 b^3 c_e \\ +48\theta b^3 a_r + 8\theta b^3 a_e - 20\mu\theta^3 b^2 c_r + 21\theta^3 b^2 c_e + 28\theta^3 b^2 c_r + 24\theta^2 b^2 a_e + 70\theta^2 b^2 a_r \\ -8\mu\theta^4 bc_r + 44\theta^3 ba_r + 25\theta^3 ba_e + 10\theta^4 bc_e + 8\theta^4 bc_r + 9\theta^4 a_r + 9\theta^4 a_e\end{array}}{2(b+\theta)(b+2\theta)(8b^2+16\theta b+9\theta^2)}, \quad (9)$$

$$p_e = \frac{\begin{array}{l}8b^4 c_e + 32\theta b^3 c_e - 2\theta b^3 c_r + 8b^3 a_e - 6\theta^2 b^2 c_r + 2\mu\theta^2 b^2 c_r + 43\theta^2 b^2 c_e + 24\theta b^2 c_e \\ +10\theta b^2 a_r + 22\theta^3 bc_e + 20\theta^2 ba_r + 4\mu\theta^3 bc_r + 25\theta^2 ba_e - 4\theta^3 bc_r + 9\theta^3 a_r + 9\theta^3 a_e\end{array}}{2b(b+2\theta)(8b^2+16\theta b+9\theta^2)}. \quad (10)$$

It is easy to see that equation (9) isn't equal to equation (4), equation (10) isn't equal to equation (5) either. In other words, it doesn't realize the optimized profits of the entire supply chain. So we have the following result.

Result 2. Without coordination, the wholesale price for the traditional retailer is equation (8) and the retail prices of traditional channel and Internet channel are equation (9) and equation (10) respectively in partially integrated dual channel.

When the supply chain utilizes revenue sharing contract, the profits of the partially integrated manufacturer and the traditional retailer are $\pi_{pim}{}^{PID} = w_{tr} d_r + (p_e - c_e) d_e + \rho p_r d_r$ and $\pi_{tr}{}^{PID} = ((1-\rho) p_r - w_{tr} - c_r) d_r - c_r d_f$. Similar to Cai [20], we introduce a fixed price difference scheme $p_r = p_e + \varepsilon$ and obtain the following theorem.

Theorem 1. The coordination strategy of

$$p_r = p_e + \frac{a_r - a_e + (2\theta + b)(c_r - c_e) - 2\mu\theta c_r}{2(b+2\theta)}, \quad (11)$$

and the revenue sharing contract with

$$w_{tr}=\frac{\begin{array}{c}2\rho\theta^2c_e-2\rho\theta^2c_r+\theta bc_r-\theta bc_e-2\mu\theta^2c_r-5\rho\theta bc_r+\rho\theta bc_e+2\rho\mu\theta^2c_r-\\ \theta a_r+\theta a_e+2\theta^2c_r-2\theta^2c_e+\rho\theta a_r-\rho\theta a_e-2\mu\theta bc_r-2\rho b^2c_r+2\rho\mu\theta bc_r\end{array}}{2b(b+2\theta)}, \quad (12)$$

fully coordinates the supply chain with partially integrated dual channel.

Proof. Given equation (11) and equation (12), the optimal p_r for the partially integrated manufacturer is equation (4) and the optimal p_e for the traditional retailer is equation (5). That is to say the equilibrium prices are same to the optimal prices for the entire supply chain.

3.3 Decentralized Dual Channel

Without coordination, the profits of the manufacturer, the traditional retailer and the Internet retailer are $\pi_m{}^{DD}=w_{tr}d_r+w_{er}d_e$, $\pi_{tr}{}^{DD}=(p_r-w_{tr}-c_r)d_r-c_rd_f$ and $\pi_{er}{}^{DD}=(p_e-w_{er}-c_e)d_e$. Where w_{tr} and w_{er} are the wholesale prices for the traditional retailer and the Internet retailer respectively. The sequence of the game is that the manufacturer sets w_{tr} and w_{er} first and then the traditional retailer and the Internet retailer set p_r and p_e respectively at the same time. We can know that $\pi_{tr}{}^{DD}$ is concave with respect to p_r and $\pi_{er}{}^{DD}$ is concave with respect to p_e. In accordance with backward induction, letting the differential of $\pi_{er}{}^{DD}$ on p_e and the differential of $\pi_{tr}{}^{DD}$ on p_r both be zero, we obtain

$$p_r=\frac{\begin{array}{c}4\theta bw_{tr}+\theta bc_e-2\mu\theta^2c_r+2b^2c_r+\theta bw_{er}+4\theta bc_r+2b^2w_{tr}+2ba_r\\ -2\mu\theta bc_r+2\theta a_r+\theta a_e+2\theta^2c_r+2\theta^2w_{tr}+\theta^2c_e+\theta^2w_{er}\end{array}}{3\theta^2+8\theta b+4b^2}, \quad (13)$$

$$p_e=\frac{\begin{array}{c}\theta a_r+2\theta a_e+\theta bc_r+\theta^2c_r+\theta bw_{tr}+\theta^2w_{tr}-\mu\theta^2c_r+2b^2w_{er}+\\ 4\theta bw_{er}+2ba_e+4\theta bc_e+2\theta^2c_e+2\theta^2w_{er}+2b^2c_e\end{array}}{3\theta^2+8\theta b+4b^2}. \quad (14)$$

Substituting equation (13) and equation (14) into $\pi_m{}^{DD}$ and letting the differentials of $\pi_m{}^{DD}$ on w_{tr} and w_{er} both be zero, we obtain

$$w_{tr}=\frac{b^2a_r-b^3c_r-3\theta b^2c_r+\theta^2a_e+\theta^2a_r+\theta ba_e-2\theta^2bc_r+2\theta ba_r+2\mu\theta^2bc_r+\mu\theta b^2c_r}{2b(b+\theta)(b+2\theta)}, \quad (15)$$

$$w_{er}=\frac{ba_e+\theta a_r+\theta a_e-b^2c_e-2\theta bc_e}{2b(b+2\theta)}. \quad (16)$$

Substituting equation (15) and equation (16) into equation (13) and equation (14), we can get

$$p_r = \frac{\begin{array}{c}2b^4c_r + 8\theta b^3c_r + \theta b^3c_e + 6b^3a_r - 2\mu\theta b^3c_r - 6\mu\theta^2b^2c_r + 10\theta^2b^2c_r + 3\theta^3a_e + 3\theta^3a_r + \\ 3\theta^2b^2c_e + 15\theta^2ba_r + 10\theta^2ba_e + 4\theta^3bc_r - 4\mu\theta^3bc_r + 2\theta^3bc_e + 18\theta b^2a_r + 5\theta b^2a_e\end{array}}{2b(b+2\theta)(2b+\theta)(2b+3\theta)}, \quad (17)$$

$$p_e = \frac{\begin{array}{c}2b^4c_e + \theta b^3c_r + 8\theta b^3c_e + 6b^3a_e - \mu\theta^2b^2c_r + 18\theta b^2a_e + 3\theta^2b^2c_r + 5\theta b^2a_r + \\ 10\theta^2b^2c_e + 10\theta^2ba_r + 15\theta^2ba_e + 2\theta^3bc_r - 2\mu\theta^3bc_r + 4\theta^3bc_e + 3\theta^3a_r + 3\theta^3a_e\end{array}}{2b(b+2\theta)(2b+\theta)(2b+3\theta)}. \quad (18)$$

It is easy to see that equation (17) isn't equal to equation (4), equation (18) isn't equal to equation (5) either. In other words, it doesn't realize the optimized profits of the entire supply chain. So we have the following result.

Result 3. Without coordination, the wholesale price for the traditional retailer and the Internet retailer are equation (15) and equation (16) respectively and the retail prices of traditional channel and Internet channel are equation (17) and equation (18) respectively in decentralized dual channel.

When the supply chain utilizes revenue sharing contract, the profits of the manufacturer, the traditional retailer and the Internet retailer are $\pi_m{}^{DD} = (w_{tr} + \rho p_r)d_r + (w_{er} + \rho p_e)d_e$, $\pi_{tr}{}^{DD} = ((1-\rho)p_r - w_{tr} - c_r)d_r - c_r d_f$ and $\pi_{er}{}^{DD} = ((1-\rho)p_e - w_{er} - c_e)d_e$. We obtain the following theorem.

Theorem 2. The revenue sharing contracts with

$$w_{tr} = \frac{\begin{array}{c}\rho\theta b^2c_e - 6\rho\theta b^2c_r - 2\rho b^3c_r + 2\rho\mu\theta b^2c_r - \rho\theta ba_e - \theta b^2c_e + \mu\theta^2bc_r - 2\theta^2bc_e + \\ 2\rho\theta^2bc_e + \theta ba_e + 3\rho\mu\theta^2bc_r - 4\rho\theta^2bc_r - \rho\theta^2a_r + \theta^2a_r - \rho\theta^2a_e + \theta^2a_e\end{array}}{2b(b+\theta)(b+2\theta)}, \quad (19)$$

$$w_{er} = \frac{\begin{array}{c}2\mu\theta b^2c_r - 2\rho b^3c_e - 66\rho\theta b^2c_e - \theta b^2c_r - 2\rho\mu\theta b^2c_r + \rho\theta b^2c_r + 3\mu\theta^2bc_r - \rho\theta^2a_e + \\ 2\rho\theta^2bc_r - 4\rho\theta^2bc_e - 2\theta^2bc_r - 3\rho\mu\theta^2bc_r - \rho\theta ba_r + \theta ba_r + \theta^2a_e - \rho\theta^2a_r + \theta^2a_r\end{array}}{2b(b+\theta)(b+2\theta)}. \quad (20)$$

fully coordinates the supply chain with decentralized dual channel.

Proof. Given equation (19) and equation (20), the solution of p_r and p_e for the traditional retailer and the Internet retailer in a Nash game are equation (4) and equation (5) respectively. That is to say the equilibrium prices are same to the optimal prices for the entire supply chain.

3.4 Horizontally Integrated Dual Channel

Without coordination, the profits of the manufacturer and the horizontally integrated retailer are $\pi_m{}^{HID} = w_{hir}(d_r + d_e)$ and $\pi_{hir}{}^{HID} = (p_r - w_{hir} - c_r)d_r - c_r d_f +$

$(p_e - w_{hir} - c_e)d_e$. Where w_{hir} is the wholesale price for the horizontally integrated retailer. The sequence of the game is that the manufacturer sets w_{hir} first and then the horizontally integrated retailer set p_r and p_e. We can know that π_{hir}^{HID} is jointly concave with respect to p_r and p_e. In accordance with backward induction, letting the differentials of π_{hir}^{HID} on p_r and p_e both be zero, we obtain

$$p_r = \frac{\theta a_e + 2\theta bw + b^2 c_r + 2\theta bc_r + ba_r + b^2 w - \mu\theta bc_r + \theta a_r}{2b(b+2\theta)}, \tag{21}$$

$$p_e = \frac{2\theta bw + b^2 c_e + 2\theta bc_e + ba_e + b^2 w + \mu\theta bc_r + \theta a_e + \theta a_r}{2b(b+2\theta)}. \tag{22}$$

Substituting equation (21) and equation (22) into π_m^{HID} and letting the differential of π_m^{HID} on w_{hir} be zero, we obtain

$$w_{hir} = \frac{a_r + a_e - bc_r - bc_e}{4b}. \tag{23}$$

Substituting equation (23) into equation (21) and equation (22), we can get

$$p_r = \frac{6\theta a_e - 2\theta bc_e + 6\theta bc_r + 6\theta a_r + 3b^2 c_r + 5ba_r - b^2 c_e + ba_e - 4\mu\theta bc_r}{8b(b+2\theta)}, \tag{24}$$

$$p_e = \frac{6\theta a_e - 2\theta bc_r + 6\theta bc_e + 6\theta a_r + 3b^2 c_e + 5ba_e - b^2 c_r + ba_r + 4\mu\theta bc_r}{8b(b+2\theta)}. \tag{25}$$

It is easy to see that equation (24) isn't equal to equation (4), equation (25) isn't equal to equation (5) either. In other words, it doesn't realize the optimal profits of the entire supply chain. So we have the following result.

Result 4. Without coordination, the wholesale price for the horizontally integrated retailer is equation (23) and the retail prices of traditional channel and Internet channel are equation (24) and equation (25) respectively in horizontally integrated dual channel.

When the supply chain utilizes revenue sharing contract, the profits of the manufacturer and the horizontally integrated retailer are $\pi_m^{HID} = w_{hir}(d_r + d_e) + \rho(p_r d_r + p_e d_e)$ and $\pi_{hir}^{HID} = (1-\rho)(p_r d_r + p_e d_e) - (d_r + d_e)w_{hir} - (d_r + d_f)c_r - c_e d_e$. We obtain the following theorem.

Theorem 3. The revenue sharing contracts can't fully coordinate the supply chain with horizontally integrated dual channel.

Proof. Given w_{hir}, the solution of p_r and p_e for the horizontally integrated retailer are

$$p_r = \frac{\theta a_r - \rho\theta a_r + \theta a_e - \rho\theta a_e - \rho b a_r + 2\theta b c_r + 2\theta b w_{hir} + b^2 w_{hir} + b a_r + b^2 c_r - \mu\theta b c_r}{2b(1-\rho)(b+2\theta)}, \quad (26)$$

$$p_e = \frac{2\theta b w_{hir} - \rho b a_e + \theta a_r - \rho\theta a_r + \theta a_e - \rho\theta a_e + \mu\theta b c_r + b a_e + b^2 w_{hir} + 2\theta b c_e + b^2 c_e}{2b(1-\rho)(b+2\theta)}. \quad (27)$$

Letting equation (26) equal to equation (4) and equation (27) equal to equation (5), we can get

$$w_{hir} = \frac{-\rho c_r (b + 2\theta - \mu\theta)}{b + 2\theta}, \quad (28)$$

$$w_{hir} = \frac{-\rho(2\theta c_e + \mu\theta c_r + b c_e)}{b + 2\theta}. \quad (29)$$

Equation (28) isn't equal to equation (29). That is to say there is no w_{hir} which can make the horizontally integrated retailer set the retail prices equal to the optimal prices for the entire supply chain.

4 Conclusion

When the Internet channel has a lower retail price, free riding becomes a popular phenomenon. In this paper we classified dual channel structures into four categories according to the entity which owns the Internet channel, and analyzed the optimal prices of dual channel for the entire supply chain and pricing equilibriums in various channel structures when the channel members decide separately. The result shows that the pricing equilibriums can't realize the maximal profits of the entire supply chain. Then we studied the application of revenue sharing contract to various channel structures and obtained the following results: The revenue sharing contract can be applied in partially integrated dual channel and decentralized dual channel and fully coordinate the entire supply chain; the revenue sharing contract can't fully coordinate the entire supply chain of horizontally integrated dual channel.

Our research has important managerial insight. We show that the revenue sharing contract has different performances in different dual channel structures. If the supply chain utilized coordination strategy of revenue sharing contract before the Internet channel being introduced, it can continue utilize the same coordination strategy in partially integrated dual channel or decentralized dual channel, but it will apply itself to find another coordination strategy in horizontally integrated dual channel. How to find such a coordination strategy should be researched further. There is another kind of free riding that the consumers search for information in Internet channel and then buy the product from traditional store in dual channel circumstance. How to coordinate the supply chain when such a free riding problem exists is another interesting research issue.

References

1. Chiang, W., Chhajed, D., Hess, J.: Directmarketing, indirect profits: Astrategic analysis of dual-channel supply-chain design. Management Science 49(1), 1–20 (2003)
2. Yao, D., Liu, J.: Competitive pricing of mixed retail and e-tail distribution chan-nels. Omega 33(3), 235–247 (2005)
3. Yan, R., Pei, Z.: Retail services and firm profit in a dual-channel market. Journal of Retailing and Consumer Services 16(4), 306–314 (2009)
4. Webb, K., Lambe, C.: Internal multi-channel conflict: An exploratory investigation and conceptual framework. Industrial Marketing Management 36(1), 29–43 (2007)
5. Cattani, K., Gilland, W., Heese, H., et al.: Boiling frogs: Pricing strategies for a manufacturer adding an Internet channel. Production and Operations Management 15(1), 40–57 (2005)
6. Xiao, J.: Study on cooperation strategy between electronic channels and retailers in dual-channel supply chain. Journal of Systems Engineering 24(6), 673–679 (2009) (in Chinese)
7. Graves, S., Kok, G.: Handbooks in operations research and management science: Supply chain management. Elsevier Publishing Company, North Holland (2003)
8. Cachon, G., Lariviere, M.: Supply chain coordination with revenue sharing: Strengths and limitations. Management Science 51(1), 30–44 (2005)
9. Jeuland, A., Shugan, S.: Managing channel profits. Marketing Science 2(3), 239–272 (1983)
10. Boyaci, T.: Competitive stocking and coordination in a multiple-channel distribution system. IIE Transactions (2005)
11. Xie, Q., Huang, P.: A quantity discount model for coordination of Internet-based hybrid channels. Systems Engineering: Theory & Practice 27(8), 1–11 (2007) (in Chinese)
12. Cai, G., Zhang, Z.: Game theoretical perspectives on dual-channel supply chain competition with price discounts and pricing schemes. International Journal of Production Economics 117(1), 80–96 (2009)
13. Kucuk, S.: The role of the Internet on free-riding: An exploratory study of the wallpaper industry. Journal of Retailing and Consumer Services 17(4), 313–320 (2010)
14. Xin, J.: How does free riding on customer service affect competition. Marketing Science 26(4), 488–503 (2007)
15. Wu, D.: Implications of reduced search cost and free riding in E-commerce. Marketing Science 23(2), 255–262 (2004)
16. van Baal, S., Dach, C.: Free riding and customer retention across re-tailers' channels. Journal of Interactive Marketing 19(2), 75–85 (2005)
17. Carlton, D., Chevalier, J.: Free riding and sales strategies for the Internet. The Journal of Industrial Economics 49(4), 441–461 (2001)
18. Yoo, W., Lee, E.: Internet channel entry: A strategic analysis of mixed channel structures. Marketing Science 30(1), 29–41 (2011)
19. Guo, Y., Qu, D.: Analysis of Pricing Policies of hybrid distribution channels in E-market. Journal of Systems Engineering 23(5), 570–576 (2008) (in Chinese)
20. Cai, G.: Channel selection and coordination in dual-channel supply chains. Journal of Retailing 86(3), 22–36 (2010)

Some Consideration of SIRMs Connected Fuzzy Inference Model with Functional Weights

Hirosato Seki and Tomoharu Nakashima

Abstract. This paper discusses the SIRMs (Single-Input Rule Modules) connected fuzzy inference model with functional weights (SIRMs model with FW). The SIRMs model with FW consists of a number of groups of simple fuzzy if-then rules with only a single attribute in the antecedent part. The final outputs of conventional SIRMs model are obtained by summarizing product of the functional weight and inference result from a rule module. In the SIRMs model of the paper, we use square functional weights, and compare with the conventional model.

1 Introduction

Single Input Rule Modules (SIRMs) connected fuzzy inference model proposed by Yubazaki et al.[12, 6, 7, 8, 9, 10] consists of several groups of single-input fuzzy if-then rules. The advantage of the SIRMs models is computational efficiency. The number of involved fuzzy if-then rules is significantly less than the conventional model where each fuzzy if-then rule has as many inputs as the problem domain at hand. The SIRMs model has been applied to the automatic control of vehicles and the stability control of inverted pendulums. Seki et al.[2, 3, 5] analyzed the property of the SIRMs inference models and also proposed extended versions of SIRMs models.

In the standard formulation of SIRMs, the weight for each rule module is considered to be constant. However, it is often necessary to have variable weights for connecting rule modules in order to accommodate high-dimensional problem domains, which is the case in most of the practical cases. From this perspective, Yi et al.[9] proposed a new SIRMs model that introduces dynamic weights for rule modules and showed the effective ness of the model for control domains.

Hirosato Seki
Kwansei Gakuin University, 2-1, Gakuen, Sanda-shi, Hyogo 669-1337, Japan
e-mail: h_seki@ieee.org

Tomoharu Nakashima
Osaka Prefecture University, address
e-mail: name@email.address

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 471–476.
springerlink.com

On the other hand, SIRMs connected fuzzy inference model with functional weights has been also proposed in [1]. In this model, the weights for connecting rule modules are variable rather than constant. In order to realize this, funcional weights such as linear expression are employed in the model.

In this paper, we consider the SIRMs model with square functional weights. Moreover, it is compared with the conventional SIRMs model, and shown the applicability of them.

2 A SIRMs Connected Fuzzy Inference Model with Functional Weights

In this section, we review the *SIRMs connected fuzzy inference model with functional weights* (SIRMs model with FW) [1] in which the weight for the inference result of each rule module is generalized to a function, where the system has n inputs and a single output, and each rule module corresponds to one of the n input attributes and has only that input attribute in the antecedent part of fuzzy if-then rules in the rule module.

The rules of the SIRMs model with FW are given as follows:

$$\begin{aligned} &\text{Rules-}1 : \{x_1 = A_j^1 \longrightarrow y_1 = y_j^1\}_{j=1}^{m_1} \\ &\qquad \vdots \\ &\text{Rules-}i : \{x_i = A_j^i \longrightarrow y_i = y_j^i\}_{j=1}^{m_i} \\ &\qquad \vdots \\ &\text{Rules-}n : \{x_n = A_j^n \longrightarrow y_n = y_j^n\}_{j=1}^{m_n} \end{aligned} \tag{1}$$

where Rules-i stands for the “i-th single-input rule module,” x_i corresponding to the i-th input item is the sole variable of the antecedent part of Rules-i, and y_i is the variable of its consequent part. A_j^i and y_j^i are, respectively, fuzzy set and real number of the jth rule of the Rules-i, where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m_i$, and m_i stands for the number of rules.

Given an input x_i^0 to Rules-i, the compatibility degree of the antecedent part in the j-th rule in Rules-i is given by (2), and the inference result y_i^0 of Rules-i is given as in (3).

$$h_j^i = A_j^i(x_i^0) \tag{2}$$

$$y_i^0 = \frac{\sum_{j=1}^{m_i} h_j^i y_j^i}{\sum_{j=1}^{m_i} h_j^i}. \tag{3}$$

Final inference result y^0 of the SIRMs model with FW is given by

$$y^0 = \sum_{i=1}^{n} w_i(x_i^0) y_i^0 \tag{4}$$

where functional weights $w_i(x_i^0)$'s stand for the importance degree for each input item x_i $(i = 1, 2, \ldots, n)$.

For wxample, the functional weight in case of using a linear expression is as follows [1]:

$$w_i(x_i) = c_i + d_i x_i \tag{5}$$

This SIRMs model is reduced to the conventional SIRMs model when functional weight $w_i(x_i)$ is replaced to a constant.

3 Learning Algorithm

This section presents the learning algorithm for the SIRMs model with FW. The learning algorithm is derived based on the gradient decent concept where the modification rules for the parameters of the SIRMs model is construct so that a pre-defined error is minimized [1].

In this paper, the functional weight in case of using a square function is used as follows:

$$w_i(x_i) = c_i x_i^2 \tag{6}$$

Let us assume that we use triangular-type membership functions for the antecedent parts of fuzzy if-then rules as follows:

$$A_j^i(x_i) = \begin{cases} 1 - |x_i - a_j^i|/b_j^i, & a_j^i - b_j^i \leq x_i \leq a_j^i + b_j^i, \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where a_j^i and b_j^i are the center and the width of the fuzzy set $A_j^i(x_i)$, respectively, and x_i is an input value for the ith attribute. The modification rules for a_j^i, b_j^i, and y_j^i at the $(t+1)$th iteration are defined as follows:

$$\begin{aligned} a_j^i(t+1) = a_j^i(t) + \alpha \cdot (y^T - y^0(t)) \cdot c_i(t) x_i^2 \\ \cdot \frac{y_j^i - y_i^0(t)}{\sum_{j=1}^{m_i} h_j^i(t)} \cdot \frac{\mathrm{sgn}(x_i - a_j^i(t))}{b_j^i(t)}, \end{aligned} \tag{8}$$

$$\begin{aligned} b_j^i(t+1) = b_j^i(t) + \beta \cdot (y^T - y^0(t)) \cdot c_i(t) x_i^2 \\ \cdot \frac{y_j^i - y_i^0(t)}{\sum_{j=1}^{m_i} h_j^i(t)} \cdot \frac{|x_i - a_j^i(t)|}{(b_j^i(t))^2}, \end{aligned} \tag{9}$$

$$y_j^i(t+1) = y_j^i(t) + \gamma \cdot (y^T - y^0(t)) \cdot c_i(t) x_i^2 \cdot \frac{h_j^i(t)}{\sum_{j=1}^{m_i} h_j^i(t)}, \tag{10}$$

$$c_i(t+1) = c_i(t) + \delta \cdot (y^T - y^0(t)) \cdot x_i^2 \cdot y_i^0(t), \tag{11}$$

where α, β, γ, δ and ε are learning rates, t represents the number of learning iterations, and sgn is a signum function that is defined as follows:

$$\mathrm{sgn} = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -1, & \text{otherwise.} \end{cases} \tag{12}$$

4 Identifying Non-linear Functions

This section examines the performance of the proposed SIRMs model for identifying the following non-linear function:

$$y_1 = (2x_1 + 4x_2^2 + 0.1)^2 / 37.21, \tag{13}$$

In the above function, we assume that the domain is $[-1,1]^2$. Thus the range is $[0,1]$.

In the computational experiments of identifying the above functions, membership functions for each attribute are generated so that each attribute is divided into five segments, and each membership function has an intersection with its next one at the membership value of 0.5. The initial value for the importance of each input attribute is specified as 0.5.

49 input-output pairs were randomly generated as training patterns, which are then used for the learning of the SIRMs models. For each experiments, the learning iteration was specified as 1000. In order to examine the performance of the trained SIRMs models, 2601 input-output pairs were generated so that each input attributes are specified with an increment of 0.04. The 2601 input-output pairs were used as test patterns where the output from the trained SIRMs for each input is compared with the corresponding output. The values of learning rates (i.e., α, β, γ, δ, ε) were specified as 0.001, 0.0001, 0.01, 0.01, and 0.1, respectively. These values were specified by trial-and-error. Mean squared error (MSE) is used as a performance measure of the proposed SIRMs in the computational experiments in this paper.

The training process is performed for the a non-linear function y_1 for ten times. The results of the computational experiments are shown in Table 1. In the tables, SIRMs-FW represents the performance of the SIRMs with square functional weights, and SIRMs the conventional model. From the results, we can see that the proposed SIRMs model performs better than the conventional model.

Table 1 Error evaluation for a nonlinear function y_1

Case	SIRMs-FW	SIRMs [4]
1	0.00905	0.0104
2	0.00810	0.0099
3	0.00912	0.0104
4	0.00901	0.0110
5	0.00968	0.0110
6	0.00790	0.0107
7	0.00858	0.0099
8	0.00985	0.0125
9	0.00850	0.0100
10	0.00758	0.0093
Average	0.00874	0.0105

5 Conclusions

In the conventional SIRMs model, the final inference result of the model for an input pattern is calculated as the weighted sum of the inference result from each rule module. Since the weight for each rule module is constant, the SIRMs model cannot be applied to highly dynamic problem domains.

On the other hand, the SIRMs model with FW employes functional weights instead of a constant value. Thus the weight for each rule module is no longer constant but variable according to input values. The learning algorithm for the SIRMs model with FW is also derived using gradient decent method. The learning algorithm enables to automatically construct an appropriate input-output mapping by the SIRMs model with FW from a set of training patterns.

In the computational experiments, a non-linear function has been used in order to examine the performance of the SIRMs model with FW using square functional weight, in this paper. Additive function has been used in the computational experiments. The experimental results have shown that the SIRMs model with square FW performs better than the conventional SIRMs model. Especially, the SIRMs model with square FW has strong nonlinearity. Therefore, this model will be able to applied to involved problems.

The inference results of the SIRMs model with square FW are strongly depended on input. Thus, input values should be normalized to apply the SIRMs model with FW to real systems.

In the well-known simplified fuzzy inference, $\Pi_{i=1}^{M}$ rules are involved in a single system while only $\sum_{i=1}^{M}$ rules in the SIRMs model. Therefore, the SIRMs model is more efficient and more effective approach than the other fuzzy inference models.

Future works include other formulations should be evaluated for functional weights of the SIRMs model with FW, and other nonlinear functions, e.g., multiplicative functions, medical diagnosis, etc., should be applied to this SIRMs model with FW.

References

1. Nakashima, T., Seki, H.: A Study on SIRMs models with linear functional weights. In: Proc. 12th International Symposium on Advanced Intelligent Systems (ISIS 2011), pp. 224–226 (2011)
2. Seki, H., Ishii, H., Mizumoto, M.: On the generalization of single input rule modules connected type fuzy reasoning method. IEEE Trans. Fuzzy Systems 16, 1180–1187 (2008)
3. Seki, H., Ishii, H., Mizumoto, M.: On the monotonicity of fuzzy-inference methods related to T–S inference method. IEEE Trans. Fuzzy Systems 18, 629–634 (2010)
4. Seki, H., Ishii, H., Mizumoto, M.: Nonlinear identification and medical diagnosis system using functional-type SIRMs connected fuzzy inference method. International Journal of Innovative Computing, Information and Control 6, 5275–5286 (2010)
5. Seki, H., Mizumoto, M.: On the equivalence conditions of fuzzy inference methods—part 1: basic concept and definition. IEEE Trans. Fuzzy Systems 19, 1097–1106 (2011)
6. Yi, J., Yubazaki, N., Hirota, K.: Stabilization control of seriestype double inverted pendulum systems using the SIRMs dynamically connected fuzzy inference model. Artificial Intelligence in Engineering 15, 297–308 (2001)
7. Yi, J., Yubazaki, N., Hirota, K.: Upswing and stabilization control of inverted pendulum system based on the SIRMs dynamically connected fuzzy inference model. Fuzzy Sets and Systems 122, 139–152 (2001)
8. Yi, J., Yubazaki, N., Hirota, K.: A new fuzzy controller for stabilization of parallel-type double inverted pendulum system. Fuzzy Sets and Systems 126, 105–119 (2002)
9. Yi, J., Yubazaki, N., Hirota, K.: A proposal of SIRMs dynamically connected fuzzy inference model for plural input fuzzy control. Fuzzy Sets and Systems 125, 79–92 (2002)
10. Yi, J., Yubazaki, N., Hirota, K.: Anti-swing and positioning control of overhead traveling crane. Inf. Sci. 155, 19–42 (2002)
11. Yu, W., Li, X.: Fuzzy identification using fuzzy neural networks with stable learning algorithms. IEEE Trans. Fuzzy Systems 12, 411–420 (2004)
12. Yubazaki, N., Yi, J., Hirota, K.: SIRMs (Single Input Rule Modules) connected fuzzy inference model. Journal of Advanced Computational Intelligence and Intelligent Informatics 1, 23–30 (1997)

Stepwise Benchmarking Path Selection in DEA

Jaehun Park, Hyerim Bae, and Sungmook Lim

Abstract. In order to identify the best performers for the benchmarking process, Data Envelopment Analysis (DEA), a methodology for measuring the relative efficiencies among homogeneous Decision-Making Units (DMUs) and for yielding a reference target for and inefficient DMU along with the corresponding efficiency gap, has been used. The general use of DEA has certain limitations that it might not be feasible for an inefficient DMU to achieve its benchmarking target in a single step. In order to overcome these problems, stepwise benchmarking target selection method has been undertaken. However, it did not consider benchmarking target selecting in optimal aspects. Thus, we propose an optimal benchmarking target selection method that inefficient DMU can select benchmarking target considering multiple benchmarking objectives using DEA. For the benchmarking objectives, the improvement feasibility and the efficiency gap are considered.

Keywords: Data Envelopment Analysis, Benchmarking target selection.

1 Introduction

Generally, identifying the best performers can be considered to be the most importance activity in the benchmarking process. In order to identify the best performers for the benchmarking process, Data Envelopment Analysis (DEA), a methodology for measuring the relative efficiencies among homogeneous Decision-Making Units (DMUs) and for yielding a reference target for and inefficient DMU along with the corresponding efficiency gap, has been widely used (Ross and Droge 2002).

However, the general use of DEA has certain limitations in the aspects of selecting the proper performers of inefficient organization. One of the problems to

Jaehun Park · Hyerim Bae
Business & Service Computing Lab., Industrial Engineering, Pusan National University, 30-san Jangjeon-dong Geumjong-gu, Busan 609-735, South Korea
e-mail: pjh3479@pusan.ac.kr, hrbae@pusan.ac.kr

Sungmook Lim
Division of Business Administration, College of Business and Economics, Korea University, Jochiwon, Yeongigun, Chungnam 339-700, South Korea
e-mail: sungmook@korea.ac.kr

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 477–484.
springerlink.com

be discussed in this research is that it might not be feasible for an inefficient DMU to achieve its benchmarking target's efficiency in a single step, especially when that DMU is far from the benchmarking target DMU. In order to overcome this limitation, Zhu (2003) proposed a stratification method by iteratively generating the efficient frontiers. Alirezaee and Afsharian (2007) proposed a layered efficiency evaluation model that provides a strategy by which an inefficient DMU can move toward a better layer. Lim *et al.* (2011) proposed a stratification benchmarking path method that can select next benchmark DMU by applying the stratification method in a context-dependent DEA.

The existing works addressed here can be considered to more realistic and effective than general DEA in aspects of providing the benchmarking targets of inefficient DMU in phase. In most of previous works, they did not consider stepwise benchmarking target in optimal aspects. Thus, this paper proposes an optimal stepwise benchmarking target selection method in DEA considering multiple benchmarking objectives. In case of benchmarking of multiple objectives, the resource improvement and the improvement direction proximity are considered. The proposed method is different from previous researches in that it suggests the benchmarking target sequentially to find the optimal aspect by considering two criteria (the resource improvement and the improvement direction proximity). This study conducts a real case experiment by benchmarking on East Asia container terminal as an application of the proposed method. The structure of this paper is organized as follows. Section 2 provides an overview of DEA for the efficiency measurement and benchmarking. Section 3 discusses the proposed method, and section 4 details our empirical study. Finally Section 5 summarizes our work.

2 Related Work

2.1 Data Envelopment Analysis (DEA)

Data Envelopment Analysis (DEA) is a linear-programming methodology that evaluates the relative efficiencies of DMUs using a set of inputs to produce a set of outputs (Zhu, 2003). The mathematical model of DEA is represented by model (1) and (2). The model (1) is a basic CCR model of DEA initially developed by Charnes, Cooper & Rhodes (1978). Here, u_r is the weight given to the r-th output, v_i is the weight given to the i-th input, n is the number of DMUs, s is the number of outputs, m is the number of inputs, k is the DMU being measured, y_{rj} is the amount of the r-th output produced by DMU j, and x_{ij} is the amount of the i-th input produced by DMU j.

DEA can suggest a reference set, which is a set of corresponding efficient units that can be utilized as a benchmark for improvement. The reference set can be obtained by dual model, as shown in (2). In model (2), θ is the efficiency score, λ_j is the dual variable, and ε is a non-Archimedean infinitesimal. By solving model (3), we can identify a composite DMU (a linear combination of DMUs) that utilizes less input than the test DMU while maintaining at least the same output levels. The optimal values of the dual variable λ_j are the coefficients for this linear combination of units. If a DMU is given an efficiency score of '1', it is considered to be efficient; an efficiency score less than '1' indicates inefficiency. For more details on model development, reader may refer to Charnes *et al.* (1978).

$$
\begin{aligned}
&Max \sum_{r=1}^{s} u_r y_{rk} \\
&s.t. \sum_{i=1}^{m} v_i x_{ik} = 1 \\
&\quad \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0; \ j = 1,\ldots,n \\
&\quad u_r, v_i \ge \varepsilon; \ r = 1,\ldots,s; \ i = 1,\ldots,m
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
&Min \ \theta \\
&s.t. \sum_{j=1}^{n} \lambda_j x_{ij} - \theta x_{ik} \le 0; \ i = 1,2,\ldots,m, \\
&\quad \sum_{j=1}^{n} \lambda_j y_{rj} - y_{rk} \ge 0; \ r = 1,2,\ldots,s, \\
&\quad \lambda_j \ge 0; \qquad j = 1,2,\ldots,n
\end{aligned}
\tag{2}
$$

3 Proposed Method

The step of the proposed method is as follows; first, all DMUs are stratified based on efficiency score to form efficient layers by applying stratification method in context-dependent DEA. The stratified layers represent the benchmarking phase of inefficient DMU that wants to improve the efficiency (Hereafter this DMU is called as evaluated DMU). Second, the benchmarking path network is constructed based on benchmarking candidate DMUs. The benchmarking candidate DMUs represent the reference DMUs, which are on each layer, of evaluated DMU. Third, the resource improvement and improvement direction proximity between each benchmarking candidate DMU are measured for the benchmarking objectives. Finally, optimal benchmarking path is searched based on benchmarking objectives using mathematical model.

3.1 Stratification of DMUs

All DMUs are stratified based on efficiency scores for the stepwise benchmarking target selection. The stratification DEA model in context-dependent DEA, which proposed by Seiford and Zhu (2003), is used to stratify DMUs.

Table 1 Supermarket Example

Store		A	B	C	D	E	F	G	H	I	J	K	L
Employee	x_1	2	4	8	3	4	5	5	6	7	6	6	7
Floor area	x_2	4	2	1	6	3	2	6	3	3	9	4	7
Sales	y	1	1	1	1	1	1	1	1	1	1	1	1

Let's consider the following supermarket example, originally introduced in Copper *et al.* (2006) and more DMUs are added for easy explanation of the procedure outlined in this paper. Table 1 consists of twelve DMUs, and each DMU consumes two inputs and yields one output.

When we apply stratification DEA method to supermarket example, we can obtain the five layers as illustrated in Figure 1. Let DMU L be an evaluated DMU, and DMU L can improve its efficiency by crossing the sequence of layers.

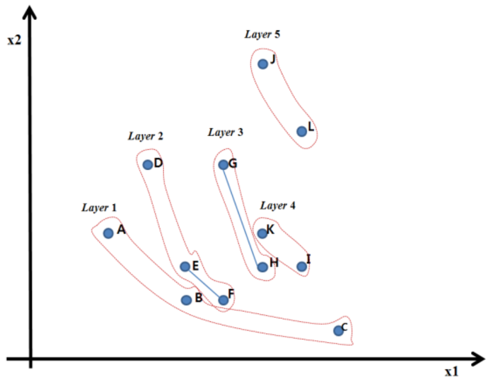


Fig. 1 Benchmarking target of stratification method

3.2 Construct Benchmarking Path Network

In the figure 1, DMU L can obtain its reference DMUs on each layer by model (2). For example, the reference DMU on layer 4 from DMU L is DMU K. The next step is a replacement of DMU K by the evaluated DMU. The reference DMUs from each evaluated DMU are shown in figure 2.

These reference DMUs can be the benchmarking candidate DMUs of evaluated DMU. Based on these benchmarking candidate DMUs, the benchmarking path network is constructed. The benchmarking path network from DMU L is shown in figure 3.

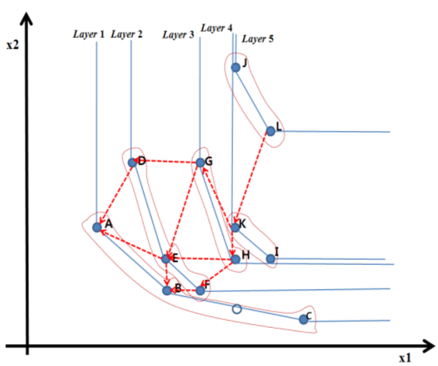


Fig. 2 The reference DMUs from each evaluated DMU

If DMU L selects DMU B as its ultimate benchmarking DMU, then DMU L have three alternatives to reach DMU B in phase (DMU L→K→H→F→B, DMU L→K→H→E→B and DMU L→K→G→E→B). Among these alternatives, optimal benchmarking path that satisfies the benchmarking objectives can be searched.

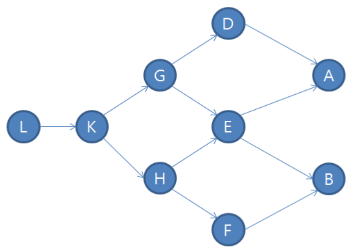


Fig. 3 Benchmarking path network from DMU L

3.3 *Measure the Benchmarking Objectives*

For the benchmarking objective, two objectives are considered in this paper; minimize the resource improvement and maximize the improvement direction proximity. Both definitions are as follows;

Definition 1. Resource improvement
The resource improvement is the improvement gap of inputs or outputs between each benchmarking candidate DMU. The resource improvement can be calculated by $G_i = x_{ik} - \alpha_i^* x_{ik}$, where k is the index of evaluated DMU and j is the benchmarking candidate DMUs on next layers from evaluated DMU. α_i^* is yielded by model (3).

$$\begin{aligned} & min \sum_{i=1}^{m} \alpha_i \\ & s.t. \\ & x_{ij}\lambda_j \leq x_{ik}\alpha_i, \quad i = 1,\ldots,m \\ & y_{rj}\lambda_j \geq y_{rk}, \quad r = 1,\ldots,s, \quad \lambda_j \geq 0 \end{aligned} \tag{3}$$

Definition 2. Improvement direction proximity
Improvement direction proximity, which proposed by Park *et al.* (2010), is employed to select the closest efficient DMU for the next step of benchmarking DMU. The formulation for identifying the improvement direction proximity of the j-th DMU (δ_j) can be represented by model (4).

$$\delta_j = 1 - \left[(\cos^{-1} \frac{\sum_{r=1}^{m} g_r h_r}{\sqrt{\sum_{r=1}^{m} g_r^2} \sqrt{\sum_{r=1}^{m} h_r^2}}) / \hat{\delta} \right] \tag{4}$$

where

g_r: Difference of the r-th input between evaluated DMU and ultimate DMU
h_r: Difference of the r-th input factors between evaluated DMU and j-th DMU
$\hat{\delta}$: User-defined threshold value (It can be defined between 90 and 180)

δ_j indicates a relative inner product between g_r and h_r. DMU j, which has the maximum δ_j, means that it is closest to the direction from the evaluated DMU to the ultimate benchmarking target.

3.4 Search the Optimal Benchmarking Path

In order to select optimal benchmarking path considering the resource improvement and the improvement direction proximity as benchmarking objectives, multi-objective model (5) is used. We define the variable $x_{jj'}$ for the $l_{jj'}$, where $l_{jj'}$ is the link between j-th DMU and j'-th DMU: if j-th DMU and j'-th DMU in $l_{jj'}$ are possible to benchmark each other, then $x_{jj'}$ is defined as 1, otherwise $x_{jj'}$ is defined as 0.

$$\begin{aligned}
Sp = \max \; & \sum_{j=1}^{J} \sum_{j'=1}^{J} \delta_{jj'} x_{jj'} \\
\min \; & \sum_{j=1}^{J} \sum_{j'=1}^{J} \sum_{i=1}^{m} G_{ijj''} \\
s.t. \; & \sum_{j'=1}^{J} x_{jj'} - \sum_{k=1}^{J} x_{kj} = \left\{ \begin{array}{ll} 1 & (j=1) \\ 0 & (j=2,3,\ldots,J-1) \\ -1 & (j=J) \end{array} \right\} \\
& x_{jj'} = 0 \;\; or \;\; 1 \;\; (j', j = 1,2,\ldots,J)
\end{aligned} \tag{5}$$

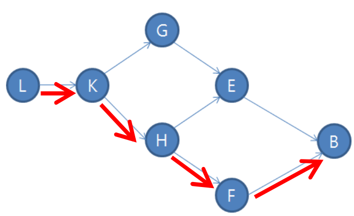


Fig. 4 Optimal stepwise benchmarking path from DMU L

If Model (5) is applied to Figure 3 under DMU B selected as ultimate DMU of DMU L, the optimal benchmarking path can be obtained as shown in Figure 4.

4 Case Study

In this study, we applied our proposed method to the real container terminal data to select optimal stepwise benchmarking path of the inefficient container terminals. The relevant data sources have been collected for 22 East Asia container terminals from *Containerization International Year Book* 2005-2007, excluding incomplete data. The numbers of berths, the lengths of berths (m), the total area of the port (km^2) and the number of cranes are used as inputs, while the total container traffic (TEU) are used as outputs for the efficiency evaluation of container terminals.

Six layers were stratified using stratification DEA method. Hongkong, Shanghai, Shenzhen and Xiamen are included as set of port in layer 1, and Kwangyang and Kobe are included as set of port in layer 6. Kwangyang and Shenzhen are selected as evaluated DMU and ultimate benchmarking DMU, respectively. The benchmarking path of Kwangyang is shown in Figure 5. Kwangyang select Osaka as first stepwise benchmarking target, and select Dalina, Keelung, Busan and finally Shenzhen, sequentially.

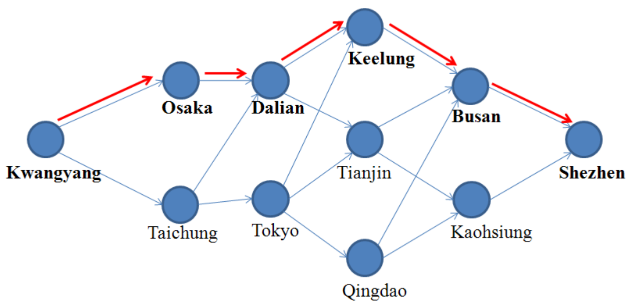


Fig. 5 Optimal stepwise benchmarking path from Kwangyang

5 Conclusions

In this paper, we proposed an optimal stepwise benchmarking target selection method in DEA considering multiple benchmarking objectives. This is a new method to remedy the drawback of the existing stepwise benchmarking, which is no consideration on optimal benchmarking target selection with multiple benchmarking objectives. For the multiple benchmarking objectives, the resource improvement and the improvement direction proximity were considered. As an application of the proposed method, the proposed method is applied to find the optimal stepwise benchmarking path for the 22 East Asian container terminal. The proposed method is different from previous researches in that the proposed method suggested the optimal stepwise benchmarking target by considering two objectives (minimize the resource improvement and maximize the improvement direction proximity).

Acknowledgement. "This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No.2010-0027309)".

References

Alirezaee, M.R., Afsharian, M.: Model improvement for computational difficulties of DEA technique in the presence of special DMUs. Applied Mathematics and Computation 186, 1600–1611 (2007)

Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision making units. European Journal of Operational Research 2, 429–444 (1978)

Lim, S., Bae, H., Lee, L.H.: A study on the selection of benchmarking paths in DEA. Expert System with Applications 38, 7665–7673 (2011)

Ross, A., Droge, C.: An integrated benchmarking approach to distribution center performance using DEA modeling. Journal of Operations Management 20, 19–32 (2002)

Zhu, J.: Quantitative models for performance evaluation and benchmarking-Data Envelopment Analysis with Spreadsheets and DEA Excel Solver. Kluwer Academic Publishers (2003)

The Automatic Two – Step Vessel Lumen Segmentation Algorithm for Carotid Bifurcation Analysis during Perfusion Examination

Marek R. Ogiela and Tomasz Hachaj

Abstract. The novel path detection algorithm and the region growing based lumen detection algorithm is the main contribution of this article. The proposed lumen segmentation method is consisted of two sub-algorithms. After preprocessing step the first algorithm detects the possible path between start and end point. In the second step it performs the thinning of previously obtained path. The second algorithm is a region – growing procedure with proper homogeneity criteria. The role of this procedure is to segment the whole lumen of considered vessel. The region growing is computed in axial slices and the seed point for growing method in each slice is a voxel taken from path computed in previous step. The proposed method was tested on six carotid arteries structures obtained from CTA examination. Our researches has shown that with our method it is possible to keep the TPR (true positive rate) in 5 from 6 considered cases on the level higher than 80% with FPR (false positive rate) below 1%.

Keywords: Lumen segmentation, carotid bifurcation, computed tomography angiography, brain perfusion maps, computer - aided diagnosis.

1 Introduction

Computed tomography angiography (CTA) is a popular medical imaging method that is often used beside standard computed tomography (CT) in acute stroke imaging [9, 10]. Imaging of the carotid arteries is important for the evaluation of

Marek R. Ogiela
AGH University of Science and Technology 30 Mickiewicza Ave, 30-059 Krakow, Poland
e-mail: mogiela@agh.edu.pl

Tomasz Hachaj
Pedagogical University of Krakow, Institute of Computer Science and Computer Methods, 2 Podchorazych Ave, 30-084 Krakow, Poland
e-mail: tomekhachaj@o2.pl

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 485–493.
springerlink.com

patients with ischemic stroke or Transient Ischemic Attack (TIA). There are many computer – aided methods that support the radiologist during the procedure of segmentation of carotid bifurcation region. The methods that have been already proposed can be divided in two groups: model – based and intensity – based algorithms. The first group takes advantage of geometric specificity of vessels, in particular the notions of orientation and tubular shape. One of possible vessels segmentation approach is utilizing tube detection filters based on analysis of volumetric image Hessian matrix eigenvalues [7]. Those methods are capable to detect any local tubular structure but are sensitive to noises and scanning artifacts. The second group is dedicated mainly for detailed extraction of continues tubular structures. Many intensity – based methods are two – step procedures. They computes the optimal path (according to given criteria) between selected points that lies inside the examined vessels and in the second step detects the vessel lumen locally for each voxel of obtained path (it is often done using a 3D active surface approach with various forces models like in [8]).

The problem of finding the path that links two points inside the lumen is commonly solved similarly to minimal path finding task. The solution may be found by energy minimizing approach (for example with fast – marching algorithm [3]) or graph – based approach (for example with Dijkstra's shortest-path algorithm [4]). The energy minimizing approach requires complex definition of energy in each voxel while graph methods need to compute the weight of edges (cost of path) between two neighbor nodes. The second approach is more intuitive and closer to the process of manual determination of shortest path in which software operator build the path by adding the voxel that differs the least from the last added. The definition of edges weight in Dijkstra's – like algorithms is also not an easy task. It should takes into account not only difference of neighbors tissues densities but also length of the path and difference between densities of each voxel in path and beginning end ending points (if the difference of densities along the path is to big it is possible that the path was found outside the vessel lumen) [4]. In order to satisfies those conditions we proposed the new path finding algorithm. The novel path detection algorithm and the region growing based lumen detection algorithm is the main contribution of this article. The proposed method was tested on six carotid arteries structures obtained from CTA examination. The Common Carotid Artery (CCA) and Internal Carotid Artery (ICA) are clinically the most relevant arteries of the Carotid Bifurcation [5]. Therefore the segmentation technique proposed in this publication is optimized for detection of those particular blood vessels.

We also generated ROC curve in order to find optimal threshold value for homogeneity criteria in region growing schema. The results of the automatic segmentation were compared to manual findings of radiologist.

The results presented in this article are extension of our previous work. Our latest researches were concentrated on automatic analysis of dynamic perfusion computed tomography maps (CTP) in the event of brain stroke. In [12] we proposed a new approach of description and analysis of potential lesions in cerebral blood flow (CBF) and cerebral blood volume (CBV) perfusion maps. In [13] we extended the possibility of automatic analysis on another type of perfusion maps (time to peak maps - TTP). In [14] we proposed the method for generation of prognostic maps that visualizes the possible evaluation of ischemia.

We decided to widen the area of our interest on CTA because the examination of carotid arteries is an important step during assessing the risk of brain stroke [6].

2 Methods

The proposed lumen segmentation method is consisted of two sub-algorithms. After preprocessing step the first algorithm detects the possible path between the start and the end point. In the second step it performs the thinning of previously obtained path. The generated path between the start and point becomes 1 voxel width keeping the same length as path from first step. The second algorithm is a region – growing procedure [2] with proper homogeneity criteria. The role of this procedure is to segment the whole lumen of considered vessel. The region growing is computed in axial slices and the seed point for growing method in each slice is a voxel taken from path from Algorithm I.

Description of symbols used in algorithms descriptions:

Freezed points:= $\varnothing$ *– already visited points.*
Narrow band$_i$*:=* $\varnothing$ *– points, that are visited in i-th step.*
Start point – starting point of the path. $\varnothing$
End point – end point od the path.
Delta value:=0 - maximal accepted difference between neighbor points.
$S(x_j)$ *- Surranding of point* x_j *with radius 1 (26 voxels).*
$V(x_j)$ *- Value of vexel density in point* x_j*.*
Path lenght - the length of the path (in voxels).
Delta value - maximal accepted difference between two voxel densities. If the difference is greater than Delta value, the considered voxel is not included into the path.

In the preprocessing step the volumetric image is convoluted with Gaussian kernel in order to remove noises and scanning artifacts. The image is then thresholded in order to remove voxels that density do not belongs to range:

[min(V(Start point), V(End point)) – 40, max(V(Start point), V(End point)) + 200]

That step eliminates the uncontrolled propagation of path detection algorithm in regions where tissues has to low or to high density to be part of examined vessel.

Algorithm I, step I – detection of path between *Start point* and *End point*.
Delta value:=-1
While (End point$\notin$*Freezed points)*
  Delta value:=Delta value+1
  i:=0
  Narrow band:= $\varnothing$
  Narrow band$_0$*:={Start point}*
  Freezed point:={(Start point, I)}

While (#Narrow band$_i$>0 $\wedge$ End point $\notin$Freezed points)
i:=i+1
$\forall x_j \in$*Narrow band$_i$*
$\forall y_k \in S(x_j)$
if(/V(x_j)- V(y_k)/<Delta value)
Narrow band$_i$:= Narrow band$_i \cup y_k$
Freezed point:= Freezed point$_i \cup (y_{k,} i)$_
Path length:=i
End algorithm I, step I

Because the algorithm stops immediately after detecting the end point it will not generat paths that are too long and it is unnecessary to add any penalization term of path length to the edges weighting function.

Algorithm I, step II – thinning of path obtained in step I. The path generated by in second step has the same length as previous path but is only one voxel width.

Path:= $\emptyset$ – path from End point to Start point
Path:=Path$\cup$End point
k:=Path length -1
i:=0
x_i:=End point
While(Start point $\notin$ Path)
$X_{i+1}:=(y_j: /V(y_j)-V(x_i)/=\min/V(y_l)-V(x_i)/, (y_l,k) \in \text{Freezed points}, y_l \in S(x_i))$
Path:= Path$\cup x_{i+1}$
i:=i+1
k:=k-1
End Algorithm I, step II

The *T* parameter in following algorithm is maximal accepted difference between two voxel densities. If the difference is greater than it, the considered voxel is not included into segmented region.

Algorithm II – the homogeneity criteria for region growing algorithm

Homogeneity criteria is a function of parameter T

The h*omogeneity criteria for voxel x_j* and voxel x_i which already belongs to the region is a function of threshold value *T* and is satisfies if and only if:

1) *Absolute value of difference of ISO values between voxel x_i and x_j is smaller or equal T.*
AND
2) *In the surrounding of voxel x_j $S(x_j)$* exists more than 9 voxels x_k for which a*bsolute value of difference of ISO values between voxel x_j and x_k is smaller or equal T.*

$x_j \in X \Leftrightarrow /V(x_i)- V(x_j)/<=T \wedge x_j \in S(x_i) \wedge x_k \in S(x_j) \wedge \#\{x_k: /V(x_j)- V(x_k)/<=T\}>9$
End Homogeneity criteria

3 Results

The proposed algorithm was tested on set of three CTA volumes of carotid artery with size 512x512x415 voxels, 512x512x425 voxels and 512x512x432 voxels scanned by SOMATOM Sensation 10 CT scanner. The distance between axial slices was 0.7 mm. The segmentation was performed on left and right carotid artery separately. Because of that the experimental set was consisted of six tubular structures.

The volume to be segmented was determined similarly as in comparison protocol in [5]. It is defined around the bifurcation slice, which was marked as the first (caudal to cranial) slice where the lumen of the CCA appears as two separate lumens: the lumen of the ICA and the lumen of the ECA (external carotid artery). The segmentation contain the CCA, starting at least 20 mm caudal of bifurcation slice, the ICA, up to at least 40 mm cranial of bifurcation slice, and the ECA, up to between 10 and 20 mm cranial of the bifurcation slice. The segmentation was performed at first on CCA – ICA section, than on CCA – ECA, the final segmentation result is the common part of those two.

The segmentation algorithm was tested with different values of *T* parameter in homogeneity criteria. Figure 1 demonstrates the influence of *T* on completeness of detected lumen. Figure 2 shows the close-up of wrongly segmented carotid bifurcation region (overestimation of vessel lumen). In Figure 3 the segmentation results of all considered cases is presented. The threshold parameter *T* is set to 12.

Receiver Operating Characteristics (ROC) graphs [1] are a useful technique for organizing classifiers and visualizing their performance. It is possible to produce

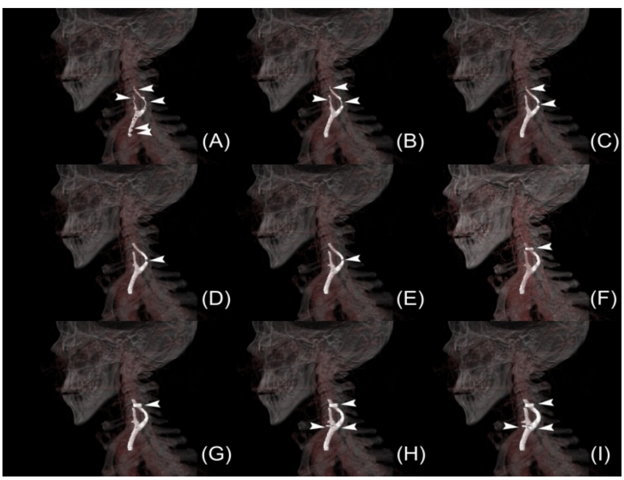


Fig. 1 Results of segmentation with various values of parameter *T* in Homogeneity Criteria. Value of T is appropriately: (A) - 4, (B) - 6, (C) - 8, (D) - 10, (E) - 12, (F) - 14, (G) - 16, (H) - 18, (I) – 50. Segmentation errors are marked with white arrows.

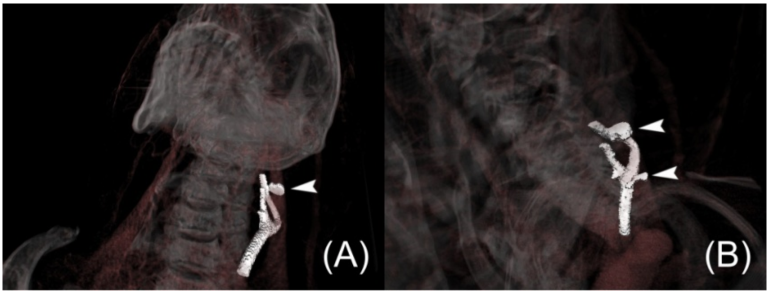


Fig. 2 Detailed view of segmentation algorithm errors. The parts of jugular vein are included to final segmentation results because of too high value of parameter *T* in homogeneity criteria. (A) T equals 14, (B) T equals 18. Segmentation errors are marked with white arrows.

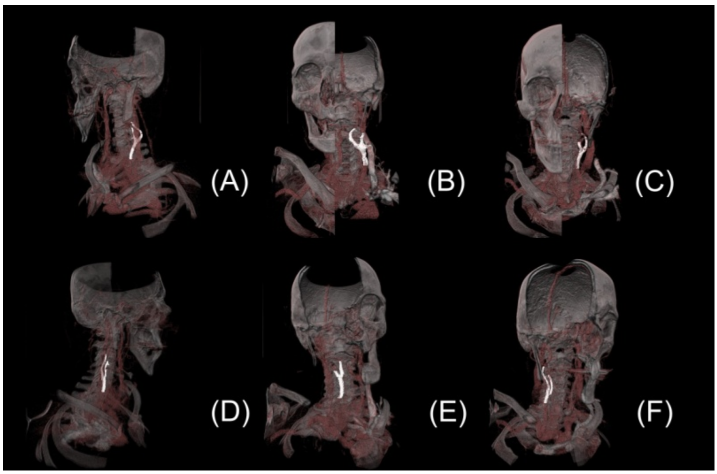


Fig. 3 Segmented carotid arteries obtained from all analyzed volumes (marked in white). The value of T in Homogeneity Criteria is 12. (A), (B), (C) – left carotid artery, (D), (E), (F) – right carotid artery.

ROC curve by performing classification with binary classifier with different threshold value. Each threshold value produces a different point in ROC space. That kind of plot can also be used to find optimal value of classifiers threshold. It is a value that corresponds to the point on ROC that maximize true positive rate (TPR) and minimize false positive rate (FPR) – the closest point to (0,1) point in ROC space. The results of the automatic segmentation were compared to manual findings of radiologist.

The proportion of voxels that belongs to arteries of interest to all other voxels (background) is like $1 \sim 10^5$. Because of that the FPR comparing to TPR is very

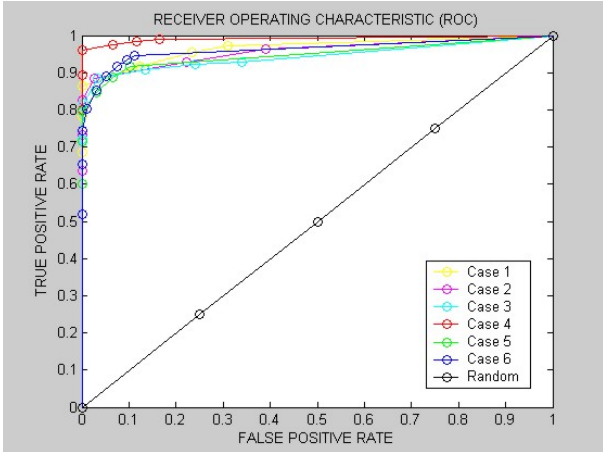


Fig. 4 The plot of receiver operating characteristic

small. That does not reflect the real impact of false positive detection (overestimation of lumen size) on quality of segmentation. In order to make ROC curve more informative we multiplied the FPR by 10^4 factor to give both parameters the same scale. The plot with scaled FPR is presented in Figure 4.

4 Discussion

In all examined cases from the test set the path detection algorithm rightly detect the curve that links the start and end point (the path lies inside the carotid lumen). The value of parameter *T* highly affects the quality of segmentation. If the threshold is to low many incompleteness (false negative errors) are introduced (Figure 1 A – E, marked by white arrows). If the *T* exceed the certain value the overestimation of lumen is more common (false positive errors, Figure 1 F - I). The most common overestimation that happens is introducing the parts of jugular vein to segmented carotid arteries (Figure 2). That is because those structures lie very close to region of interests and may have similar tissue density. The determination of optimal *T* value for all considered cases is not an easy task. Our researches has shown that with $T = 12$ it is possible to keep the TPR in 5 from 6 considered cases on the level higher than 80% with FPR below 1% (Figure 4). That threshold level gives smooth and easy to interpreted vessels segments not disturbed by extensive structure (Figure 3). The threshold on level $T = 12$ also mach in most considered cases the optimal point on ROC curve (the closest point to (0,1) point in ROC space). Because of that we decided to choose that level as the baseline for the further future research.

5 Conclusions

The new segmentation algorithm proposed in this article is a convenient and easy to use method that might be applied not only for carotid arteries but also for any type o intensity – based segmentation of continues structures. The principles of the algorithm can be easily described what is especially valuable for the methods that are introduced to medical society. The initialization procedure is very simple (only two points per continues vessel are required to start the segmentation). The algorithm can also be adapted for segmentation task of longer continues parts of vascular system simply by adding more initial points or by modifying the adaptive parameter T of Homogeneity Criteria. The initial validation results looks very promising. They have shown the possibility of obtaining high TPR value and in the same time keeping FPR on low level (Figure 4). In order to make reliable comparison to other contemporary algorithms the intensive researches on bigger data sets are require. On that moment the results we get states our algorithm among of the best state of the art vessels segmentation algorithms [16].

The our method has also some drawbacks. A homogeneity criterion does not allow segmenting very narrow vessel structures (with continues parts with diameter of 2 voxels or less). Also increasing of the tolerance value of homogeneity criteria may cause improper growth of the segmentation area. This situation can be observed in Figure 2 where jugular vein was segmented as a part of examined artery.

Our goal for the future is validation of proposed algorithm on larger dataset and if the result of evaluation will be on acceptable rate we will use this algorithm as the baseline for the further researches on automatic diagnosis of carotid structures. In order to accomplish this task we are planning to create appropriate semantic description of carotid artery similarly to those proposed in [11, 15]. After correct identification of possible lumen abnormality we will try to integrate the results with already developed by us CTP diagnosis framework. That approach will allow us to create more complex and complete diagnostic records that might be very helpful for radiologist in decision - making process.

Acknowledgments. This work has been supported by the Ministry of Science and Higher Education, Republic of Poland, under project number N N516 511939.

References

1. Fawcett, T., Graphs, R.O.C.: Notes and Practical Considerations for Researchers, vol. 31, pp. 1–38, HPL-2003-4. Citeseer (2003-2004)
2. Lu, Y., Jiang, T., Zanga, Y.: Region growing method for the analysis of functional MRI data. NeuroImage 20, 455–465 (2003)
3. Krissian, K., Garcia, S.A.: A Minimal Cost Path and Level Set Evolution Approach for Carotid Bifurcation Segmentation. In: Proc. of MICCAI 2009 Workshop: Carotid Lumen Segmentation and Stenosis Grading Challenge (2009)

4. Freiman, M., Frank, J., Weizman, L., Nammer, E., Shilon, O., Joskowicz, L., Sosna, J.: Nearly automatic vessels segmentation using graph-based energy minimization. In: Proc. 3rd 3D Segmentation in the Clinic: Carotid Lumen segmentation challenge workshop, 12th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2009, London, UK (2009)
5. Hameeteman, R., Zuluaga, M., Joskowicz, L., Freiman, M., Van Walsum, T.: Carotid lumen segmentation and stenosis grading challenge, 4.2, 5.1 (May 2009), `http://cls2009.bigr.nl/download/evaluationframework2009.pdf`
6. Steinman, D.A., Poepping, T.L., Tambasco, M., Rankin, R.N., Holdsworth, D.W.: Flow Patterns at the Stenosed Carotid Bifurcation: Effect of Concentric versus Eccentric Stenosis. Annals of Biomedical Engineering 28(4), 415–423, doi:10.1114/1.279
7. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale Vessel Enhancement Filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
8. Cuisenaire, O., Virmani, S., Olszewski, M.E., Ardon, R.: Fully automated segmentation of carotid and vertebral arteries from contrast enhanced CTA. In: Proceedings of SPIE (2008), vol. 6914, pp. 69143R-69143R-8. Spie (2008)
9. Scharf, J., et al.: Improvement of Sensitivity and Interrater Reliability to Detect Acute Stroke by Dynamic Perfusion Computed Tomography and Computed Tomography Angiography. Journal of Computer Assisted Tomography 30(1), 105–110 (2006)
10. Josephson, S.A., Bryant, S.O., Mak, H.K., Johnston, S.C., Dillon, W.P., Smith, W.S.: Evaluation of carotid stenosis using CT angiography in the initial evaluation of stroke and TIA. Neurology 63(3), 457–460 (2004)
11. Trzupek, M., Ogiela, M.R., Tadeusiewicz, R.: Intelligent image content semantic description for cardiac 3D visualisations. Engineering Applications of Artificial Intelligence 24, 1410–1418 (2011)
12. Hachaj, T., Ogiela, M.R.: Automatic detection and lesion description in cerebral blood flow and cerebral blood volume perfusion maps. Journal of Signal Processing Systems for Signal, Image, and Video Technology 61(3), 317–328 (2010)
13. Hachaj, T., Ogiela, M.R.: A system for detecting and describing pathological changes using dynamic perfusion computer tomography brain maps. Computers in Biology and Medicine 41, 402–410 (2011)
14. Hachaj, T., Ogiela, M.R.: CAD system for automatic analysis of CT perfusion maps. Opto-Electronics Review 19(1), 95–103 (2011), doi:10.2478/s11772-010-0071-2
15. Trzupek, M.: Semantic Modelling of Coronary Vessel Structures in Computer Aided Detection of Pathological Changes. In: Tjoa, A.M., Quirchmayr, G., You, I., Xu, L. (eds.) ARES 2011. LNCS, vol. 6908, pp. 220–227. Springer, Heidelberg (2011)
16. Hameeteman, K., et al.: Evaluation framework for carotid bifurcation lumen segmentation and stenosis grading. Medical Image Analysis 15(4), 477–488 (2011)

The Benefit of Cooperation between Retailers under Supply Uncertainties

Jing Hou and Youquan Sun

Abstract. The focus of this paper is placed on evaluating the benefit of cooperation for two retailers with the same product who face suppliers with yield uncertainties. Two cooperation strategies are analyzed: one is inventory pooling and the other is centralized purchasing. The retailers' expected profit functions under the two cooperation strategies with stochastic deliveries are first obtained, and then the optimal order quantities are identified and compared. Finally, the sensitivity of the benefits of cooperation to various input factors is examined through numerical examples. The study contributes to the literature by providing a better understanding of risk sharing and by shedding insights on the value of cooperation under supply risks.

1 Introduction

In many industries, the output quantity of a production process is not deterministic and is influenced by many factors[1], such as capacity uncertainty, weather and working conditions. Companies that face supply uncertainty has to find ways of risk transferring or sharing to minimize the possible lost such as larger penalty cost and lower customer service level due to unsatisfied demand. It is well adopted by various industries, for example, IT industry, food industry and energy supply chain, that multiple sourcing, backup sourcing and inventory pooling are effective strategies to mitigate supply risks. Both multiple sourcing and backup sourcing intend to use the order from other suppliers as a buffer against uncertainty of one supplier. While sometimes, it is not easy to find other appropriate supplier but could cooperate with other retailers who sell same product by supply risk sharing.

Inventory pooling has been proved to be one of the efficient ways for the buyers to mitigate unexpected risks and also one common way in real world. It is

Jing Hou
Business School, Hohai University, Nanjing, Jiangsu 221000, China
e-mail: penguinhj@163.com

Youquan Sun
Petrochina Dalian Marketing Company, Dalian, Liaoning 116021, China
e-mail: sunyq99@petrochina.com.cn

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 495–504.
springerlink.com

common for retailer who has multiple stores selling the same product to use inventory pooling against demand uncertainty. For two or more retailers, if they face supply uncertainty from different suppliers of the same product, inventory pooling can work efficiently as a way of risk pooling and sharing.

Such inventory stocking points decouple the supply chain so that the uncertainties of supply can be shielded[2]. We consider two kinds of risk sharing strategies: inventory pooling and centralized purchasing. Under centralized purchasing, the retailers not only pool their inventory, but also make purchasing decisions together. Orders are placed to their suppliers to maximize the total profit, and deliveries are stocked in the same place before transported to each retailer's store. It can be seen that centralized purchasing is a format of inventory pooling of closer cooperation. And centralized is also implemented by many large companies to deal with various risks. In what follows, we will use inventory pooling and centralized purchasing to differentiate the two cooperation strategies.

Our paper tries to examine the benefit of cooperation between two retailers facing supply uncertainty risks. As will be discussed in the following literature review, the problem studied in this paper has not been fully addressed in the literature. We attempt to figure out how exactly the inventory pooling and centralized purchasing could mitigate the supply risks by comparing the two retailers' order decisions and expected profits. Specifically, the profit functions of the retailers are first obtained, and then the optimal decisions are derived for each cooperation strategies. Finally, the effects of the two cooperation strategies are examined numerically with different input parameters.

Supply uncertainty or yield uncertainty in supply chain context has been the subject of many researches recently, while most of them focus on the coordination between suppliers and buyers. For example, Güler and Bilgiç (2009) propose two contracts to coordinate an assembly system under random yield and random demand[3]; He and Zhang (2008) also examine several risk sharing contracts that distribute the random yield risk among one supplier and one retailer[4], and their subsequent research study risk sharing contract with the consideration of a secondary market[5]; Kelle et al. (2009) study the buyer–supplier cooperation and negotiation under random yield[6]; Xiang et al. (2009) discuss the collection pricing decision in a remanufacturing system[7]; Xu (2010) studies the management problems of production and procurement in a decentralized supply chain consisting of one supplier and one manufacturer through option contracts[8]; Giri (2011) considers a single-product single-period inventory model under dual sourcing[9]. Tang et al. (2011) investigate a newsvendor problem in which the price decision will be postponed and determined upon recognition of random yield and prior to realizing demand uncertainties[10]. Readers are referred to the work of Tang et al. (2011) for extensive literature review of random yield.

In the field of risk pooling and centralized purchasing, many researchers investigate the effects of mitigating demand uncertainties[11-13] and different lead times[14] or price discount. For example, Munson and Hu (2010) compute lot sizes under four different purchasing frameworks when offered prices from either all-units or incremental quantity discount schedules. The four frameworks represent some of the most common configurations of centralized vs. decentralized pricing, purchasing, warehousing, and delivery[15].

Most of the articles related to on supply chain risk management are focused on the demand-side risk, especially demand uncertainty. And for those related to supply risk often considers only one method of mitigation strategy. Besides, how the supply risk pooling has benefit the retailers and how supply uncertainty affects the decisions of centralized purchasing have not been thoroughly studied. Therefore, our paper contributes to the literature of risk pooling in that we examine the value of supply risk pooling and compare the two cooperation strategies: inventory pooling and centralized purchasing. We show under the two different cooperation methods: (1) how the optimal order quantities are different; (2) the retailers' expected profits increase; and (3) closer cooperation brings larger profits.

2 Model Analysis

We consider two retailers not far from each other and face demand of D_1 and D_2 of the same product individually each cycle. Retailer 1 orders a quantity of Q_1 units from supplier 1, and can receive Q_1r_1units due to yield uncertainty of supplier 1. Similarly, retailer 2 orders Q_1 units from supplier 2, but can only receive Q_2r_2 units. Assume $0 \le r_i \le 1$, with a density function of $f_i(r_i)$ having a mean of u_i ($i = 1$, 2). Besides, the following set of notation is used: the unit selling price of the product is determined by the market and denoted as s; the wholesale price of supplier i is w_i; the unit inventory holding cost is the same for both retailers as h, and the unit penalty cost is p.

2.1 Models without Cooperation

If the two retailers do not cooperate and each retailer makes its own purchase decision and holds inventory separately, then for retailer 1, its expected profit can be written as:

$$\begin{aligned}\Pi_1 &= E[s\min(D_1,Q_1r_1) - w_1Q_1r_1 - h(Q_1r_1 - D_1)^+ - p(D_1 - Q_1r_1)^+] \\ &= s[Q_1\int_0^{D_1/Q_1} r_1 f_1(r_1)dr_1 + D_1\int_{D_1/Q_1}^{1} f_1(r_1)dr_1] - p\int_0^{D_1/Q_1}(D_1 - Q_1r_1)f_1(r_1)dr_1 \\ &\quad - wQ_1u_1 - h\int_{D_1/Q_1}^{1}(Q_1r_1 - D_1)f_1(r_1)dr_1\end{aligned} \tag{1}$$

Thus we have $\dfrac{d^2\Pi_1}{dQ_1^2} = -(s+h+p)\dfrac{D_1^2}{Q_1^3}f(\dfrac{D_1}{Q_1}) < 0$, which indicates that Π_{r1} is a strictly concave function with respect to Q_1. Therefore, similarly, suppose each retailer's order can be fulfilled by their respective supplier, then the optimal order quantity for retailer i is:

$$Q_{i(i=1,2)}{}^* \Rightarrow \int_0^{D_i/Q_i^*} r_i f_i(r_i)dr_i = \frac{(w_i+h)u_i}{s+h+p} \quad (2)$$

2.2 Models under Inventory Pooling

Suppose the two retailers also makes its own purchase decision from their suppliers respectively, but they cooperate in the form that their inventory is placed together somewhere near both of them. Each retailer pays for its own supplier, and if its supplier cannot satisfy him/her, he/she could buy from the other retailer at its wholesale price if there is over-stock for the other one. In this case, inventory is pooling and also the supply risks are shared.

2.2.1 The Expected Profit of Each Retailer under Inventory Pooling

We will first obtain the expected profit of retailer 1, which is analyzed under several different cases:

(1) If $D_1 - Q_1 r_1 > 0$, and $D_2 - Q_2 r_2 > 0$, that means both retailers are lack of inventory to satisfy individual demand, then penalty cost is $p\int_0^{D_1/Q_1}(D_1 - Q_1 r_1) f_1(r_1)dr_1 \cdot \int_0^{D_2/Q_2} f_2(r_2)dr_2$, holding cost is zero and purchasing cost is $w_1 Q_1 u_1$, while the revenue is $s\int_0^{D_1/Q_1} Q_1 r_1 f_1(r_1)dr_1 \cdot \int_0^{D_2/Q_2} f_2(r_2)dr_2$.

(2) If $D_1 - Q_1 r_1 > 0$, and $D_2 - Q_2 r_2 < 0$, in this case, retailer 1 is lack of inventory while retailer 2 has excess inventory, then retailer 1 could buy additional $Q_2 r_2 - D_2$ at a price of w_2:

(a) if $D_1 - Q_1 r_1 + D_2 - Q_2 r_2 > 0$, then retailer 1 is still short of $D_1 - Q_1 r_1 + D_2 - Q_2 r_2$ items; the purchasing cost is $w_1 Q_1 u_1 + w_2 \iint_{Q_1 r_1 + Q_2 r_2 < D_1 + D_2, Q_2 r_2 > D_2} (Q_2 r_2 - D_2) f_1(r_1) f_2(r_2) dr_1 dr_2$, holding cost is zero and penalty cost is $p \iint_{Q_1 r_1 + Q_2 r_2 < D_1 + D_2, Q_2 r_2 > D_2} (D_1 + D_2 - Q_1 r_1 - Q_2 r_2) f_1(r_1) f_2(r_2) dr_1 dr_2$, while the revenue is $s \iint_{Q_1 r_1 + Q_2 r_2 < D_1 + D_2, Q_2 r_2 > D_2} (Q_1 r_1 + Q_2 r_2 - D_2) f_1(r_1) f_2(r_2) dr_1 dr_2$.

(b) else it only buys a quantity of $D_1 - Q_1 r_1$, with both penalty cost and holding cost as zero, purchasing cost as $w_1 Q_1 u_1 + w_2 \iint_{Q_1 r_1 + Q_2 r_2 \ge D_1 + D_2, Q_1 r_1 < D_1} (D_1 - Q_1 r_1) f_1(r_1) f_2(r_2) dr_1 dr_2$, the revenue is $sD_1 \iint_{Q_1 r_1 + Q_2 r_2 \ge D_1 + D_2, Q_1 r_1 < D_1} f_1(r_1) f_2(r_2) dr_1 dr_2$.

(3) If $D_1 - Q_1 r_1 < 0$, and $D_2 - Q_2 r_2 < 0$, that means both retailers can satisfy their demand respectively and has excess inventory, then purchasing cost is $w_1 Q_1 u_1$, holding cost is $h \int_{D_1/Q_1}^{1} (Q_1 r_1 - D_1) f_1(r_1) dr_1 \cdot \int_{D_2/Q_2}^{1} f_2(r_2) dr_2$ and penalty cost is zero, while the revenue is $sD_1 \int_{D_1/Q_1}^{1} f_1(r_1) dr_1 \cdot \int_{D_2/Q_2}^{1} f_2(r_2) dr_2$.

(4) If $D_1 - Q_1 r_1 < 0$, and $D_2 - Q_2 r_2 > 0$, retailer 1 has excess inventory while retailer 2 cannot satisfy its demand, then retailer 2 could buys additional $Q_1 r_1 - D_1$ at a price of w_2:

(a) if $D_1 - Q_1 r_1 + D_2 - Q_2 r_2 > 0$, then retailer 2 is still short of $D_1 - Q_1 r_1 + D_2 - Q_2 r_2$ items; purchasing cost is $w_1 Q_1 u_1 - w_1 \iint_{Q_1 r_1 + Q_2 r_2 < D_1 + D_2, Q_1 r_1 > D_1} (Q_1 r_1 - D_1) f_1(r_1) f_2(r_2) dr_1 dr_2$, holding cost is zero and penalty cost is zero, the revenue is $sD_1 \iint_{Q_1 r_1 + Q_2 r_2 < D_1 + D_2, Q_1 r_1 > D_1} f_1(r_1) f_2(r_2) dr_1 dr_2$.

(b) else retailer 2 only buys $D_2 - Q_2 r_2$, with penalty cost as zero, purchasing cost as $w_1 Q_1 u_1 - w_1 \iint_{Q_1 r_1 + Q_2 r_2 \ge D_1 + D_2, Q_2 r_2 < D_2} (D_2 - Q_2 r_2) f_1(r_1) f_2(r_2) dr_1 dr_2$, and holding cost as $h \iint_{Q_1 r_1 + Q_2 r_2 \ge D_1 + D_2, Q_2 r_2 < D_2} (Q_1 r_1 + Q_2 r_2 - D_1 - D_2) f_1(r_1) f_2(r_2) dr_1 dr_2$, the revenue is $sD_1 \iint_{Q_1 r_1 + Q_2 r_2 \ge D_1 + D_2, Q_2 r_2 < D_2} f_1(r_1) f_2(r_2) dr_1 dr_2$.

Similarly, we can obtain retailer 2's expected profit, which is somewhat symmetrical to retailer 1's and omitted here.

2.2.2 The Benefits of Inventory Pooling for the Retailers

We will deduce the benefits of inventory pooling by comparing the various costs and revenue with and without inventory pooling.

(1) First, compare the inventory holding cost, the difference is

$$
\begin{aligned}
& h \int_{D_1/Q_1}^{1} (Q_1 r_1 - D_1) f_1(r_1) dr_1 + h \int_{D_2/Q_2}^{1} (Q_2 r_2 - D_2) f_2(r_2) dr \\
& - h \iint_{Q_1 r_1 + Q_2 r_2 > D_1 + D_2} (Q_1 r_1 + Q_2 r_2 - D_1 - D_2) f_1(r_1) f_2(r_2) dr_1 dr_2 \\
& = h \cdot \{ [\int_{D_1/Q_1}^{1} (Q_1 r_1 - D_1) f_1(r_1) dr_1 - \int_{0}^{1} \int_{D_1/Q_1 + (D_2 - Q_2 r_2)/Q_1}^{1} (Q_1 r_1 - D_1) f_1(r_1) dr_1 f_2(r_2) dr_2] \\
& + [\int_{D_2/Q_2}^{1} (Q_2 r_2 - D_2) f_2(r_2) dr_1 - \int_{0}^{1} \int_{D_2/Q_2 + (D_1 - Q_1 r_1)/Q_2}^{1} (Q_2 r_2 - D_2) f_2(r_2) dr_2 f_1(r_1) dr_1] \}
\end{aligned}
\quad . \tag{3}
$$

Now consider

$$\int_{D_1/Q_1}^{1}(Q_1r_1-D_1)f_1(r_1)dr_1-\int_0^1\int_{D_1/Q_1+(D_2-Q_2r_2)/Q_1}^{1}(Q_1r_1-D_1)f_1(r_1)dr_1f_2(r_2)dr_2$$

$$=\int_0^1\int_{D_1/Q_1}^{1}(Q_1r_1-D_1)f_1(r_1)dr_1f_2(r_2)dr_2-\int_0^1\int_{D_1/Q_1+(D_2-Q_2r_2)/Q_1}^{1}(Q_1r_1-D_1)f_1(r_1)dr_1f_2(r_2)dr_2,$$

$$=\int_0^1 f_2(r_2)dr_2\int_{D_1/Q_1}^{D_1/Q_1+(D_2-Q_2r_2)/Q_1}(Q_1r_1-D_1)f_1(r_1)dr_1$$

for $(D_2-Q_2r_2)/Q<0$, we have

$$\int_{D_1/Q_1}^{1}(Q_1r_1-D_1)f_1(r_1)dr_1-\int_0^1\int_{D_1/Q_1+(D_2-Q_2r_2)/Q_1}^{1}(Q_1r_1-D_1)f(r_1)dr_1f_2(r_2)dr_2$$

, where $Q_1r_1-D_1<0$,

$$=-\int_0^1 f_2(r_2)dr_2\int_{D_1/Q_1+(D_2-Q_2r_2)/Q_1}^{D_1/Q_1}(Q_1r_1-D_1)f_1(r_1)dr_1$$

and $\int_0^1 f_2(r_2)dr_2\int_{D_1/Q_1+(D_2-Q_2r_2)/Q_1}^{D_1/Q_1}(Q_1r_1-D_1)f_1(r_1)dr_1>0$;

otherwise for $(D_2-Q_2r_2)/Q>0$, then $\int_0^1 f_2(r_2)dr_2\int_{D_1/Q_1}^{D_1/Q_1+(D_2-Q_2r_2)/Q_1}(Q_1r_1-D_1)f_1(r_1)dr_1>0$.

Therefore, $\int_{D_1/Q_1}^{1}(Q_1r_1-D_1)f_1(r_1)dr-\int_0^1\int_{D_1/Q_1+(D_2-Q_2r_2)/Q_1}^{1}(Q_1r_1-D_1)f_1(r_1)dr_1f_2(r_2)dr_2>0$ always holds. Similarly,

$$\int_{D_2/Q_2}^{1}(Q_2r_2-D_2)f_2(r_2)dr_2-\int_0^1\int_{D_2/Q_2+(D_1-Q_1r_1)/Q_2}^{1}(Q_2r_2-D_2)f_2(r_2)dr_2f_1(r_1)dr_1>0\cdot$$

Therefore, the total expected holding cost definitely decreases under inventory pooling.

(2) Second, compare the under-stock penalty cost, the difference is

$$\begin{aligned}&p\int_0^{D_1/Q_1}(D_1-Q_1r_1)f_1(r_1)dr_1+p\int_0^{D_2/Q_2}(D_2-Q_2r_2)f_2(r_2)dr_2\\&-p\iint_{Q_1r_1+Q_2r_2\le D_1+D_2}(D_1+D_2-Q_1r_1-Q_2r_2)f_1(r_1)f_2(r_2)dr_1dr_2\\&=p\int_{D_1/Q_1}^{1}(Q_1r_1-D_1)f_1(r_1)dr_1+p\int_{D_2/Q_2}^{1}(Q_2r_2-D_2)f_2(r_2)dr_2\\&-p\iint_{Q_1r_1+Q_2r_2>D_1+D_2}(Q_1r_1+Q_2r_2-D_1-D_2)f_1(r_1)f_2(r_2)dr_1dr_2\end{aligned}\tag{4}$$

According to the analysis of holding cost, the total expected under-stock penalty cost also decreases under inventory pooling.

(3) Third, compare the expected profit, we have

$$s[Q_1\int_0^{D_1/Q_1} r_1 f_1(r_1)dr_1 + D_1\int_{D_1/Q_1}^{1} f_1(r_1)dr_1 + Q_2\int_0^{D_2/Q_2} r_2 f_2(r_2)dr_2 + D_2\int_{D_2/Q_2}^{1} f_2(r_2)dr_2 \\ -(D_1+D_2)\iint_{Q_1r_1+Q_2r_2>D_1+D_2} f_1(r_1)f_2(r_2)dr_1dr_2 - \iint_{Q_1r_1+Q_2r_2<D_1+D_2}(Q_1r_1+Q_2r_2)f_1(r_1)f_2(r_2)dr_1dr_2] \quad . (5)$$

2.3 Models under Centralized Purchasing

Under centralized purchasing, the retailers not only pool their inventory, but also purchase together. In this case, order decisions from two suppliers are made to maximize their total profits.

Therefore, decisions of Q_1 and Q_2 are made to maximize the total profits:

$$\Pi = s\iint_{Q_1r_1+Q_2r_2\geq D_1+D_2}(D_1+D_2)f_1(r_1)f_2(r_2)dr_1dr_2 + s\iint_{Q_1r_1+Q_2r_2<D_1+D_2}(Q_1r_1+Q_2r_2)f_1(r_1)f_2(r_2)dr_1dr_2 \\ -w_1Q_1u_1 - w_2Q_2u_2 - h\iint_{Q_1r_1+Q_2r_2\geq D_1+D_2}(Q_1r_1+Q_2r_2-D_1-D_2)f_1(r_1)f_2(r_2)dr_1dr_2 \\ -p\iint_{Q_1r_1+Q_2r_2<D_1+D_2}(D_1+D_2-Q_1r_1-Q_2r_2)f_1(r_1)f_2(r_2)dr_1dr_2 \quad (6)$$

The profit of each retailer depends on how they allocate the total inventory. In this paper, we focus on the benefit of cooperation for two retailers, and how the profit is split will be studied in further research.

3 Numerical Analysis

Now we have deduced the expected profits of two retailers under three situations: without cooperation, under inventory pooling and under centralized purchasing. To better understand the benefit of cooperation, we will use numerical examples to show how the benefit change with various key parameters. It is evident that the specific expression of $f_i(r_i)$ (i =1, 2) must be know in order to analyze the properties of Π_i. Hence, for illustration purpose, we assume both r_1 and r_2 follow uniform distribution, and $f_i(r_i)$ has a support on $[A_i, B_i]$ with $0 \leq A_i < B_i \leq 1$ (i =1, 2). Basic parameters are chosen randomly as $D_1 = 100$, $D_2 = 150$, $w_1 = 10$, $w_2 = 12$, $p = 5$, $h = 1$, $s = 20$, $A_1 = 0.4$, $B_1 = 1.0$, $A_2 = 0.5$, $B_2 = 1.0$. In this case, supplier 2 has larger wholesale price but lower supply risk compared to supplier 1.

(1) The impact of order quantities on the total profits

Because of the complexity of the function of Π_r, we use numerical example to illustrate the properties of Formula (6) with different values of (Q_1, Q_2) as in Figure 1, which illustrates that Π is a concave function with respect to both Q_1 and Q_2.

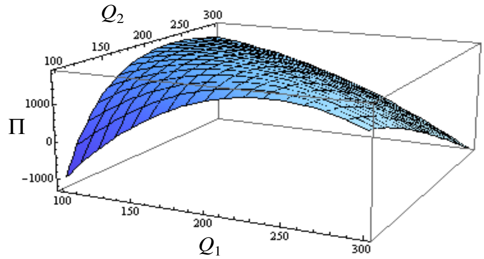


Fig. 1 The total profits of two retailers with different (Q_1, Q_2)

(2) The impact of cooperation on the optimal order quantities
As under inventory pooling, the retailers only place their inventory together, the optimal order quantities do not change. Table 1 lists the optimal order quantities under three situations with respect to different value of (p, h):

Table 1 The optimal order quantities with different (p, h)

(p, h)	(Q_1, Q_2) without cooperation or under inventory pooling	(Q_1, Q_2) under centralized purchasing
(5, 1)	(139, 189)	(187, 156)
(6, 1)	(141, 191)	(187, 158)
(7, 1)	(142, 193)	(187, 160)
(5, 2)	(136, 187)	(184, 156)
(5, 3)	(134, 185)	(182, 157)

It can be seen that, centralized purchasing tends to order more from supplier 1. That is, the supplier with lower wholesale price is better to be considered as the major supplier, even with larger supply risk. Besides, as unit penalty cost increases, or as unit holding cost decreases, both order quantities under two situations increase. That means, the retailers should order more if the penalty cost or the holding cost increases.

Table 2 The expected profits with different (p, h)

(p, h)	(Π_1, Π_2, Π) without cooperation	(Π_1, Π_2, Π) under inventory pooling	(Π_1, Π_2, Π) under centralized purchasing
(5, 1)	(722, 883, 1605)	(807, 982, 1789)	(807-851, 982-1026, 1833)
(6, 1)	(711, 867, 1578)	(801, 975, 1776)	(801-840, 975-1014, 1815)
(7, 1)	(700, 853, 1553)	(795, 966, 1761)	(795-833, 966-1004, 1799)
(5, 2)	(714, 876, 1590)	(798, 970, 1768)	(798-848, 970-1020, 1818)
(5, 3)	(706, 868, 1574)	(789, 957, 1746)	(789-846, 957-1014, 1803)

(3) The impact of cooperation on the total profit
First, by comparing the first two columns of Table 2, we know that retailer 2 could obtain more increased profit under inventory pooling than retailer 1. The reason for this phenomenon is retailer 2 faces larger demand and supply uncertainty, thus inventory pooling could mitigate supply risk more effectively. Second, as the cooperation becomes closer, the total profit increases from the first column to the last. Third, it is noted that in the third column, only the possible range of Π_1 and Π_2 are given to ensure the retailers could get larger profits than under inventory pooling, and the final values are determined through their negotiation. Therefore, the retailers should try to collaborate with each other while competing, which is the situation of "co-opetition".

(4) The benefit of cooperation
We define the benefit of cooperation to be the difference between the total profit without cooperation and that under inventory pooling or under centralized purchasing. By increasing A_1 from 0.01 to 0.4, we can plot the benefit of cooperation with different variance of r_1.

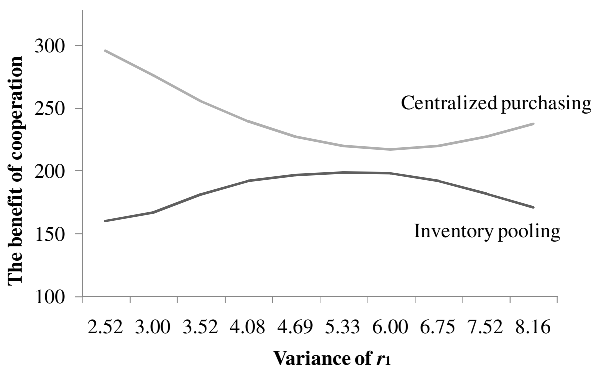


Fig. 2 The benefit of cooperation with different variance of r_1

It is clear from Figure 2 that, with higher cooperation degree (centralized purchasing compared to inventory pooling policy), the benefit of cooperation becomes more apparent. It is also interesting to find that the benefit doesn't increase of decrease with supplier 1's supply uncertainty, while denoting a non-linear curve. This phenomenon tells us larger supply risk doesn't always mean larger benefit of cooperation.

4 Concluding Remarks

As various risks increase, more and more companies choose to cooperate for risk sharing and pooling. In this paper, we focus on two retailers facing suppliers with yield uncertainty. Two kinds of cooperation methods are discussed: inventory pooling and centralized purchasing. Numerical analysis show that closer cooperation brings larger total benefit for the retailers, while larger supply risk doesn't always mean larger cooperation benefit.

Further studies are required to consider how the benefit is allocated between the retailers, and may investigate how both demand uncertainty and supply uncertainty affect the decisions. It would also be interesting to compare the transshipment strategy and inventory pooling or centralized purchasing, with the consideration of transportation costs.

References

1. He, Y., Zhang, J.: Random yield risk sharing in a two-level supply chain. International Journal of Production Economics 12(2), 769–781 (2008)
2. Lee, H.L.: Aligning Supply Chain Strategies with Product Uncertainties. California Management Review 44(3), 105–119 (2002)
3. Güler, M.G., Bilgiç, T.: On coordinating an assembly system under random yield and random demand. European Journal of Operational Research 196(1), 342–350 (2009)
4. He, Y., Zhang, J.: Random yield risk sharing in a two-level supply chain. International Journal of Production Economics 112(2), 769–781 (2008)
5. He, Y., Zhang, J.: Random yield supply chain with a yield dependent secondary market. European Journal of Operational Research 206(1), 221–230 (2010)
6. Kelle, P., Transchel, S., Minner, S.: Buyer–supplier cooperation and negotiation support with random yield consideration. International Journal of Production Economics 118(1), 152–159 (2009)
7. Li, X., Li, Y., Cai, X.: Collection Pricing Decision in a Remanufacturing System Considering Random Yield and Random Demand. Systems Engineering - Theory & Practice 29(8), 19–27 (2009)
8. Xu, H.: Managing production and procurement through option contracts in supply chains with random yield. International Journal of Production Economics 126(2), 306–313 (2010)
9. Giri, B.C.: Managing inventory with two suppliers under yield uncertainty and risk aversion. International Journal of Production Economics 133(1), 80–85 (2011)
10. Tang, O., Musa, S.N., Li, J.: Dynamic pricing in the newsvendor problem with yield risks. In: International Journal of Production Economics (February 4, 2011) (in Press Corrected Proof)
11. Liu, K., Zhou, Y., Zhang, Z.: Capacitated location model with online demand pooling in a multi-channel supply chain. European Journal of Operational Research 207(1), 218–231 (2010)
12. Bar-Lev, S.K., Schulte-Geers, E., Stadje, W.: On the demand pooling anomaly in inventory theory. Operations Research Letters 39(6), 411–413 (2011)
13. Kang, J.H., Kim, Y.D.: Inventory control in a two-level supply chain with risk pooling effect. International Journal of Production Economics 135(1), 116–124 (2012)
14. Wanke, P.F., Saliby, E.: Consolidation effects: Whether and how inventories should be pooled. Transportation Research Part E: Logistics and Transportation Review 45(5), 678–692 (2009)
15. Munson, C.L., Hu, J.: Incorporating quantity discounts and their inventory impacts into the centralized purchasing decision. European Journal of Operational Research 201(2), 581–592 (2010)

Understanding Post-acceptance Usage Behaviors—An Ambidexterity View

Yumei Luo, Hong Ling, Chenghong Zhang, and Zhengchuan Xu

Abstract. Recent information systems (IS) publications reveal an emerging interest in studying post-acceptance system usage behaviors. Compared to adoption and initial use, understanding of IS post-acceptance use is still at its early stage. To further develop knowledge about this phenomenon, this study reviews a variety of post-acceptance usage and ambidexterity concepts. An ambidexterity view is then proposed as an alternative perspective to understand these post-acceptance behaviors.

Keywords: post-acceptance use, ambidexterity, learning, innovation.

1 Introduction

Organizations have made huge investments in information technology over the last 30 years, resulting in many, if not most, intra-organizational work systems being IT-enabled. As a result, organizations' IT investment has been rising rapidly. Since the 1980s, organizations have spent to 50% of their new capital investment on IT-related activities[1]. However, the yields, compared to the heavy investment, seem far from satisfactory. Existing evidence strongly suggests that organizations underutilize the functional potential of the majority of this mass of installed IT applications [2]. Certain reports from industrial consultants have found a positive relationship between profitability of organizations and the degree of utilization of the implemented IS [3].

Mere acceptance cannot unleash the full potential of IT investments; thus, more attention has been paid recently to individuals' behaviors at the post-acceptance stages, which we call post-acceptance behaviors [4,2]. Mere acceptance cannot unleash the full potential of IT investments; thus, post- acceptance behaviors have

Yumei Luo · Hong Ling · Chenghong Zhang · Zhengchuan Xu
School of Management, Fudan University, ShangHai, China

Yumei Luo
School of Business Management and Tour management, Yunnan University, Kunming, China
e-mail: luoyumei@fudan.edu.cn

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 505–513.
springerlink.com

significant implications for organizations that seek to enhance their workers' job performance and thereby reap the full benefit from the high costs of IT infrastructure [5-7] . Accordingly, an increasing number of information systems studies have been performed in the past few years to explain various post-acceptance behaviors. Nevertheless, as similarities and differences exist among these concepts and outcomes, clearer conceptual understandings regarding these post-acceptance use concepts and the effect of them should benefit future research. Therefore, this paper reviews existing literature on post-acceptance IS usage behaviors and the nature of ambidexterity, discusses similarities and differences among these concepts, and develops ambidextrous behaviors insights regarding post-acceptance usage behaviors.

We organize the paper as follows. First, we present a view of post-acceptance behavior. We identify two aspects of post- acceptance behavior: routine use and innovative use. Second, reviews ambidextrous theory in prior literature. Then we develop a mapping exploitation and exploration in ambidexterity with routine use and innovative use in post-acceptance stage.

And develop some propositions which demonstrate how different usage behaviors affect individual performance. Finally, we conclude with implications for future research.

2 Literature Review

2.1 IS Implementation Process and Post-acceptance Usage Behaviors

The IS implementation process model was first conceived as consisting of six stages-initiation, adoption, adaptation, acceptance, routinization and infusion stages [6] (see Fig.1).

Initiation	Adoption	Adaptation	Acceptance	Routinization	Infusion

Fig. 1 IS implementation process model

Adoption and acceptance are two different stages in the IS implementation process. Adoption implies an organization's decision to allocate and ensure resources needed for the change and users' first time decision of whether to use an IS or not. Acceptance means that an organization's devotion of efforts to induce users to employ the implemented IS at work[6] and a user's IS implementation outcomes as better work performance, improved productivity, and user satisfaction [8].

Such user adoption and acceptance are important for initial IS success, however, the true return on IS investment more depends on the extensive an intensive use by employees during later stages of IS implementation process: the routinization and infusion stages [2,9,10].

According to Saga and Zmud(1994), routinization describes the state in which IS use is no longer perceived as out of the ordinary but actually becomes a normal part of the work processes, and infusion refers to the process of embedding an IS deeply and comprehensively in work processes. Therefore, routine use at the individual level can be conceived as usage behaviors perceived by employees as normal [10,11]. Infusion refers to the stage where the fullest potential of an IS has been integrated with an organization's operational and management processes [12]. From a user's viewpoint, the potential value of an IS could be realized through three alternative usage behaviors: extended use, integrated use, and emergent use [10]. **Extended use** is users' applying more of IS features to support a more comprehensive set of tasks at work [10,11]. Extended use represents a form of exploitative use. **Integrated use** refers to users' utilizing IS to establish or enhance work flow linkages among a set of tasks at work. **Emergent use** means applying IS to accommodate tasks that were not feasible or recognized prior to the application of IS at work. Emergent use, similar to individual feature extension[2] and trying to innovative with IT [13], essentially represents a form of explorative use. So, routinization stage is associated with routine use and infusion stage is associated with extended use and emergent use.

Routinization and infusion, which follow the acceptance stage, are conceived together as the post-acceptance stage [14] (See table 1).

Table 1 Mapping of Post-Acceptance Usage Behaviors and IS implementation Stages

<table>
<tr><td colspan="4">Organizational IS Implementation Stages where Individual Usage Behavior Occur</td></tr>
<tr><td colspan="2">Stage</td><td>Description of Stage</td><td>Possible Usage Behaviors</td></tr>
<tr><td rowspan="6">Post-acceptance</td><td rowspan="3">Rou-tini-zation</td><td rowspan="3">IS use is no longer perceived as out-of-ordinary but actually becomes part of the behavioral routine</td><td>Routine use</td></tr>
<tr><td>Habit use</td></tr>
<tr><td>Continual use</td></tr>
<tr><td rowspan="3">Infu-sion</td><td rowspan="3">Embedding an IS application deeply and comprehensively within an individual's or an organization's work systems</td><td>Extended/Deep use</td></tr>
<tr><td>Integrated use</td></tr>
<tr><td>Emergent use</td></tr>
</table>

Burton Jones (2006) provides that Innovative use includes extended use and emergent use. First, employees may engage in extended use or endeavor to use more of the available IS functions to support their work [15,10,11]. Incorporation of more IS features usually lead to better individual performance, and at the same time, more effective utilization of the implemented IS. Second, employees may engage in emergent use or experiment with the IS and apply it innovatively to enhance their job performance [4]. Emergent use further helps leverage the value potential of the implemented IS to an advanced level [2]. We, therefore, believe that infusion stage is associated with innovative use.

Given our focus on IS use by employees in their organizational contexts, we define post-acceptance usage behaviors for their work processes so as to focus on those behaviors for job-related purposes rather than other objectives. The functional complexity of modern IS allows employees to apply these technologies extensively and/or creatively to support task activities more fully [8,4,15,16]. In a typical situation in which a user applies an IS for a given task, the user's cognitive resources are limited [17]. The employee can display the both usage behaviors within a period of time.

2.2 Individual Behavioral Ambidexterity

Ambidexterity research has usually described organizational mechanisms that enable firms to simultaneously address exploitation and exploration. Studies have predominantly suggested that organizations pursuing exploration and exploitation simultaneously obtain superior financial performance [18,19].

In recent years, the concept of organizational ambidexterity has gained momentum in research on organizations. Large-scale empirical studies provide evidence of organizational ambidexterity's generally positive association with firm performance. The antecedents of ambidexterity include structural ambidexterity [18], informal network [20], and leadership-based [21,19]. Studies have started to explore how environmental [22,23] and organizational moderators [24,19] affect the interrelations between ambidexterity, its antecedents, and performance outcomes.

In most of these studies, the tensions that ambidexterity creates are resolved at the next organizational level down [25]. Consequently, a business unit may become ambidextrous by creating two functions or subdivisions with different foci[26]. A manufacturing plant may become ambidextrous by creating two different teams, one in charge of exploration and another in charge of exploitation [27], and a single team may become ambidextrous by allocating different roles to each individual [28]. In sum, research has suggested that structural mechanisms are used to enable ambidexterity, whereas most individuals are seen as focused on either explorative or exploitative activities.

However, Gibson and Birkinshaw(2004) describes business unit contexts that enable employees to conduct both explorative and exploitative activities. The important difference is that these studies assume that ambidexterity is rooted in an individual's ability to explore and exploit. Similarly, Mom et al. (2007) shows that some managers simultaneously engage in high levels of exploitative and explorative activities. So we have reason to believe that an individual (e.g. an employee) can show such ambidextrous behavior which refer to conducting both exploitive and explorative activities.

3 An Ambidexterity View of Post-acceptance IT Use

This study departs from March (1991) by conceptualizing exploration and exploitation at the individual level. March (1991, p. 71) considers the relation between exploration, which includes 'things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation', and exploitation, which includes 'such things as refinement, choice, production, efficiency, selection, implementation, execution' in organizational learning. In this study, we apply the distinction between exploration and exploitation to learning and innovation, albeit of different types.

While exploitation strategy refers to the refinement and extension of existing resources and competencies, explorative strategy describes organizations' experimentation with new alternatives [29]. Essentially, exploration and exploitation differ in the amount or type, rather than in the presence or absence, of learning [17].

Studies of organizational learning indicate that the essence of exploitative activities is creating reliability in experience [30-32] which is associated with deepening an individual's existing knowledge base[32]. Such exploitative activities of individual include using and refining their existing knowledge [32], applying, improving, and extending existing competences, technologies, processes and products [29].

Routine use indicate accumulated experience, albeit in an incremental manner [17]. At the individual level, the repetition of a certain set of usage procedures in order to comply with normal work process, deepens an individual's existing knowledge (e.g. using the same feature of IS each day) and involves some learning. March(1991) argue that all activity includes at least some learning. Routinization of behavior among individuals is a special form of exploitation that concerns very little learning [17] .

The essence of explorative activities is creating variety in experience [30-33] which is associated with broadening an individual existing knowledge base [32], experimenting with new approaches towards technologies, business processes, or markets [33]. Extended use describes employees' use of new or additional IS feature to accommodate more tasks [15]. Innovative use describes employees' application of IS in novel ways to support task performance or individuals' discovery of ways to apply IS feature that go beyond the ways delineated by the designers or implementers [2]. According to type or amount of learning, emergent use involves more dramatic learning and expands users' knowledge with regard to the potential of the installed IS than extended use, but these activities are all linked with new knowledge and incremental or radical innovation at the individual level [26,17,33]. So at the individual level innovative use (e.g. extended use and emergent use) is associated with explorative activities.

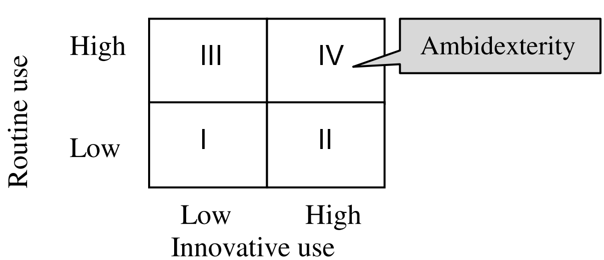


Fig. 2 Types of Exploitation and Exploration use in the post-acceptance stage

Routine use and innovative use form two side of ambidexterity at usage behaviors at individual level. To more fully understand the performance implications of ambidexterity in usage behaviors in the post-acceptance stage, we build on two dimensions: routine use and innovative use. When an individual simultaneously engage in high level routine use and innovative use, we believe that the individual has ambidextrous abilities (see fig.2).

4 Propositions

In figure 2, in Quadrant Ⅰ, an employee is only acceptance IS and begin use it, but he is not understand and know well the characteristics of IS. So we argue that Quadrant Ⅰ belongs to acceptance stage not post-acceptance stage. Quadrant Ⅱ, Ⅲ, and Ⅳ are belong to the post-acceptance stage.

Furthermore, applying the ambidexterity hypothesis to usage behaviors also implies that extreme positions along the exploitative-explorative use continuum may not be tenable: In Quadrant Ⅱ, an individual that engage in innovative use to the exclusion of routine use are likely to find that they suffer the costs of experimentation without gaining many of its benefits. They exhibit too many undeveloped new ideas, technologies and functions and too little distinctive competence. Conversely, In Quadrant Ⅲ, an individual who use all of its existing technology and functions unlikely to enhance it performance and likely to find themselves trapped in suboptimal stable equilibria. Scholars have pointed to the shortcomings inherent in focusing too much on one side or the other of this duality. And both types of actions are iteratively self-reinforcing. Because of the broad dispersion in the range of possible outcomes, exploration often leads to failure, which in turn promotes the search for even newer ideas and thus more exploration, thereby creating a “failure trap”. In contrast, exploitation often leads to early success, which in turn reinforces further exploitation along the same trajectory, thereby creating a “success trap.” In short, exploration often leads to more exploration, and exploitation to more exploitation. So, we propose that a curvilinear relationship between an employee’s routine use or innovative use and his performance.

Proposition 1: An inverted U-shaped relationship exists between an employee’s routine use and his performance.

Proposition 2: An inverted U-shaped relationship exists between an employee’s innovative use and his performance.

As a result, In Quadrant Ⅳ, maintaining an appropriate balance between innovative use and routine use is a primary factor in system value and prosperity.

In the context of our framework, we consider an individual’s ability to effectively reconcile tensions that arise from pursuing exploration and exploitation to be a necessary, but not sufficient, condition for ambidexterity in post-acceptance behaviors. Because we predict that balancing routine activities and innovation activities can contribute to enhanced performance [34]. The degree to which individual IS managers simultaneously engage in exploration and exploitation (i.e. ambidexterity) is found to be positively associated with individual performance[35]. March arguments, notwithstanding the adaptation benefits of both exploration and exploitation, the interplay between the two occurs in the form of a zero-sum game where exploration and exploitation compete for scarce resources, attention, and organizational routines; accordingly, logic dictates that exploration and exploitation be viewed as two ends of a continuum. But Katila and Ahuja (2002) conceptualized exploitation and exploration as orthogonal variables and the interaction between them will have a positive impact on new-product development. The two kinds of arguments are about the scarcity of resources. Some resources, such as information and knowledge, may be infinite [36]. Some resources, such as division of labor and allocation of resources, may be finite for individual [17]. In post-acceptance stage, employee is to not only obtain knowledge but also take time and energy. Compare to knowledge, time and energy is limited. So we propose that a curvilinear relationship between an employee's ambidexterity behavior in post-acceptance stage and his performance.

Proposition 3: An employee's ambidextrous use is more positively impact on his performance than either routine use or innovative use.

Proposition 4: An inverted U-shaped relationship exists between an employee's ambidextrous use and his performance.

5 Implications

As the IS field has accumulated a rich body of knowledge about adoption and initial acceptance of information technologies, there is an increasing interest in investigating post-acceptance technology usage behaviors. This study reviews the stage model of IS implementation and a variety of post-acceptance usage concepts in extant literature, maps these behaviors against their corresponding stages, and proposes ambidexterity view of these behaviors.

Based on an understanding of ambidexterity as an individual ability to simultaneously balance different activities in a trade-off situation [18], we develop a theoretical model linking different usage behaviors and component of them to individual performance. Our baseline proposition is that there is an inverted U-shaped relationship between routine use or innovative use and individual performance.

Complementary to extant knowledge of post-acceptance use, this paper offers additional insights into this phenomenon by discussing the differences between different post-acceptance use concepts and by categorizing them into two types that are differentiated by the nature of learning on an IS.

Meanwhile, studies of IS acceptance have consistently pointed out the true value of IS depends on post-adoption use. The view of prior research is that post-acceptance behaviors are indeed complex and sophisticated. So a deeper understanding of the behaviors and the impact on the performance offers potentially valuable insight for theory development.

References

1. Westland, J.C., Clark, T.H.K.: Global Electronic Commerce Theory and Case Studies. Universities Press (2001)
2. Jasperson, J., Carter, P.E., Zmud, R.W.: A comprehensive conceptualization of post-adoptive behaviors associated with information technology enabled work systems. MIS Quarterly 29(3), 525–557 (2005)
3. AberdeenGroup: The Precision Marketing Benchmark Report: How Top Performers Turbo-Charge Their Investment. Aberdeen Group (2006)
4. Ahuja, M.K., Thatcher, J.B.: Moving beyond intentions and toward the theory of trying:effects of work of work environment and gender on post-adoptation information technology use. MIS Quarterly 29(3), 427–459 (2005)
5. Kim, S.S., Son, J.-Y.: Out of dedication or constraint?a dual model of post-adoption phenomena adn its empirical test in the context of online services. MIS Quarterly 33(1), 49–70 (2009)
6. Cooper, R.B., Zmud, R.W.: Information technology implementation research: a technological diffusion approach. Management Science 36(2), 123–139 (1990)
7. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of inforamtion technology: toward a unifed view. MIS Quarterly 27(3), 425–478 (2003)
8. Agarwal, R., Karahanna, E.: Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. MIS Quarterly, 665–694 (2000)
9. Bhattacherjee, A., Premkumar, G.: Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. MIS Quarterly, 229–254 (2004)
10. Saga, V.L., Zmud, R.W.: The nature and determinants of IT acceptance, routinization, and infusion, pp. 67–86. Elsevier Science Inc. (1993)
11. Schwarz, A.H.: Defining information technology acceptance: a human-centered, management-oriented perspective (2003)
12. Jones, E., Sundaram, S., Chin, W.: Factors leading to sales force automation use: A longitudinal analysis. Journal of Personal Selling and Sales Management 22(3), 145–156 (2002)
13. Ahuja, M.K., Thatcher, J.B.: Moving beyond intentions and toward the theory of trying: Effects of work environment and gender on post-adoption information technology use. MIS Quarterly, 427–459 (2005)
14. Hsieh, J.J.P.A., Wang, W.: Explaining employees' Extended Use of complex information systems. European Journal of Information Systems 16(3), 216–227 (2007)
15. Saga, V.L., Zmud, R.W.: The nature and determinants of IT acceptance, routinization and infusion. Diffusion Transfer and Implementation of Information Technology 45(A-45), 67–86 (1994)
16. Wang, W., Hsieh, P.A.: Beyond routine: Symbolic adoption, extended use, and emergent use of complex information systems in the mandatory organizational context (2006)
17. Gupta, A.K., Smith, K.G., Shalley, C.E.: The interplay between exploration and exploitation. The Academy of Management Journal ARCHIVE 49(4), 693–706 (2006)
18. Gibson, C.B., Birkinshaw, J.: The antecedents, consequences, and mediating role of organizational ambidexterity. The Academy of Management Journal, 209–226 (2004)

19. Lubatkin, M.H., Simsek, Z., Ling, Y., Veiga, J.F.: Ambidexterity and performance in small-to medium-sized firms: The pivotal role of top management team behavioral integration. Journal of Management 32(5), 646–672 (2006)
20. Gulati, R., Puranam, P.: Renewal through reorganization: The value of inconsistencies between formal and informal organization. Organization Science 20(2), 422–440 (2009)
21. Beckman, C.M.: The influence of founding team company affiliations on firm behavior. The Academy of Management Journal ARCHIVE 49(4), 741–758 (2006)
22. Auh, S., Menguc, B.: Balancing exploration and exploitation: The moderating role of competitive intensity. Journal of Business Research 58(12), 1652–1661 (2005)
23. Jansen, J.J.P., Van Den Bosch, F.A.J., Volberda, H.W.: Exploratery Innovation, Exploitative Innovation, and Performance: Effects of Organizational Antecedents and Environmental Moderators. Management Science 52(11), 1661 (2006)
24. Atuahene-Gima, K.: Resolving the capability: Rigidity paradox in new product innovation. Journal of Marketing, 61–83 (2005)
25. Raisch, S., Birkinshaw, J., Probst, G., Tushman, M.L.: Organizational ambidexterity: Balancing exploitation and exploration for sustained performance. Organization Science 20(4), 685–695 (2009)
26. Benner, M.J., Tushman, M.L.: Exploitation, exploration, and process management: The productivity dilemma revisited. Academy of Management Review 28(2), 238–256 (2003)
27. Adler, P.S., Goldoftas, B., Levine, D.I.: Flexibility versus efficiency? A case study of model changeovers in the Toyota production system. Organization Science, 43–68 (1999)
28. Jansen, J.J.P., George, G., Van den Bosch, F.A.J., Volberda, H.W.: Senior team attributes and organizational ambidexterity: The moderating role of transformational leadership. Journal of Management Studies 45(5), 982–1007 (2008)
29. March, J.G.: Exploration and exploitation in organizational learning. Organization Science 2(1), 71–87 (1991)
30. Bontis, N., Crossan, M.M., Hulland, J.: Managing an organizational learning system by aligning stocks and flows. Journal of Management Studies 39(4), 437–469 (2002)
31. Holmqvist, M.: Experiential learning processes of exploitation and exploration within and between organizations: An empirical study of product development. Organization Science, 70–81 (2004)
32. Levinthal, D.A., March, J.G.: The myopia of learning. Strategic Management Journal 14(S2), 95–112 (1993)
33. McGrath, R.G.: Exploratory learning, innovative capacity and managerial oversight. Academy of Management Journal, 118–131 (2001)
34. Teece, D.J.: Dynamic capabilities and strategic management: organizing for innovation and growth. Oxford University Press, USA (2009)
35. Vidgen, R., Allen, P., Finnegan, P.: Towards' Open'IS Managers: An Exploration of Individual-Level Connectedness, Ambidexterity, and Performance, pp. 1–10. IEEE (2011)
36. Varian, H.R., Shapiro, C.: Information rules: a strategic guide to the network economy. Harvard Business School Press, Cambridge (1999)

Voice Communication Aid with Personal Digital Assistant for Autistic Children

Ippei Torii, Kaoruko Ohtani, Nahoko Shirahama,
Takahito Niwa, and Naohiro Ishii

Abstract. “Let’s Talk!” is a new AAC (Augmentative and Alternative Communication) application for personal digital assistant for autistic children. This new and remarkable application has many particular advantages comparing to existing AAC. We especially focused on an easy and simple manipulation. By tapping a symbol on a screen of a PDA with this application, a user can show his/her thoughts with pictures and sounds to others easily. There are 2 modes which can be switched depending on different situations of users. It has 120 symbols based on daily life and a user can also create the original page with new icons made by pictures or sound .The operation of “Let’s Talk!” is simple and easy. When a user chooses one from 12 categories on the top page, more specific symbols will be appeared and voice sound will come out by touching a symbol. For example, if a user chooses “Eat” category and “bread” for a symbol, the voice will say “I want to eat bread.” There are 2 modes which can be switched depending on different situations of users. On “Supportive Mode”, a supporter shows the application to a user. On Self-use Mode, a user can tell what he/she wants directly with categories and symbols. It is possible to make original icons with a camera or a voice recorder in PDA. A user also can customize an original page by arranging icons he made or existing symbols.

1 Introduction

Lately, PDA (Personal Digital Assistant) is getting attention as assistant tools for communication. Big appeals of the mobile tool are not only e-mail or camera functions but also users can customize it for their own convenience by installing

---

Ippei Torii · Kaoruko Ohtani · Takahito Niwa · Naohiro Ishii
Dept. of Information Science, Aichi Institute of Technology, 1247 Yachigusa, Yakusa-cho, Toyota, Aichi, 470-0392, Japan
e-mail: {mac,x08062xx,ishii}@aitech.ac.jp, kohtani@comcast.net

Nahoko Shirahama
MiaiYogo School, 1-51 Namimatsu, Miai-cho, Okazaki, Aichi, 444-0802, Japan
e-mail: shara@ab.cyberhome.ne.jp

J. Watada et al. (Eds.): Intelligent Decision Technologies, Vol. 2, SIST 16, pp. 515–523.
springerlink.com

many applications. Many assistant applications for PDA, such as Drop Talks[3], Voice4u[4], Tap to Talk[5], aimed to help autistic children who have communication disorder have already introduced. Those communication assistant tools are called AAC (Augmentative and Alternative Communication) or VOCA (Voice Output Communication Aid). Although many studies about VOCA [1][2] had been made and school educational fields have adopted these tools, they are not come into general use because of the high price and complicated operations. Therefore, we worked on the development of a new communication assistant tool with PDA and contrived an inexpensive and easy use application, “Let’s Talk!”, for autistic children to be able to use a high-tech equipment with no hesitancy.

2 Purpose of the Study

Bondy and Frost who developed PECS (Picture Exchange Communication System) say it is useful to teach children basic communication to start from how to express their requests.[6] People usually have strong need to communicate when they request something to others. And they really feel the joy of communication when their requests are satisfied. We discussed with teachers of special support schools who actually deal with autistic children about the situations or methods to develop this application, and narrowed down the request situations which may stimulate their desire to communicate with others. We focused to express basic requests such as "I want to do (something)." or "I need (something)." with simple operations. It is the big advantage of this application to express a request with only a few actions. We also thought to unite the design and technology of PDA will make it possible to break the barrier of communication. So we aimed to use the technology which supports design, and to make design which puts the life into technology.

3 Construction of the System

3.1 Usability of the Application

We focused on a simple manipulation without complicated explanation to develop this application. The reaction area is wide and the volume of the voice and sound is very clear and much enough to be able to hear outdoor or in a crowded place. These distinguished characters will let physically impaired children to use those AAC more easily. They can communicate with others whenever and wherever by using pocket-able PDA without carrying a special piece of equipment. It is also expected the unique contents of this application may create a chance to communicate with others. Autistic children will be accepted in the society by communication with the application and it will lead to increase the quality of life of them.

3.2 Two Modes of the Application

There are 2 modes on this application. First, on "Supportive Mode", a supporter who helps a person who has a disability starts communication. Second, on "Self-use Mode", a person who has a disability can show his/her request by voice with this application. By switching the modes depending on different situations of users, more effective communication will be expected.

3.3 Supportive Mode

It is difficult for children with significant intellectual challenges to choose what they really need to say among too many choices. On Supportive Mode of the application, a supporter starts to pick symbols fitting to the situation and the condition of a user who needs assistant for communication. A user can make a choice which is suitable for his request easily. When a user touches the symbol what he/she wants to tell, a simple2 words sentence comes out by voice sound. To limit the numbers of symbols picked by a supporter from 1to 4 makes it easy to choose.

===== Example =====
If a supporter guesses a user want to eat something.
1) A supporter chooses “Eat” among 12 categories.
2) A supporter estimates what a user wants to eat and
pick “cereal”, ”bread”, ”soup”, ”spaghetti” and
shows it to a user.
3) A user touches “cereal” and voice sound says “I
want to eat cereal.”

3.4 Self-use Mode

If a user understands how to manipulate this application, he/she can tell what he/she wants directly with categories or symbols in Self-use Mode. A user may feel the joy of understanding by trying to tell his/her thoughts to others with this mode and it will turn into more effective communication. On this mode, we adopted some distinguished ideas such as the wide reacting area which makes it easy to operate for people with disabilities.

===== Example =====
If a user tries to tell he/she wants to eat something.
1) A user chooses “Eat” among 12 categories.
2) A user touches what he/she wants to eat =”cereal”.
3) The voice sound says, “I want to eat cereal.”

3.5 Original Page

This application has about 120 icons which fit our daily life, but more icons may be required depending on a user's situation. On the other hand, if there are too many icons, it will take much time to choose and avoid the easy operation. The "Make" page of this application will provides more effective communication without confusion. On this page, a user can make his/her original icons with a camera or a voice recorder the PDA has or using an illustration or sound founded on web sites. It is sure to broaden the communication if there are some original icons which suitable with a user's dairy life, such as the places he/she goes frequently, familiar people or the dairy activities. While keeping its simple operations, easier and quicker communication is expected by using those original icons. Original Page is customizable to put icons which are created by user or existing symbols freely. We upgraded the application seven times in 3 months from the release date (April 23, 2011) to July 23. We have improved this application with adopting requests or correcting the bugs from the real users' voice through SNS (Social Network Service) such as Twitter and Face book. Communication is essential needs for human. If people with speech disorders can show their thoughts and feelings without a spoken language, it will be possible that people can understand and relate each other more deeply. The most important aim of this application is to let people with such disorders to feel the joy of communication and encourage them to communicate with others. We consider it will lead the society where people support each other.

4 Design of the System

The main purpose for the development of this application is to make it possible for autistic children to contact with others quickly and effectively. Therefore, it is important to figure out what they really need. We adopted ideas and opinions from the teachers of special support schools who contact with autistic children everyday. We made several test manufactures and asked the teachers to use them in real classes with the children. In Self-use mode and Supportive mode, we do not use the font in PDA, but show the title of symbols as a picture to unify the design of the application. However, on Original Page, there are two ways to create a symbol; to put a new symbol which is made by a user, or to use an existing symbol. This system makes it possible to use two symbols which were created indifferent ways at the same time, but on the other hand, there will be a lack of unity of design if the title as a picture and the title as a font are mixed up together. Therefore, we adopted the method to show the letter information separately by trimming the symbol to regular size automatically, when an existing symbol is loaded on Original Page. The information of title and sound which is required when an existing symbol is loaded on Original Page will be picked up from the XML file where the information is described previously. Original Page and interaction of programming files and XML files are shown in Fig. 1.

To load the sound file, pictures, and title form XML file (called it a plist in Integrated Development Environment), first the former picture and the name of sound file are pull out from the plist. Second, a new name of the file of a symbol that is trimmed and prepared is created based on the obtained date and time. Third, the picture is trimmed and a new file is created.

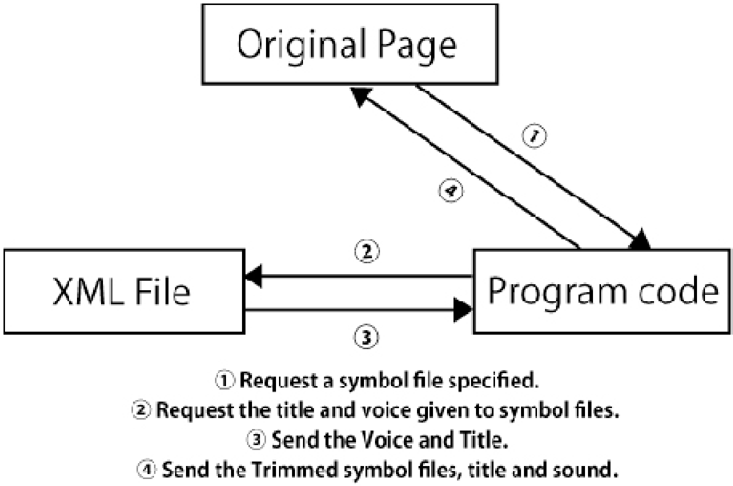


Fig. 1 Relation of Original Page and the XML file

To use this method brings some benefits. One of the benefits is the file of Supportive mode or Self-use mode will not need to be changed. It is a big progress to make it possible to put the existing symbols to Original Page without damaging the appearance, since we concentrated to keep good design qualities of this application. At the same time, the number of picture files is reduced by modifying and using the pictures of Supportive mode and Self-use mode. On iPhone AppStore, an application which is more than 20MB cannot be uploaded to iPhone or iPad without connecting through PC or Wi-Fi. The number of people who possess smart phones or high-performance music players is increased lately because of the popularity of smart mobile phone. But not all of the people who possess those high-performance equipments are familiar with computer. Also it is hard to say that the infrastructure to connect Wi-Fi is widespread among the general public. Therefore, we considered it is important to make it possible to install applications to the devices conveniently and immediately by using 3Gcircuit. It is effective to keep the size of the application under 20MB by using this method. In clinical trials, a subject child sometimes erased a symbol by mistake, or there were some children interested in erasing symbols. And the problem was found that pop-up alert was displayed repeatedly when switching the pages several times. It was happened because the reacting time of the delete button was too short. So we reset the length of the time to delete from1 second to 3 seconds. But another problem was occurred that the liquid panel reacted to a slight movement or vibration of a finger, and it caused to block to delete. We prepared the flag variable and set the initial value to 0, and added a postscript as below:

(1) When an icon is touched, set Flag to 1 and start to count the timer.
(2) When a finger leaves from an icon, set Flag to 0again. Flag also returns to 0 when a page moves by swiping.
(3) If Flag is 1 when the count of the timer becomes 3seconds, the alert to confirm to delete is displayed and Flag returns to 0 at the same time.

"Let's Talk!" is presumed to be used in noisy situation such as in a town or a classroom. Therefore, it is important that a user can communicate with others clearly in such circumstances. Many communication applications do not have enough volume or have the clipping noise which will prevent smooth communication in noisy places. Therefore, we tried one way after another to solve this problem. Usually 16 bit is suitable to record voice, but we adopted 24 bit instead. When we use 24 bit, we can enlarge or shrink the voice date in a wider range than 16 bit. The other reason to use 24 bit for recording is the sound quality does not deteriorate when we edit the voice data with an equalizer. To edit the voice data, we normalized there corded data (Fig.2) first. Then we calculated the average level and equalized all voice data. We show the process of normalization in Fig.3.

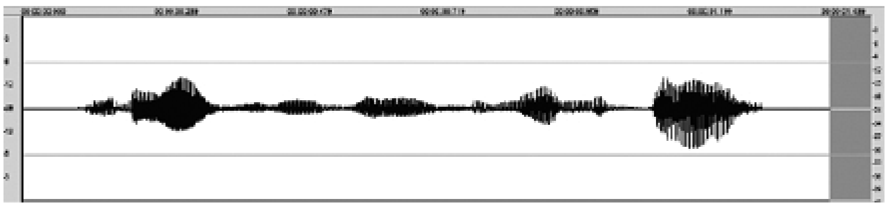


Fig. 2 The wave form of recorded voice data

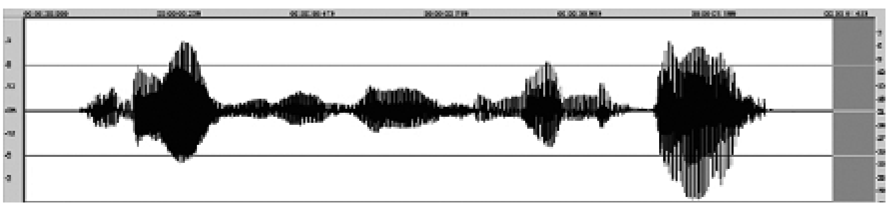


Fig. 3 The wave form after normalization

Many of autistic children have hyperacusis. They are over sensitive especially in a high register. Therefore, we cut a high register and tried to make the sound easy to listen for anyone. We applied an equalizer to change the frequency characteristics of normalized voice signal and cut a high register. We are in the audio waveform in Fig.4 are used in applications.

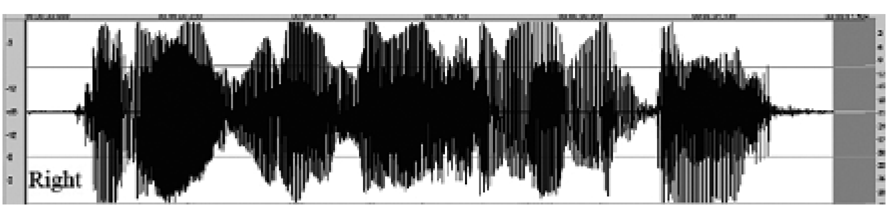


Fig. 4 The wave form of voice data on the application

The speakers of smart phones sometimes provide quite different sound compare with ordinary audio equipment. So unexpected noise will obstruct clear sound even we record a high quality voice data. Therefore, we prepared many samples of voice besides the data as stated above and modulated them. As a result, we heard many positive opinion such as, “I can listen the voice very clearly in the noisy place.”

5 Experiments at School for Handicapped Children

Introducing "Let's Talk" on the experimental basis was carried out at MiaiYogo School, which is the school for handicapped children. This application attracted the attention of most of the children and they really enjoyed to use it. They liked cartoon-like characters and were interested in the operation since it looked like a game for them. Some of the children who had never talked before started to communicate with teachers or parents with the application. It is obvious “Let’s Talk!” made great effect to those children with speech disorders. We made a survey about the reference between “Let’s Talk!” and the behavior of a child with assistance of MiaiYogo School.

5.1 Case Study 1

The subject of investigation: 8 year-old boy with autism who does not have ability to speak.

This student tried to tell his teacher he want to have another plate at lunch with “Let’s Talk!” Fig.5 shows how his behavior had changed when he started to use this application.

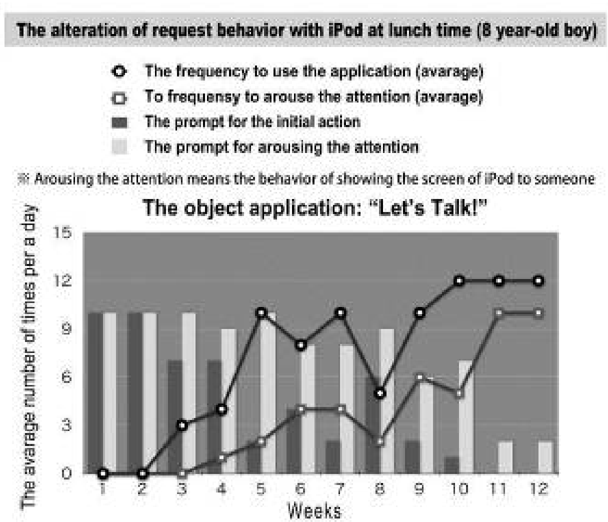


Fig. 5 The alteration of request behavior with iPod at lunch time (8 year-old boy)

It is clear that his motivation to communicate with the teacher had been increased by using iPad with this application. His classroom teacher reported that he often had offensive movement such as scratching or spitting to other students since he got in the school. But those behavior had been decreased dramatically after he started to learn how to express his thoughts to others in a proper way with this application. Rately, he tries to start communication from him with PECS (Picture Exchange Communication System) and iPod to let others to know his requests or needs.

5.2 *Case Study 2*

The subject of investigation: 10 year-old boy with autism who does not have ability to speak.

This student started to use iPad with “Let’s Talk!” when he tried to tell the teacher he want more, need help or reduse the quantity of the food at lunch time since the beginning of May 2011. According to his classroom teacher, he had learned how to manipulate this application quickly, and made his own icons by himself. He brings his own iPad anytime and uses it as a communication tool.

In these two case studies, it is considered that those students were satisfied as their teachers understood what they needed, even their requirements had not been fulfilled. It is important that they know someone understands their requests and empathizes with them. Many teachers of the school said most of the children seemed to enjoy their life at school more after introducing “Let’s Talk!”. Some of them use Fun page to make a funny sound and laugh with their friends. If they can have fun with this application, it will vitalize their life.

6 Conclusion

We developed a new communication assistant tool with PDA, “Let’s Talk!”, for autistic children. If autistic children feel the joy of communication by using this application, they will be strongly motivated to try to understand others thoughts. This application may have much possibility to be used by not only autistic children but also people who have problems of communication because of some diseases, such as pharyngeal cancer, cerebral palsy from a stroke, or senile dementia. We believe this application will help all the people with or without any disabilities to live in better life. And we also think if people can communicate with each other regardless of disabilities, it will provide new human resources and encourage developing the society where people support each other.

References

[1] Nakamura, K.: Effects and limitations of communication aids. Journal of Communication aids for People with Cognitive Disabilities. Higher Brain Function Research, 194–200 (2001)

[2] Sakai, S.: VOCA application to Autistic Children. National Welfare Foundation for Disabled Children, 29–32 (1999), http://voice4uaac.com/jp/
[3] Drop Talk, http://droplet.ddo.jp/droptalk/
[4] Voice4u, http://voice4uaac.com/jp/
[5] Tap To Talk, http://www.taptotalk.com/
[6] Frost, L., Bondy, A.: PECS: The Picture Exchange Communication System (2005)

Author Index

Subject Index