Thanaruk Theeramunkong
Susumu Kunifuji
Virach Sornlertlamvanich
Cholwich Nattee (Eds.)

LNAI 6746

Knowledge, Information, and Creativity Support Systems

5th International Conference, KICSS 2010
Chiang Mai, Thailand, November 2010
Revised Selected Papers

Springer

Lecture Notes in Artificial Intelligence 6746

Subseries of Lecture Notes in Computer Science

Thanaruk Theeramunkong Susumu Kunifuji
Virach Sornlertlamvanich Cholwich Nattee (Eds.)

Knowledge, Information, and Creativity Support Systems

5th International Conference, KICSS 2010
Chiang Mai, Thailand, November 25-27, 2010
Revised Selected Papers

Springer

Series Editors

Volume Editors

Thanaruk Theeramunkong
Cholwich Nattee
Thammasat University, Sirindhorn International Institute of Technology
131 Moo 5 Tiwanont Road, Bangkadi, Muang, Pathumthani 12000, Thailand
E-mail: {thanaruk, cholwich}@siit.tu.ac.th

Susumu Kunifuji
Japan Advanced Institute of Science and Technology
1-1 Asahi-dai, Nomi-shi, Ishikawa 923-1292, Japan
E-mail: kuni@jaist.ac.jp

Virach Sornlertlamvanich
National Electronics and Computer Technology Center
National Science and Technology Development Agency
112 Thailand Science Park, Phahonyothin Road, Klong 1, Klong Luang
Pathumthani 12120, Thailand
E-mail: virach.sornlertlamvanich@nectec.or.th

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-24787-3 e-ISBN 978-3-642-24788-0
DOI 10.1007/978-3-642-24788-0
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011938974

CR Subject Classification (1998): I.2, H.3-5, J.1, K.6, I.4, K.4

LNCS Sublibrary: SL 7 – Artificial Intelligence

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

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

Preface

The series of the International Conference on Knowledge, Information and Creativity Support Systems (KICSS 2010) has been held annually since 2006. Starting in Thailand (2006), the KICSS conferences have been held in Japan (2007), Vietnam (2008), and Korea (2009). KICSS 2010, the fifth in the series, was held in Chiang Mai, Thailand, during November 25–27, 2010. KICSS 2010 is a peer-reviewed international conference in the areas of knowledge science, knowledge engineering, information systems, and creativity support systems. It provides an international forum for researchers as well as industry practitioners to share their new ideas, original research results, and practical development experiences from all knowledge science-related areas including creativity support, decision science, knowledge science, data mining, machine learning, databases, statistics, knowledge acquisition, automatic scientific discovery, data/knowledge visualization, and knowledge-based systems. This year, we received 72 research papers from various countries including Argentina, China, Colombia, India, Indonesia, Iran, Japan, Korea, Malaysia, Poland, Thailand, and Vietnam. Every submission was rigorously reviewed by two or three reviewers with a single-blind protocol. The initial results were discussed among the reviewers and finally judged by the Program Committee Chairs. The Program Committee members were deeply involved in a highly selective process. As a result, only 25 papers (34.72%) of the 72 submissions were accepted as regular papers and 12 (16.67%) of them were accepted as short papers. In addition, nine papers (12.5%) were accepted for posters. Among these accepted papers, we selected papers with high quality to include in this LNAI/LNCS volume. The criteria of selecting the papers are that (1) the selected papers were recommended by the Session Chairs who listened to the presentation and judged that they are suitable to include in the proceedings, (2) not more than 32% of the submitted papers were selected, and (3) at least 30% of its content is different from the original one. Finally, we included 23 papers in this volume.

We highly appreciate the Program Committee members and external reviewers for their technical effort in providing straightforward scientific comments and impartial judgments in the review process of KICSS 2010. Many thanks are given to the keynote speaker and invited speakers. We thank Hiroshi Motoda, our General Chair, for his useful guidance and great effort in various spectrums related to conference arrangement. We are also grateful to the committee members who helped in the local arrangements, Ekkarat Boonchieng, Rattasit Sukhahuta (Chiang Mai University, Thailand), Thepchai Supnithi (NECTEC, Thailand), and Supot Maigunta (Rajamangala University of Technology Lanna, Thailand), for their unlimited help toward the success of the conference. We would like to give special thanks to Nattapong Tongtep for his help in the publishing process of this Volume, and to Wirat Chinnan and Swit Phuvipadawat for their support

with the KICSS 2010 conference website. Juniar Ganis, Jakkrit Techo, Nichnan Kittiphattanabawon, Nongnuch Ketui, Nattapong Tongtep, Thawatchai Suwana pong, Thatsanee Charoenporn, Ithipan Methasate, Tanatorn Tanantong, and Eakasit Pacharawongsakda helped a lot in the organization of this conference. While the arrangement of KICSS 2010 involved so many people, we would like to express an additional thank you to the contributors who helped at KICSS 2010 but whose names cannot be listed.

We greatly appreciate the support from various institutions. The conference was organized by the Sirindhorn International Institute of Technology (SIIT), Thammasat University (TU, Thailand), and co-organized by the National Electronics and Computer Technology Center (NECTEC, Thailand) and the Japan Advanced Institute of Science and Technology (JAIST, Japan). It was sponsored by the National Electronics and Computer Technology Center (NECTEC, Thailand), the Thailand Convention and Exhibition Bureau (TCEB), and the Air Force Office of Scientific Research/Asian Office of Aerospace Research and Development[1] (AFOSR/AOARD). Finally, we wish to thank all authors and all conference participants for their contribution and support.

March 2011

Thanaruk Theeramunkong
Susumu Kunifuji
Virach Sornlertlamvanich
Cholwich Nattee

[1] AFOSR/AOARD support is not intended to express or imply endorsement by the U.S. Federal Government.

Organization

General Chair

Hiroshi Motoda	Osaka University, Japan

Program Committee Co-chairs

Thanaruk Theeramunkong	SIIT, Thammasat University, Thailand
Susumu Kunifuji	JAIST, Japan
Virach Sornlertlamvanich	NECTEC, Thailand

Organizing Committee Co-chairs

Thanaruk Theeramunkong	SIIT, Thammasat University, Thailand
Susumu Kunifuji	JAIST, Japan
Virach Sornlertlamvanich	NECTEC, Thailand
Ekkarat Boonchieng	Chiang Mai University, Thailand
Rattasit Sukhahuta	Chiang Mai University, Thailand

Members

Pakinee Aimmanee	SIIT, Thammasat University, Thailand
Marut Buranarach	NECTEC, Thailand
Thatsanee Charoenporn	NECTEC, Thailand
Krisana Chinnasarn	Burapa University, Thailand
Tsutomu Fujinami	JAIST, Japan
Choochart Haruechaiyasak	NECTEC, Thailand
Tessai Hayama	JAIST, Japan
Hideaki Kanai	JAIST, Japan
Canasai Kruengkrai	TCL, NICT, Thailand
Verayuth Lertnattee	Silpakorn University, Thailand
Ithipan Methasate	NECTEC, Thailand
Ekawit Nantajeewarawat	SIIT, Thammasat University, Thailand
Cholwich Nattee	SIIT, Thammasat University, Thailand
Chutima Pisarn	Prince of Songkla University, Thailand
Annupan Rodtook	Ramkamhaeng University, Thailand
Mikifumi Shikida	JAIST, Japan
Samerkae Somhom	Chiang Mai University, Thailand
Kritsada Sriphaew	Tokyo Institute of Technology, Japan
Kozo Sugiyama	JAIST, Japan

KICSS 2010 Program Committee

Members

Pakinee Aimmanee	SIIT, Thammasat University, Thailand
Mohammad Shafkat Amin	Wayne State University, USA
Marut Buranarach	NECTEC, Thailand
Rui Camacho	Porto University, Portugal
Michelangelo Ceci	University of Bari, Italy
Krisana Chinnasarn	Burapa University, Thailand
Bing Guo	Sichuan University, China
Tessai Hayama	JAIST, Japan
Hideaki Kanai	JAIST, Japan
Hiroyuki Kawano	Nanzan University, Japan
Saori Kawasaki	JAIST, Japan
Asanee Kawtrakul	Kasetsart University, Thailand
Yasushi Kiyoki	KEIO University, Japan
Canasai Kruengkrai	TCL, NICT, Thailand
Chung-Hong Lee	National Kaohsiung University of Applied Sciences, Taiwan
Ee-Peng Lim	Singapore Management University, Singapore
Sanparith Marukatat	NECTEC, Thailand
Ithipan Methasate	SIIT, Thammasat University, Thailand
Taneli Mielikainen	Nokia Research Center Palo Alto, USA
Taizo Miyachi	Tokai University, Japan
Riichiro Mizoguchi	Osaka University, Japan
Jun Munemori	Wakayama University, Japan
Maybin Muyeba	Manchester Metropolitan University, UK
Cholwich Nattee	SIIT, Thammasat University, Thailand
Tadashi Nomoto	National Institute of Japanese Literature, Japan
Taisuke Ogawa	JAIST, Japan
Manabu Okumura	Tokyo Institute of Technology, Japan
Tansel Ozyer	TOBB University, Turkey
Chaveevan Pechsiri	Dhurakij Pundit Universtiy, Thailand
Michael Schmidt	Fluid Operations AG, Germany
Giovanni Semeraro	University of Bari, Italy
Xiaowei Shao	University of Tokyo, Japan
Akira Shimazu	JAIST, Japan
Thepchai Supnithi	NECTEC, Thailand
Xijin Tang	Chinese Academy of Sciences, China
Nattakorn Thubthong	Chulalongkorn University, Thailand
Takeaki Uno	National Institute of Informatics, Japan
Chai Wutiwiwatchai	NECTEC, Thailand
Hsin-Chang Yang	National University of Kaohsiung, Taiwan
Min Yao	Zhejiang University, China

Takaya Yuizono	JAIST, Japan
Shichao Zhang	Guangxi Normal University, China
Yang Zhang	Northwest A&F University, China
Kaidi Zhao	Amazon.com Inc., USA

Secretariats

Chair

Thatsanee Charoenporn	NECTEC, Thailand

Members

Juniar Ganis	SIIT, Thammasat University, Thailand
Peerasak Intaraapaiboon	SIIT, Thammasat University, Thailand
Nongnuch Ketui	SIIT, Thammasat University, Thailand
Nichnan Kittiphattanabawon	SIIT, Thammasat University, Thailand
Ithipan Methasate	NECTEC, Thailand
Eakasit Pacharawongsakda	SIIT, Thammasat University, Thailand
Chawisa Phumdontree	SIIT, Thammasat University, Thailand
Sudchaya Saengthong	SIIT, Thammasat University, Thailand
Thawatchai Suwanapong	SIIT, Thammasat University, Thailand
Tanatorn Tanantong	SIIT, Thammasat University, Thailand
Jakkrit Techo	SIIT, Thammasat University, Thailand
Nattapong Tongtep	SIIT, Thammasat University, Thailand

Conference Web Administrators

Wirat Chinnan	SIIT, Thammasat University, Thailand
Swit Phuvipadawat	Tokyo Institute of Technology, Japan
Thanasan Tanhermhong	SIIT, Thammasat University, Thailand

Hosts and Sponsors

Sirindhorn International Institute of Technology (SIIT), Thammasat University, Thailand
National Electronics and Computer Technology Center (NECTEC), Thailand
Japan Advanced Institute of Science and Technology (JAIST), Japan
Chiang Mai University (CMU), Thailand

Sponsors and Supporters

The Air Force Office of Scientific Research, Asian Office of Aerospace Research and Development (AFOSR/AOARD), USA
Japan Creativity Society (JCS), Japan
Thailand Convention and Exhibition Bureau (TCEB), Thailand

Table of Contents

Presentation Support System Providing the Function of Promoting Comment Collection in Limited Time

Hui Cheng, Tessai Hayama, and Susumu Kunifuji

School of Knowledge Science, Japan Advanced Institute of Science and Technology
{niniparty,t-hayama,kuni}@jaist.ac.jp

Abstract. In recent years, the use of presentations as an efficient method for communication has been proven to play a significant role in professional research activities, especially from the standpoint of education and business in enterprises. In these scenarios, this method is intended to clearly convey information from the presenter to the audience in a limited time, and to garner active feedback from the audience by means of human interaction. However, for the majority of presentations, it is difficult to stimulate the audiences to comment and discuss the topic, as well as to receive sufficient comments in a limited time. For these reasons, we have attempted to devise a system that provides a function for collecting the maximum possible comments in a time-limited presentation. By increasing the possibilities offered to presenters for handling such situations, and to remind both the presenters and audiences of the pace of the presentation, we conducted an experiment to prove that the approach used in our system is effective.

Keywords: presentation support system, meeting support with limited time, comment collection, human interaction.

1 Introduction

In recent years, presentations have been used in every aspect of life as a method of efficient communication. They have been proven to play a significant role in professional research activities, especially from the standpoint of education and business in enterprises. In these scenarios, a method for clearly conveying the information from the presenter to the audience in a limited time is needed. In addition, this method must be designed to elicit active feedback from the audience. However, for the majority of presentations, it is difficult to stimulate audiences to comment and discuss the topic, as well as to receive sufficient comments in a limited time. Sometimes audiences intend to communicate and comment with each other while watching the same content, however, they may miss presenter's message if discuss during the time-limited presentation. Therefore, how to cope with this issue is the subject in this paper.

Since the significance of obtaining information and comments from presentation, lots of researchers are devoted to presentation supporting work. In Japan, there is a workshop named "WISS", one aim of this workshop is researching on-line chat system in presentation for promoting discussion from 1997 and produced many excellent researches. For example, Kurihara and his partners developed AirTransNote which conducted a practical environment of realizing augmented classroom with wireless digital

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 1–12, 2011.

pens[1] and Presentaiton Sensei which is a presentation training system using speech and image processing[2]. In recent years, the majority of research studies on presentation methods in the groupware field have changed to concentrate on increasing the quantity of comments to promote discussion by using available media, such as mobile phones and papers. Such as Kobayashi and Nishimoto's system, which is an enhanced chat system crossing media[3]. Yet another example is from Doi, who designed a lecture supporting system[4] on the basis of the annotation system CollabSticky[5] to collect comments using various media. In this study, more available media, such as mobile phones and papers, are utilized. In an anonymous environment, audiences feel more comfortable in participating in the presentation. Davison's system[6], for example, provides a presentation support system for collecting comments in a convenient input form by using parallel, non-oral communication channels. This type of design can prevent strong personalities from dominating the discussion in an anonymous environment. Another example is Nishida's system[7], considering the disadvantage of an anonymous environment which has a high possibility of producing irresponsible statements in conversations, Nishida developed a tool that can convert from an anonymous to a non-anonymous environment with the agreement of the other group members. Compared with Davison, Nishida's system has performed well in terms of collecting comments and promoting discussion in a practical scenario. As time-limited factor, researches on facilitating communication while watching real-time content are proposed, such as Nishida's "On-Air Forum" [8]. Altogether, until now, through the approach of reducing an audience's psychological burden and providing more media to stimulate the audience to join the discussion, the research studies on presentation support systems have had a good effect. Nonetheless, problems remain, such as improving an audience's understanding and responding to the desire of audience members to immediately write down their opinions on the presentation. More importantly, time limitation of presentation has been neglected in previous research studies, particularly in seminars on educational themes and product-development meetings. Therefore, it is essential to carry out presentation research to increase the number of comments in a limited time.

In this study, we focus on the motivation of that audience can listen to the presentation while sharing comments with other audiences, without too much mental burden, and an audience to contribute comments in a limited time and the development of a presentation support system that increases the number of comments in a limited time. In this system, in order to create an environment in which audiences can express opinions in a limited time, we have come up with the following two specific approaches for obtaining the target:

1) the presenter handles the situation of fielding comments from the audience.
2) both the presenter and the audience are aware of the presentation process.

Through these approaches, we attempt to achieve the effect in which the presenter can manage the audience's commenting time to ensure that they have enough time to raise comments. As a result, we have tested the system and it has reached a good effect on achieving understand in presentation by the reflected consciousness and contributing comments by using airtime.

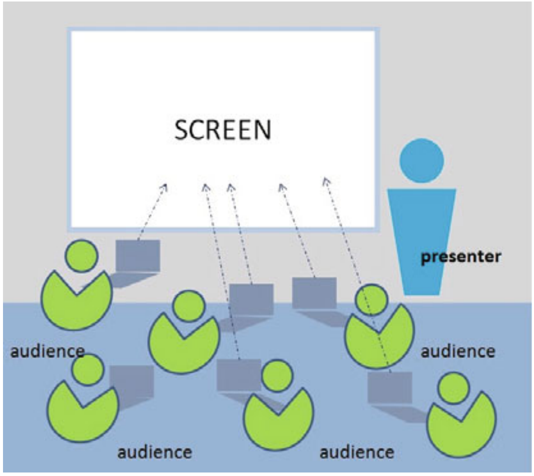


Fig. 1. Target presentation environment

1.1 Target Configuration of Presentation

The target presentation environment for this research study is as follows: 1) the presenter uses a PC and projects a magnified slide image on a large screen, 2) every audience member uses a PC, 3) each PC is connected to the network and 4) it is possible for the audience members to send their comments to the screen at any time during the presentation, as shown in Figure 1. An environment setting such as Figure 1 is generally used in related research studies as Hatanaka'system[9] and Kobayashi's system[10] and so on. Also this kind of presentation has been frequently put into practical use in academic conferences, educational scenarios and business in enterprises in recent years. Besides presentation support, education support is also considered to be another target of this research field.

1.2 Approaches

Our approaches for comment collection in a time-limited presentation are as follows:

1) The function of the presenter of handling the audience situation:
   (a) Providing a possibility of handling a number of potential comments from the audience.
   (b) Providing a possibility of handling the most interesting contents in the presentation.
2) The function of reminding both the presenter and the audience of the pace of the presentation:
   (a) Showing a timer and the page number of the current slide.
   (b) Showing the quantity of potential comments.

Fig. 2. Memo number editing bar for presenter

We observe that, in a real presentation situation, audience members write down some simple keywords. These keywords are helpful in capturing intriguing ideas that suddenly enter the audience member's mind while listening. Meanwhile, during the question time which is the last part of a presentation, audience members refer to these keywords as a reminder for formulating their final comments and formal questions.

As shown in Table 1 and as mentioned above, in a common presentation scenario, we try to take advantage of the simplicity of memos in order to develop a presentation support system. The distinctive features of memos, such as clarity, brevity and time-saving, help audiences capture their original ideas. These memos function as a reminder for the audience members to recall their thoughts and formulate their final comments.

For the function of the presenter of handling the audience situation, memos written by the audience members are captured and counted in a memo column. Viewing the count in the memo column helps the presenter handle the audience situation. Based on this number, the presenter can attempt to wait for the audience members to edit comments. In addition, the memo column can also help the presenter capture the most interesting contents in the presentation and make any necessary adjustments.

For the function of reminding both the presenter and the audience of the pace of the presentation, our design includes a timer, slide numbers and the content of the memos in order to assist the presenter in handling the balance of the presentation effectively.

2 System Implementation

2.1 System Description

In this system, by utilising the simplicity and clarity of the memos, we designed a presentation environment that provides the number of contributions and facilitates the collection of memos. We added some figure boxes that display on the presenter's screen in order to facilitate communication and reach a consensus regarding the degree of understanding between the presenter and the audience. These figure boxes include the number and content of the audience members' memo contributions, and a time distribution showing a timer and the state of the slideshow.

As Figures 3 and 4 show, this system consists of two separate interfaces, for the presenter and the audience, which are constituted of a data management section and a server section.

Here, we introduce the system separately based on the viewpoints of the presenter and the audience, respectively.

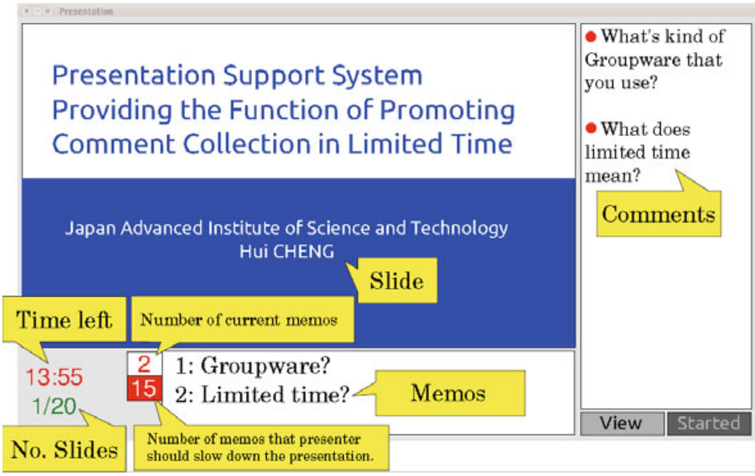


Fig. 3. Interface for presenter

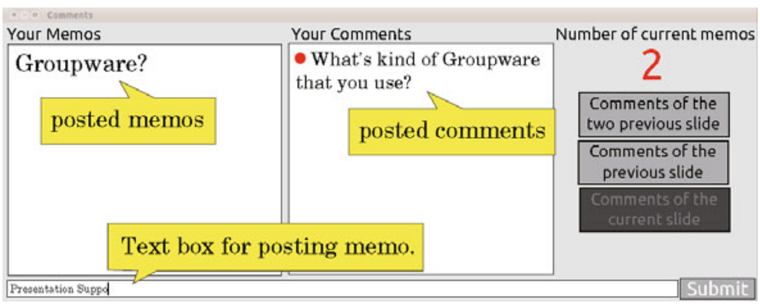


Fig. 4. Interface for audience

For the presenter, we designed two interfaces, as shown in Figures 2 and 3. Before beginning the presentation, the presenter must upload the slide with the presentation environment by using the view button (see Figure 3). We use extraction programming to achieve an inset construction between ordinary PowerPoint software and this presentation commenting system. In real usage, the presenter is required to set a number as shown in Figure 2. This number reminds the presenter of the number of memos that have arrived, then he or she will attempt to wait for the audience members to submit comments by slowing down the presentation pace or by consciously putting the points he or she intends to stress after the audience members' submissions, and also provide hints to the audience of the time for submitting comments. This number can be set according to the number of audience members or the number of slides, as desired by the presenter.

During a presentation, through the use of the timer and the state of the slideshow progress, the time-distribution function provides a hint to the presenter. In the meantime, the number of current memos implies the number of potential comments. This number is displayed together with the content of the memos, with the intention of signalling the potential need to slow down the presentation. Here, we emphasise that the memos which are sent by the audience members can be easily understood because they consist of simple words. The number of current memos is revised when new memos arrive. Also, when the number of current memos reaches a multiple of the number which is set before the beginning of the presentation by the presenter, the background colour changes from yellow to red, in order to gain the attention of both the presenter and the audience members, and to remind the audience members that it is time to submit comments.

For the audience members, we designed the memo column on the basis of the scenario that the audience members record their ideas or questions on paper as usual. In this way, any short content that they think of about the presentation can be written down in the Text Box for posting memos. In addition, the memos which the audience members have wrote down are sent to both the Your Memos column of the audience interface and the Memos column of the presenter interface. When the number of memos becomes a multiple of the number being set, together with the change in the background colour, it reminds the audience of the timing of comment editing. At that time, the audience members can edit the comment and send it to the Comments column by consulting the memos accumulated in the left column.

According to the above design, the presenter uploads the slide before the beginning of the presentation. When he or she clicks the Start button, the presentation starts. Meanwhile, as the slides progress, the audience members contribute comments to the relevant slide. In particular, the audiences can contribute their comments to the current slide or to the last or next-to-last slide. By means of this design, we intend to provide the audience members with the possibility of contributing comments to any slide. All comments are displayed in the Comment column as Anonymous.

On the other hand, as the presenter proceeds with his or her slides using the presentation environment, he or she can be reminded of the state of progress of the presentation. Therefore, the presenter can carry out a balanced presentation that does not exceed the allotted time.

By means of the design introduced above, we intend that this presentation support system can collect comments in a limited time and that it functions according to the following three aspects:

1) By skimming over the memos (i.e. the contents of the memos), the presenter can grasp the nature of comments from the audience members.
2) By skimming over the progress state which includes the timer and the progression of slides, the presenter can manage the speed of his or her presentation, so as not to exceed the allotted time.
3) By skimming over the numbers which are shown, the audience members can contribute comments in a gathering time.

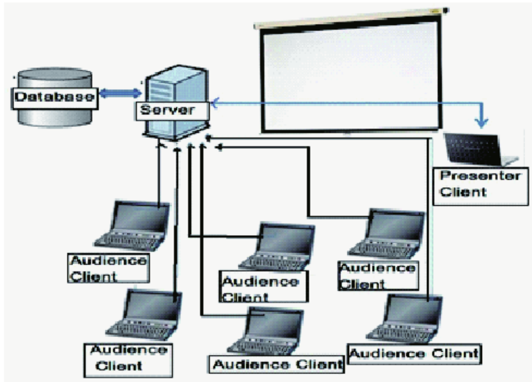


Fig. 5. System construction

2.2 System Configuration

As shown in Figure 5, our system configuration consists of two kinds of clients and a server. The interface section is implemented by Java GUI, the communication section between the clients and the server by Java RMI and the database section by MySQL. All data is controlled by the server.

In the logging section, the input user ID and password are sent to the server. After the ID and password are authenticated by the database of the server, the interfaces appear separately, as shown in Figures 3 and 4. The audience members edit memos and contribute comments as shown in Figure 4, and all the memos and comments are sent to the server. Then, along with the time and user data, new memos and comments are recorded in the database. Meanwhile, as new data, the new memos and comments are sent to the presenter's client system. After receiving the data, the new memos and comments are shown in relevant columns. Furthermore, by utilising Microsoft PowerPoint in this system environment, we use image-extraction programming to convert the slide file into a JPG file. From the above, we obtain a system configuration which enables the audience members to send memos and comments, and which shows them the interface of the presenter client.

3 Experiment

3.1 Purpose and Procedure

In this study, we implemented a presentation support system to enable audience members to easily comment on the presentation. We also conducted an experiment which proved that the approach of our system is effective. Hereinafter, we summarise the approaches of our experiment as:

1) Whether the presenter can finish the presentation in the limited time.
2) Whether the audience can contribute comments by using the memo function during the presentation time.

3) Whether the presentation time can be managed properly by means of reminders of the number of memos submitted.

We tested the first experimental approach by comparing the actual time that the presenter spent with the limited time. Through analysing the history of the timing of both the memos and comments, we attempted to check the feasibility of the second experimental approach. As for the third approach, we investigated the relationship between the time distribution of comments and the progression of each slide during the entire presentation.

We attempted to carry out the experimental procedure twice. For experimenters, we invited a master student who has experience as a presenter in academic conferences, and some other master students who have experience as audience members in academic conferences. In both the experiments, the presenter was required to use the same slides which he or she had used in his or her academic conference experience. The time limit was 10 minutes as in the real academic conference that he or she had experienced. At the same time, we invited 17 audience members for the first experiment and another 10 members for the second one. In addition, to emphasise the test effects of the function of handling the understanding of the audience members by the presenter, before experiment B, we explained not only the instructions for using the system but also some tips for promoting interaction, such as:

For the presenter, it is desirable to:

1) Finish the presentation in 10 minutes.
2) Provide some time for the audience members to contribute comments.
3) Discuss the contributed comments while the audience members are editing.

For the audience members, it is desirable to:

1) Contribute memos or comments of their thoughts during the presentation.
2) Contribute the meaning of comments clearly to make sure that other audience members understand what they want to express.
3) Contribute as many comments as they can.
4) Contribute comments when the presenter is talking about the contributed comments, and edit memos at any time.

In addition, we conducted a questionnaire survey after the presentation to gather data on the experimenter and the usability of this system.

3.2 Results and Discussion

In this experiment, the first presentation was completed in 9 minutes and 54 seconds, and the second one was completed in 9 minutes and 56 seconds. Both these timings were close to the limit of 10 minutes (see Table 2). In addition, according to the results of the questionnaire survey, we understood that the presenter could control the pace of the presentation by reminders from the timer and the progress of slides. It is observed that the function of displaying the timer and the presentation pace helps in supporting the presenter complete the presentation in a limited time.

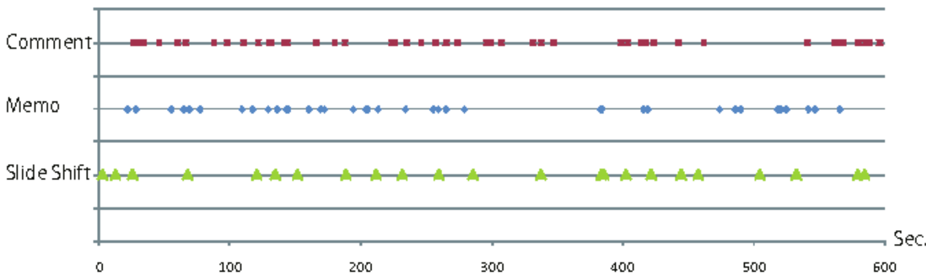


Fig. 6. Comment, memo and slide shift in the time-line of experiment A

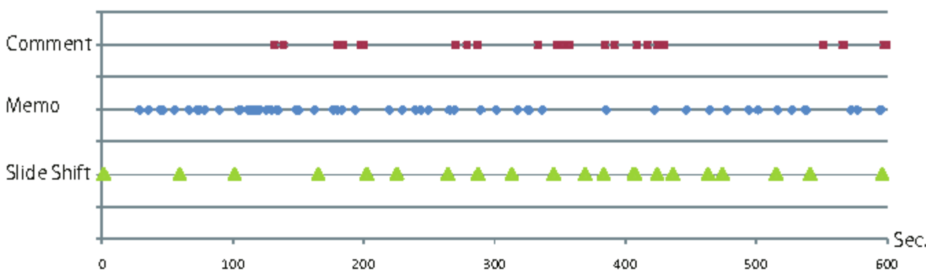


Fig. 7. Comment, memo and slide shift in the time-line of experiment B

In both the experiments, for the presentations which dealt with the development of a convenient page-turning tool for e-book, we received both the memo contributions and the comment contributions(see Table 8). As shown in Figures 6 and 7, the audiences submitted memos during the entire presentation, while comments were contributed almost always around the time when the slides were changed. Comparing the two experiments, experiment A and experiment B, we observed that the comments contributed by the audience members gathered around the presentation time in experiment B. Furthermore, we compared the average number of comments per audience and observed that the audience members contributed more memos in experiment A than in experiment B, the approach of which was stepped up. However, from the average number of memos (2.59 in A and 5.6 in B), a large margin was observed between A and B because the comments were contributed around the presentation time and because more memos were submitted in experiment B. Although we understood that there was not sufficient time for the audience members to contribute comments, on the contrary, to a certain extent this proved that experiment B provided a possibility to make the audience pay more attention when listening to the presentation; we plan to continue to conduct experiments to prove this point. Based on the above analysis, we have come to the conclusion that the memo function provides a possibility of intensively contributing comments using the presentation time during page turning.

Table 1. Differences between memo and comment

	Memo	**Comment**
Purpose	To seize the fleeting ideas coming across the mind in the presentation by recording events or observations.	To add more information, observations or statements on the basis of memo.
Characteristic	Brief and clear.	Complete and substantial.
Function	As a reminder for audiences to help recalling the thoughts formulated during previous listening.	As a feedback to show audiences' opinion to other companions and the presenter as well.
Influence	Having little influence for audiences for consuming short time.	Having a comparatively great effect for audiences for consuming long time.

Table 2. Number of comments, memos and audiences, and presentation spent time in experiment

Experiment	**Number of comments**	**Number of memos**	**Number of audiences**	**Presentation spent time**
A	52	44	17	9:54
B	25	56	10	9:56

As is shown in Table 4 and Table 5, by analysing the timing of comment submission, we understood that the proportion of the comments to the current slide in A is 71% and that in B is 56%. These results prove that this system is effective in collecting real-time comments. In both experiments A and B which use the same presenter and the same slides, there are six slides which keep a margin on more than two comments

Table 3. Investigation of timing on comment submission

Experiment	**Comments of the current slide**	**Rate of comments**	**Comments of the previous slide**	**Comments of the two previous slide**
A	37	0.71	9	6
B	14	0.56	3	8

Table 4. Investigation of timing on comment submission

Experiment	**Comments of the current slide**	**Comments of the comments**	**Comments of the two previous slide**
A	37	9	6
B	14	3	8

Table 5. Rates of comment submission

Experiment	**Rate of comments**
A	0.71
B	0.56

For e-book reader, turning over the pages is achieved by button operation at the present time. Physical paging operation which is performed by paper medium in reading sentences is not equipped. This system perceive the reader's paging operation and synchronize with visual paging expression by sticking transparent sheet and magnetism sensor on existing e- book reader. As a result, reading from e-book reader is expected to be easier.

(a) Content of the experimental presentation

Memo examples	Comment examples
Plastic?	How to fix the plastic sheet on the touch panel?
Magnetic sensor?	Why using magnetic sensor?
Different?	Compare with other book-leader, what is different?
Physical operation?	What's Physical operation?

(b) Memo contributions and comment contributions

Fig. 8. Examples of presentation content, memos and comments

being submitted. Also, on five pages in the six slides, the presenter spent 17.6 seconds on average which was perceived to be quite long. The above analysis proved that the presenter can handle the pace of presentation consciously by means of reminders related to the condition of comment submission of the audience members.

4 Conclusion and Limitation

The purpose of our paper is to determine the way of enhancing effective communication in a realistic presentation scenario. In order to achieve this aim, we focussed on enhancing the motivation of audience members to contribute their comments in a limited time period. We developed a presentation support system which collects comments in a time-limited way through memo contribution and collection. To prove the feasibility of this system, we conducted an experiment. We found that it was feasible, under time-limited conditions, for both the presenter and audience to understand the progress of the presentation. It was also viable to provide reminders of the pace of presentation to both the presenter and the audience. Therefore, this presentation supporting system effectively provides the possibility of submitting comments on the presentation in a time-limited way. In this paper, we emphasise on to the way of increasing the quantity of comments. In our future work, we will shift our focus of research on the ways to increase the number of effective comments and to improve the quality of comments in order to gather more useful feedback.

Acknowledgment. Our research is partly supported by a grant-in-aid for Scientific Research (B) (20300046) from the Ministry of Education, Culture, Sport, Science and Technology of Japan.

References

1. Miura, M., Kunifuji, S., Sakamoto, Y.: Practical Environment for Realizing Augmented Classroom with Wireless Digital Pens. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part III. LNCS (LNAI), vol. 4694, pp. 777–785. Springer, Heidelberg (2007)

2. Kurihara, K., Goto, M., Ogata, J., Matsusaka, Y., Igarashi, T.: Presentation Sensei: A Presentation Training System using Speech and Image Processing. In: Proc. of ACM ICMI International Conference on Multimodal Interfaces, pp. 358–365 (2007)
3. Kobayashi, T., Nishimot, K.: An Enhanced Chat System That Allows Cross-Media References. In: WISS 2009, pp. 101–106 (2009) (in Japanese)
4. Doi, K., Hirashima, D., Takagi, M., Mochizuki, M., Teshigawara, Y.: Development of a Lecture Supporting System to Collect Comments using Various Media. In: WBE 2009, pp. 294–300 (March 2009)
5. Doi, K., Inoue, M., Tasaka, M., Hiroshima, D., Teshigawara, Y.: A proposal of an Annotation System CollabSticky Focused on Collecting Comments in Presentation. In: DICOMO 2007, pp. 159–164 (July 2007)
6. Davison, R., Briggs, R.: GSS for presentation support. Commun. ACM 43(9), 91–97 (1993)
7. Nishida, T., Igarashi, T.: Bringing Round-Robin Signature to Computer-Mediated Communication. In: ECSCW 2007, pp. 219–230 (September 2007)
8. Nishida, K., Kurihara, K., Goto, M.: On-Air Forum: A System for Facilitating Communication While Watching Real-time Content. In: WISS 2009, pp. 95–100 (2009) (in Japanese)
9. Akihiro, H., Madoka, Y., Hiroyuki, T., Yahiko, K.: Experimental Chat Sessions to Encourage Active Communication in Lectures. IEIC Technical Report 100(91), 7–12 (2000)
10. Kobayashi, T., Nishimot, K.: Temporal Relations between Text Messages and Spoken Utterances in A Face-To-Face Communication with A Text Chat. IPSJ 2006(123), 2006-HI-121(10), 67–74 (2006)

Experiments in Automatic Assessment Using Basic Information Retrieval Techniques

Md Maruf Hasan

School of Technology, Shinawatra University, Thailand
maruf@siu.ac.th

Abstract. In *Information Retrieval* (IR), the similarity scores between a query and a set of documents are calculated, and the relevant documents are ranked based on their similarity scores. IR systems often consider queries as short documents containing only a few words in calculating document similarity score. In *Computer Aided Assessment* (CAA) of narrative answers, when model answers are available, the similarity score between Students' Answers and the respective Model Answer may be a good quality-indicator. With such an analogy in mind, we applied basic IR techniques in the context of automatic assessment and discussed our findings. In this paper, we explain the development of a web-based automatic assessment system that incorporates 5 different text analysis techniques for automatic assessment of narrative answers using vector space framework. The experimental results based on 30 narrative questions with 30 model answers, and 300 student's answers (from 10 students) show that the correlation of automatic assessment with human assessment is higher when advanced text processing techniques such as Keyphrase Extraction and Synonym Resolution are applied.

Keywords: Computer Aided Assessment (CAA), Information Retrieval (IR), Intelligent Text Analysis, and Natural Language Processing.

1 Introduction

With the advent of WWW and ubiquitous access to huge amount of information, research in Information Retrieval (IR) gained new momentum in the recent years. Several new techniques and tools are developed to aid the IR tasks. In the context of IR, the similarity scores between a query and a set of documents are calculated, and the relevant documents are retrieved and ranked based on their similarity scores. IR systems often consider queries as *short* documents containing only a few words. Similarity scores between documents are the indicators of *relevance*. In Computer Aided Assessment (CAA) of *narrative* answers, when model answers are available, the similarity score between the Students' Answers (SA) and respective Model Answer (MA) may be a good indicator of *quality*. With such an analogy in mind, we applied basic IR and text analysis techniques in the context of automatic assessment, and discussed our findings. In this paper, we explain the development of a web-based automatic assessment system that incorporates 5 different text analysis techniques for automatic assessment of narrative answers using a vector space representation.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 13–21, 2011.

Online education is gaining widespread interest mainly because of the flexibility it offers in terms of scheduling, effective usage of teaching time and learning pace. Therefore, assessment of students' understanding in the online environment is becoming a major concern. Many researchers are paying attention to how to improve the online assessment system and which techniques should be adopted in an online assessment system. Most online assessment systems resort to just True/False, Multiple Choice, and Fill-in-the-Blank type of questions due to their simplicity (direct matching). How-ever, it is a general opinion in the field that such objective tests *without* narrative questions is *not* sufficient in fully evaluating the knowledge of a learner. Assessment of non-narrative answers is straight forward. However, the proposed online assessment system aims to integrate assessment of objective questions and narrative questions *together* to assess comprehensive knowledge and understanding of the learners.

Automatic assessment of narrative answers is a field that has attracted attentions to online education community in the recent years. Several systems have been developed under academic and commercial initiatives using statistical, Natural Language Processing (NLP), and Information Extraction (IE) techniques. Some researchers also used clustering, semantic networks and other hybrid approaches [1]. In our research, we experimented with 5 different text analysis techniques (typically used for basic IR tasks) for automatic assessment of narrative answers using the *vector space framework*. We use Uni-gram, Bi-gram, TF.IDF, Keyphrase Extraction, and Keyphrase Extraction along with Synonym Resolution [2][3][4] to represent model answers and student's answers as vectors; and then similarities between the model answer and the students' answer are computed by using cosine similarity [5].

The Web-based online assessment system integrates the assessment of both objective and narrative answers. The system is easy-to-use and modular in design - new assessment modules can be added easily and existing ones can be revised as necessary. The system allows teachers to set test papers manually, or automatically (according to specific criteria from a question-bank); and students to take the test online. The automatic assessment may be performed instantly using any or all of the proposed text analysis techniques. The automatic scores may also be validated and finalized by the teacher.

2 Brief Overview of the Online Assessment System

We focused on developing a Web-based online assessment system that is modular and easy-to-use for both teachers and students. The system is developed using ASP.NET and Microsoft SQL Server. Some key features of the system are explained in this section.

Teacher: The teacher interacts with the system by entering questions along with their model answers. The teacher can also set the test paper, exam schedule and define the full marks for each question. For the assessment of narrative answers, the teacher may choose one or more text analysis techniques. Currently, we incorporated 5 text analysis modules which will be discussed later. For certain approaches (such as Keyphrase Extraction), the teacher also needs to train a model automatically using available machine learning algorithms.

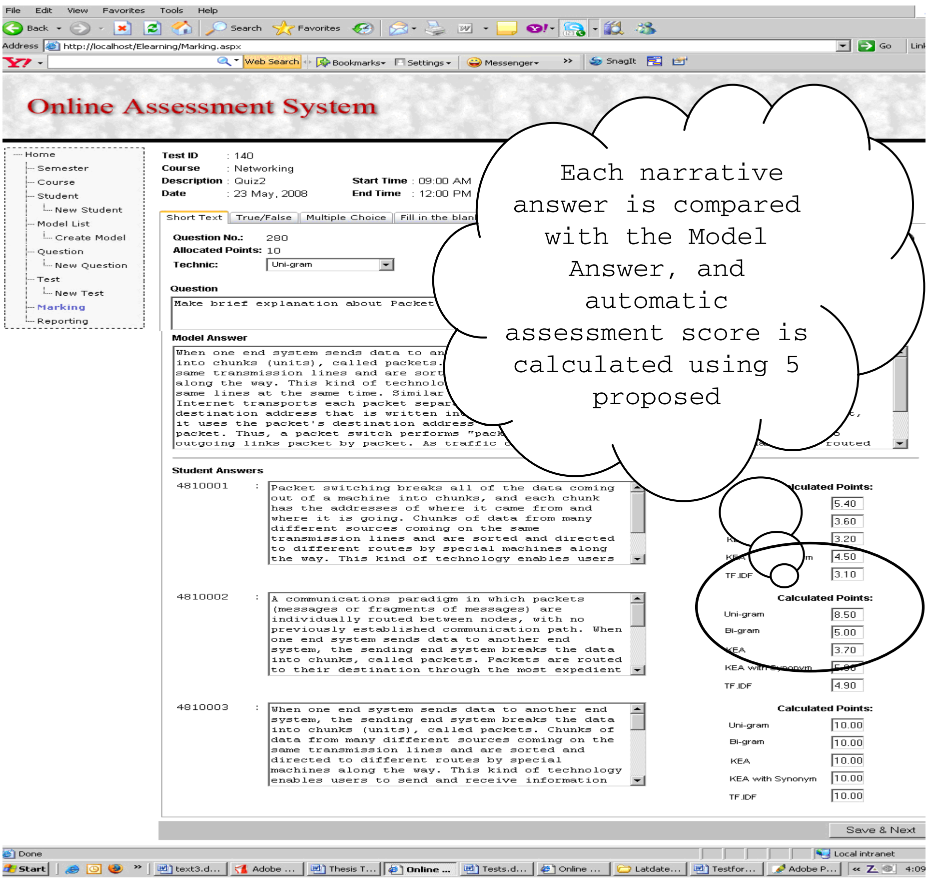


Fig. 1. Similarity Scores of Student's Answers Calculated for all 5 Proposed Techniques

Student: The students have to register (or authenticate) themselves to take the test by giving their personal details (or password). Upon authentication, they need to choose from the available courses to get the relevant test paper and answer each question from the test online. The student sessions end with submission of their answers. It is possible to restrict the duration of a test.

System: The system stores several crucial in-formation in the backend database, which include students' information, course information, set of questions associated with each course, customized tests composed by the teachers (manually and automatically). Automatic selection of test question from the so-called question-bank can be random or filtered by some criteria (such as level of difficulty).

Based on the assessment techniques chosen, the system makes comparison between Stu-dents' Answers (SA) and the respective Model Answers (MA); and calculates the cosine similarity score. Figure 1 shows the similarity scores of a narrative question using all five proposed techniques for Teacher's perusal and validation.

The online system also generates report for each test-taker which includes course and student information and score for that test as validated by the teacher. Nevertheless, the system archives *historical data* (e.g., students' answers along with teacher validated scores), which may be further analyzed to augment the quality of automatic assessment through incremental and adaptive machine learning techniques.

3 The Proposed Text Analysis and IR Techniques

In this research, we experimented with *five* different text analysis and IR techniques for automatic assessment of narrative answers using vector space framework. They include Uni-gram, Bi-gram, TF.IDF weight of word, Key-phrase Extraction, and Keyphrase Extraction along with Synonym Resolution. We represented both model answers and student's answers as vectors; and then similarities between the model answer and the students' answer are computed by using cosine similarity. For each approach, some preprocessing was necessary.

Document Pre-processing: The major preprocessing steps include tokenization, stop-word removal [6] and stemming [7]. For our experiment setup, document pre-processing is performed before applying Uni-gram, Bi-gram and TF.IDF based approaches [8]. Additional preprocessing was not necessary for KEA based approaches since KEA has stemmer and stop-word removal built-in. For the *Keyphrase with Synonym Resolution* approach, we also use a thesaurus for synonym resolution.

TF.IDF Weights: The TF.IDF weight is a weight often used in information retrieval and text mining in term weighting. The TF.IDF weight weighs each word (or phrase) in the text document according to how unique it is [9]. The similarity score between the query and the documents in a document collection shows how *relevant* a document is for that particular query. We also used TF.IDF weights[1] in calculating

[1] In the context of IR, IDF weighs *rare* terms highly, hence *multiplied* with TF. In the context of automatic assessment, where *Model Answer* is modeled as a *Query* against a set of Students Answers, apparently, TF/IDF weights (unlike TF.IDF) appear to be more appropriate (i.e., *common* terms should get more attention in an optimistic assumption that majority of students would be using similar terminologies to answer narrative question). However, in our experiments, we noticed that TF/IDF weights did not perform as expected perhaps due to the small document-set involved.

term weights. The similarity score between the model answer and a student answer is supposedly indicative to the *quality* of the students answer.

In our experiment, the term frequency (TF) and inverse document frequency (IDF) weights are calculated using the following formula:

Term Frequency (TF):

$$TF_{i,j} = \frac{f_{i,j}}{\max_j f_{i,j}} \tag{1}$$

where $f_{i,j}$ = frequency of term i in collection j. Collection consists of 11 documents (1 MA and 10 SAs).

Inverse Document Frequency (IDF):

$$IDF_t = \log(1 + \frac{N}{n_t}) \tag{2}$$

where N = 11 documents consisting of 1 MA and 10 SAs; and n_t = the number of documents containing the term, t.

Keyphrase Extraction Algorithm: KEA [2] is a keyphrase extraction tool based on supervised machine learning techniques [3][4][10]. For the fourth and fifth approach in this experiment, we use KEA to extract key-phrases from MA and SA to make similarity comparison.

KEA's Keyphrase extraction algorithm has two phases. In the training phase, KEA creates a keyphrase extraction model for identifying key-phrases, using training documents (documents with manually assigned keyphrases). We use 50 documents from *computer network domain* to train KEA. Manually assigned keyphrases were available for these training documents. In the extraction phase, using the trained model, we extract keyphrases from the students' answers and model answers. More details about KEA, including the features and algorithms can be found in [2][3]. In summary, KEA uses high level linguistic features and supervised learning with or without a thesaurus. Desired numbers of keyphrases to be extracted automatically can also be specified in KEA.

The Vector Space Framework: The Vector Space Model (VSM) [5] is a flexible representation framework for documents and queries to be represented as vector of terms. Terms could be single word, bi-gram and key-phrase etc. Term vectors may contain Boolean term-weights (1 and 0 to represent presence or absence of a term) as well as complex weighting schemes (such as TF.IDF). The Boolean weighting scheme is not suitable for partial matching. Therefore, we used frequency-based weighting schemes; and the similarity between MA and SA vectors is measured using cosine similarity [9] as follows:

$$sim(SA_i, MA) = \frac{SA_i . MA}{|SA_i||MA|} = \frac{\sum_j w_{i,j} \ X \ w_{MA,j}}{\sqrt{\sum_j w^2_{i,j}} \sqrt{\sum_j w^2_{MA,j}}} \tag{3}$$

4 Experiments and Evaluations

In our experiment, we used 30 narrative questions with 30 model answers (MA). A group of 10 students attempted all 30 questions and therefore, student's answer set (SA) consists of a total of 300 documents. For the first three approaches (i.e., *Uni-gram, Bi-gram* and *TF.IDF*), we preprocessed both MAs and SAs and represent them as respective MA and SA vectors. Similarity scores are then computed using co-sine measure. For the last two approaches (i.e., *Keyphrase Extraction* and *Keyphrase with Synonym Resolution*), we trained a KEA keyphrase extraction model using a set of 50 domain-specific documents with manually annotated keyphrases. Keyphrase vectors for MA and SA are computed with or without synonym resolutions. As before, we measure cosine similarity between SA vectors with the relevant MA vector.

For each 300 student's answer, human assessments were available. The correlation between human assessment and automatic assessments using all 5 approaches are calculated using *Pearson's Correlation* as follows:

$$Correlation(x, y) = \frac{CoVariance(x, y)}{StdDev(x) \times StdDev(y)} \tag{4}$$

The correlations between human and automatic assessment for 30 questions using all 5 automatic assessment approaches are presented in Figure 2.

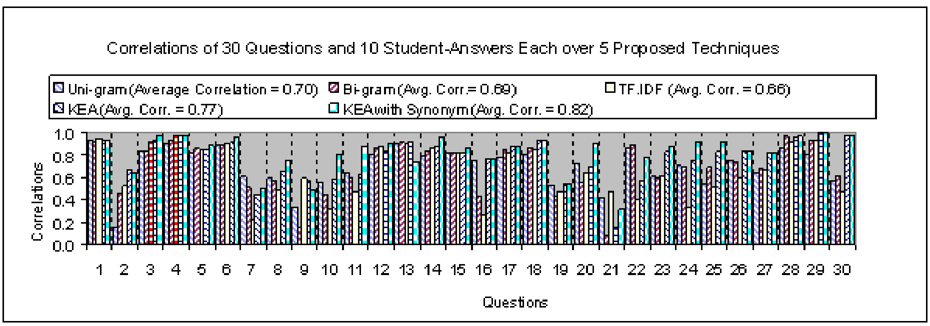


Fig. 2. Correlation of 5 Proposed Approaches with Human Judgment (30 Questions)

As shown in Figure 2, the highest average correlation is achieved with KEA with synonym resolution. This is due to the fact that KEA uses high-level features and synonym resolution that captures differences in term usage. The correlation value for Uni-gram is relatively high. However, in a VSM representation (i.e., a *bag-of-word* representation), Uni-gram based assessment may be greatly vulnerable to abuse.

In general, the TF.IDF weights for terms are accurately estimated from a huge dataset (collection), and therefore, from our experimental setup it is hard to make any justifiable comments about its performance in CAA due to small dataset used. Neither TF.IDF nor TF/IDF weight showed any better correlations.

Although it is not directly comparable, for reference purpose, we present a table showing the correlation values of other existing systems [11] and those of ours in Table 1.

Table 1. Average Correlations of existing systems along with our proposed approaches

System	**Techniques**	**Correlation**	**Language**
AEA	LSA, PLSA or LDA	0.75	English
Apex Assessor	LSA	0.59	French
Automark	Information Extraction	0.59	English
IEMS	Pattern matching	0.80	English
Jess	Pattern matching	0.71	Japanese
MarKLt	NLP, Pattern matching and statistics	0.75	English
PEG	Linguistic features	0.87	English
SEAR	N/A	0.45	English
Atenea	Statistical, NLP and LSA	0.56	English
Uni-gram	*N-gram word*	**0.70**	*English*
Bi-gram	*N-gram word*	**0.69**	*English*
TF.IDF	*TF.IDF weight*	**0.66**	*English*
KEA	*Keyphrase Extraction*	**0.77**	*English*
KEA with Synonym	*Keyphrase+ Synonym*	**0.82**	*English*

The average correlations of our approaches are listed at the bottom 5 rows of Table 1 (c.f., Fig. 2). The average correlations for all 5 approaches are significant in comparison with the existing systems. The proposed approaches are rather *straight-forward* implementation of IR techniques for CAA task to justify our key assumption about the similarity between IR and CAA tasks. Therefore, based on our experimental result, we argue that it is worthy to investigate the proposed approaches further using additional refinements deemed suitable for the purpose of automatic assessment. New algorithms using other IR techniques (e.g., Latent Semantic Analysis) can also be tried. In our implementation of the online assessment sys-tem, each technique is implemented as a separate module that is easy to revise. Moreover, it is also easy to incorporate new assessment modules in the system due to its modular architecture.

5 Conclusions and Future Work

We experimented with a set of 30 narrative questions and 300 students' answers using 5 different automatic assessment techniques. The average correlations between human assessment and automatic assessments for each approach are comparable with those of the existing systems. We observe that *KEA with Synonym Resolution* yields the highest correlation. For automatic assessment it is necessary to integrate high-level linguistic features. Similar observations are also prevalent in Information Retrieval tasks. The advances in natural language processing will certainly have implications in modeling IR as well as Computer Aided Assessment of narrative answers.

The online assessment system outlined here is based on a modular design. Therefore, the existing approaches can be easily augmented further. New approaches can also be integrated with ease. Moreover, the online assessment system allows us to archive *historical data* (students' answers along with human validated grades). For TF.IDF weighting scheme a larger dataset is desirable. Furthermore, supervised machine learning algorithms on such historical datasets may help us exploring some *adaptive* assessment algorithms in future. The idea behind *relevance feedback* mechanism employed in some IR systems is to take advantage of human relevance judgments as feedback in reformulating the query for better retrieval results. In the context of CAA, such feedback mechanisms should be experimented with the help of the teacher-validated *historical dataset*.

It should also be noted that automatic assessment is thus far not a substitute of humans with computers, but an aid to help the teachers to cut down their time or to cross-check their own assessment which may be prone to inadvertent human errors or omissions. The online automatic assessment system we developed is therefore practical and useful. We wish to refine our algorithms to incorporate some high-level linguistic features using state-of-the-art natural language processing and machine learning techniques [12][13].

Acknowledgement. The author would like to thank his graduate student, *Mrs. Ohmar Thwin* for her help in data preparation, experiments and system implementation as part of her M.Sc. thesis.

References

1. P'erez, D., Postolache, O., Alfonseca, E., Cristea, D., Rodriguez, P.: About the effects of using Anaphora Resolution in assessing free-text student answers. In: Proceedings of Recent Advances in Natural Language Processing Conf., Borovets, Bulgaria, pp. 380–386 (2005)
2. Keyphrase Extraction Algorithm (KEA), `http://www.nzdl.org/KEA/`

3. Frank, E., Paynter, G.W., Witten, I.H., Gutwin, C., Nevill-Manning, C.G.: Domain-specific keyphrase extraction. In: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, pp. 668–673. Morgan Kaufmann, California (1999)
4. Witten, I.H., Paynter, G.W., Frank, E., Gutwin, C., Nevill-Manning, C.G.: KEA: Practical Automatic Keyphrase Extraction. In: Processing of the Fourth ACM Conference on Digital Libraries, pp. 254–255. ACM Press, Berkeley (1999)
5. Salton, G., Wong, A., Yang, C.S.: A Vector Space Model for Automatic Indexing, Communications. The ACM 18(11), 613–620 (1975)
6. Jarmasz, M., Szpakowicz, S.: Roget's thesaurus and semantic similarity. In: Processing of the International Conference on Recent Advances in Natural Language Processing, Borovets, Bulgaria, pp. 212–219 (2003)
7. Porter Stemming, `http://www.comp.lancs.ac.uk/computing/research/stemming/general/porter.htm`
8. N-gram Models, `http://en.wikipedia.org/wiki/N-gram/`
9. Zhang, Y., Gong, L., Wang, Y.: An improved TF-IDF approach for text classification. In: Network & Information Center, School of Electronic & Information Technology, pp. 49–50. Shanghai Jiaotong University, Shanghai (2004)
10. Weka Machine Learning Suite, `http://weka.sourceforge.net/`
11. Pérez, D., Alfonseca, E., Rodriguez, P.: About the effects of combining Latent Semantic Analysis with natural language processing techniques for free-text assessment. Revista Signos 38(59), 325–343
12. Larkey, L.S.: Automatic Essay Grading Using Text Categorization Techniques. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Melbourne, Australia, pp. 90–95 (1998)
13. Rosé, C.P., Roque, A., Bhembe, D., VanLehn, K.: A hybrid text classification approach for analysis of student essays. In: HLT-NAACL Workshop on Building Educational Applications Using Natural Language Processing, Edmonton, Canada, pp. 68–75 (2003)

Relevant Piece of Information Extraction from Presentation Slide Page for Slide Information Retrieval System

Tessai Hayama and Susumu Kunifuji

1-1, Asahidai, Nomishi, Ishikawa, Japan
{t-hayama,kuni}@jaist.ac.jp
http://www.jaist.ac.jp/~t-hayama/

Abstract. An approach to create a slide-data reuse system is to provide minimum information needed from a slide page to meet users' needs. Previous techniques for slide data processing have not handled each piece of information within a slide page accurately: some ruled out figures or tables, and some yielded meaningless text strings in which incomplete short sentences were connected. Thus, it is necessary to recognize the role of each piece of information to select strongly related pieces of information, and then to display them properly. In this paper, we describe a method that extracts relevant information from a slide page for effective search.

Keywords: Human Interface, Information Extraction, Presentation Slide, Information Searching.

1 Introduction

The widespread use of electronic presentations is increasing the number of slides that businesses accumulate. Since these slides are often stored and reused as e-learning or web content, data stored on slides are rapidly becoming one of industry's most important information resources. Therefore, it is necessary to develop a practical method for reusing these data.

One approach to create a slide data reuse system is to provide minimum information needed from a slide page to meet users' requirements. If possible, a slide search engine should provide a list of results including more slide pages with more legible slide information, so that users can easily find slide data that meets their requirements.

Previous studies of techniques for slide searching have mainly examined search algorithms and methods of presenting the search results in research domains such as database engineering and educational engineering. For example, Min[5] developed a method of aligning a paper's sections and its slide pages and proposed using sentences from a paper with a corresponding slide page image in each search result. Guo et al.[2] developed a slide search system using image retrieval. Yokota et al.[6] and Okamoto et al.[7] developed a multimedia database system

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 22–31, 2011.

for educational materials. Their systems provide a slide search function that uses text data within a slide page as an index for lecture video-clip retrieval. Because previous techniques for slide data processing have converted pieces of information within a slide page into simple character string(s) or a page image, systems adopting these techniques could not deal accurately with the pieces of information. Thus, an approach is needed that recognizes the role of each piece of information, selects strongly related pieces of information, and then displays them properly. Some systems ruled out figures and tables within a slide page, and some yielded meaningless text strings consisting of connected incomplete short sentences.

In this paper, we describe a method that extracts relevant pieces of information from a slide page for effective slide information searching. The proposed method presents the extracted pieces of information according to certain presentation regions with the objective of providing data on the content of each slide page accurately and compactly.

2 Approach

We consider a typical keyword search system in which a keyword is input and slide pages including keyword-matched sentences are output as items on a list of search results. In this section, we explain information included in a slide page and then describe our approach, which extracts relevant pieces of information from a slide page.

2.1 Information within a Slide-Page

A slide created by presentation software includes information consisting of primitive objects, such as text, pictures, lines, and basic diagrams. The primitive objects are categorized as the title, body text, figures, tables, and decorative elements, which provide a basic representation that supports understanding.

Slide pages usually include a title that summarizes the content; objects other than the title are laid out for easy understanding of the content. The layout generally represents relationships between the pieces of information within a slide page using position and distance. For example, indentation is often used to represent a hierarchical relationship between objects; the same phrases can be seen differently, for example, as a sentence or as a part of a figure, depending on their positions. The meaning of an incomplete sentence may be clarified by the sentence's relationship with other piece(s) of information.

2.2 Toward Extracting Relevant Pieces of Information

As described above, information within a slide page is generally laid out in basic categories such as title, body text, figures, and tables. The layout is for a large-scale display. Certain pieces of information should be extracted from a slide

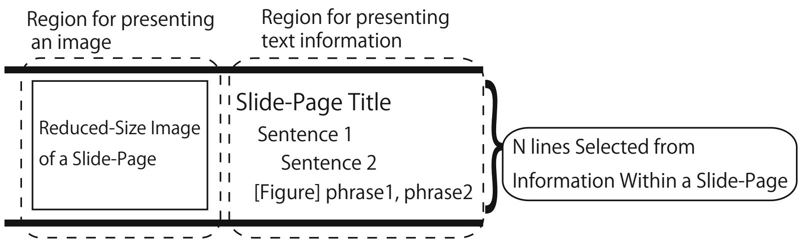


Fig. 1. Proposed interface for information presentation of a slide page

page and then restructured considering the layout on the page and the size of the display region.

In this restructuring, the relationships between the extracted pieces of information must be maintained, because it often is difficult to understand an isolated piece of information within a slide page. Providing other piece(s) of related information can support understanding.

When information within a slide page is presented, it must be legible. It is almost impossible to present all information within a slide page in a small region, especially figures and tables. The piece of information most strongly associated with user requirements should be maintained as legible as possible. Also, it is often useful to estimate the content of a slide page from certain characteristics of the page, such as the amount of information it contains and the design of the page. Thus, the extracted pieces of information should include legible text and make users aware of such meta-information.

3 Proposed Method

We propose a method that extracts information from a slide page according to the size of the presentation region and presents the extracted information in an interface. We explain the information presentation interface and information selection method in sections 3.1 and 3.2, respectively. In addition, we explain the benefits of the proposed method.

3.1 Information Presentation Interface

We propose the interface shown in Figure 1 to present pieces of information from a slide page.

This interface includes regions for presenting an image and for presenting text. The region for presenting an image is assigned a reduced-size image of a slide page. The image size is determined by the number of lines of text in the other region. The region for presenting text is assigned selected pieces of information from within the slide page. The title sentence of the slide page is assigned to the first line, and to the remaining lines, pieces of information described by the attributes “sentence,” “figure,” and “table” are assigned. If information with the attribute "sentence" includes a long character string, the number of characters is reduced by a rule: if a retrieval key phrase is included, a certain number of characters around the key phrase are clipped; otherwise, a certain number of characters from the beginning are clipped. Information with the attribute "figure" or "table" is replaced by an attribute name ("[Figure]" or "[Table]") at the head of the line and a character string is included after the attribute name. Thus, information in character strings included in the region can remain legible. In addition, pieces of information in the region are presented using indentation to clearly represent the relationships between them.

In the next section, we explain a method for selecting certain pieces of information from a slide page while maintaining the relationships between them.

3.2 Information Selection Method

The information selection method calculates the selection priority of pieces of information within a slide page and then selects a certain amount of information from a slide page according to the priority. We believe that not only retrieval key phrases but also other piece(s) of related information should be selected to facilitate accurate understanding of information. We use the relationships between pieces of information as revealed in the layout structure to select such information. We consider that piece(s) of information in upper level of the structure from target information support to discern the context that the target information is used in. And also, pieces of information in the lower level of the structure from the target information may support to reveal details of the target information.

The procedures for information selection are as follows.

- **Step 0)** As preprocessing, structure pieces of information within a slide page extracted using a structure extraction tool[3], which assigns attributes such as "title," "sentence," "figure," and "table" to each object within the slide page, and then arranges each object in a hierarchical structure. The root node of the hierarchical structure is assigned to the object with the attribute "title."
- **Step 1)** Select a piece of information in the highest level of the hierarchy from among those that include the retrieval key phrase. As pieces of information are selected, the selection order priority is to the right.
- **Step 2)** Focus on information in the next lower level in the hierarchy.
  2-1) Select the rightmost piece of information from those one level below each piece of information selected in the previous step.
  2-2) Select the rightmost piece of information from those one level below each piece of information not selected in the previous step.
  2-3) Repeat the selection process of step 2-2 until selections are made at one level below all piece(s) of information selected in the previous step.
- **Step 3)** Select piece(s) of information through the root node of the structure from the piece of information that includes retrieval key phrase, if the piece of information including retrieval key phrase is not in the root node.
- **Step 4)** Go back to position of information focused in step 2 and then repeat the selection process of step 2 until no information remains to be selected.

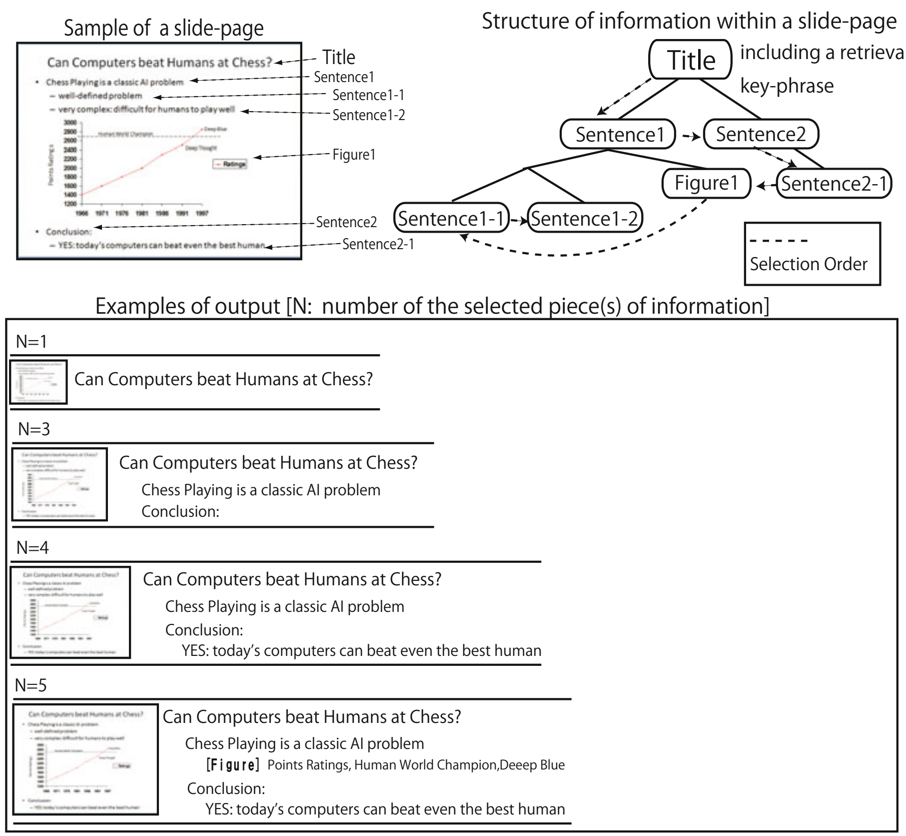


Fig. 2. Examples of the application of the proposed information selection method

– **Step 5)** Select unselected piece(s) of information at the hierarchical level of step 1. If pieces of information are selected, the selection order priority is to the right.
– **Step 6)** Repeat steps 2, 3 and 4 until no information remains to be selected.

The selection order of this method indicates selection priority. A certain number of pieces of information within a slide page are extracted, and the relationships among them are retained according to the selection order.

Figure 2 shows examples illustrating the application of the information selection method. In the first step, using the slide information structure extraction tool, information within a slide page is represented as a hierarchical structure, the root of which is assigned to the piece of information with the attribute "Title." In step 1, "Title," which includes the retrieval key phrase, is selected. Step 2 focuses on pieces of information one level below the "Title" hierarchy; "Sentence1" is selected in step 2-1, and "Sentence2" is selected in 2-2. Since the piece of information including retrieval key phrase is assigned to the root node. step 3

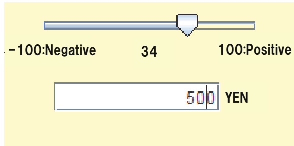


Fig. 3. Inteface for evaluating a slide-page content based on WTP and EU

is skipped. Step 4 goes back to focus point of the structure focused in the step 2 and then focuses on pieces of information in the next level in the hierarchy, and the procedures of step 2 are performed; "Sentence2-1" is selected in step 2-1 and "Figure1" is selected in step 2-2. Then step 2 is applied again, focusing on pieces of information in the next level in the hierarchy. The procedures of step 2 are performed; "Sentence1-1" is selected in step 2-1, and "Sentence1-2" is selected in step 2-2. Since there is no information in the next level of the hierarchy to select, steps 5 and 6 are skipped, and information selection is complete. The resulting selection priority for selecting pieces of information from the slide page is in the order "Title," "Sentence1," "Sentence2," "Sentence2-1," "Figure1," "Sentence1-1," and "Sentence1-2."

3.3 Expected Effects

We expect the following results of applying the proposed method to slide data.

- Because slide page images as well as text data are presented, users are aware of the presence of figures/tables and the amount of information within each slide page, and hence, they can grasp the content more accurately.
- Because a piece of information including a key phrase is presented along with related information, maintaining the relationships between the pieces of information, users can understand text more easily.

4 Experiment

4.1 Purpose and Procedures

The proposed method provides information for more accurate understanding of the content of a slide page in a certain area on the screen. In our experiment, we investigated the usefulness of the proposed method by comparing it with the presentation of simple text. The comparison was achieved by evaluating the content of each slide page. If the presentation method provides a more accurate understanding of slide page content, the content presented by the method can be evaluated more accurately. Thus, we calculated the correlation between the evaluation values assigned to the content of each slide page as acquired by each presentation method and those assigned to the original slides. The larger the correlation is, the more effective the presentation method is for correct understanding of the content.

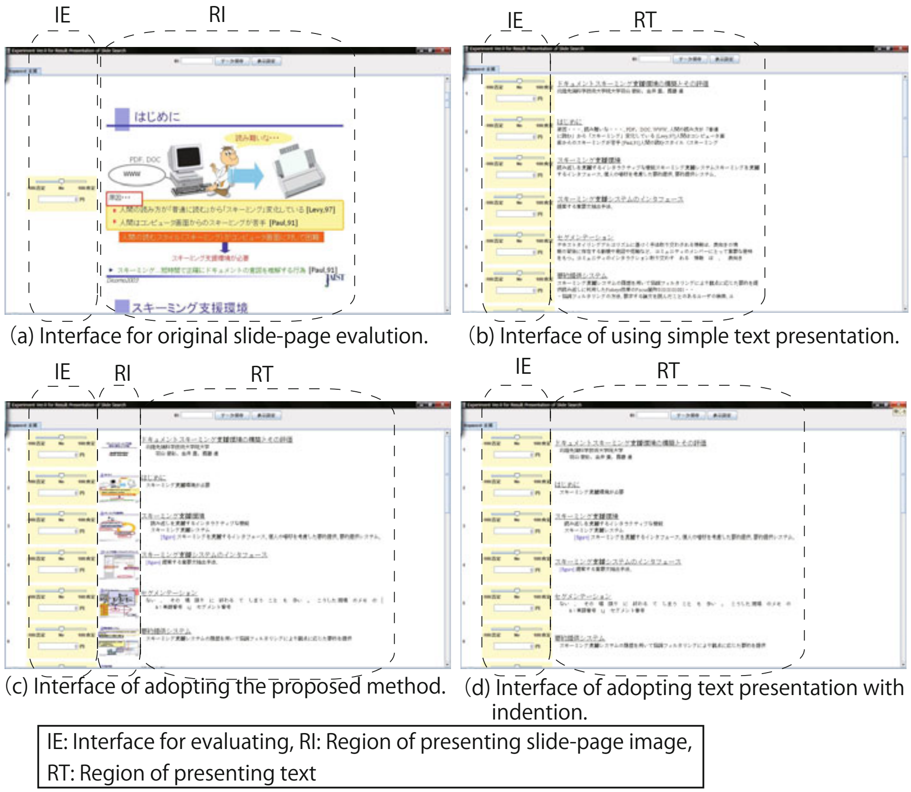


Fig. 4. System interfaces for each the presentation method used in the experiment

We consider that an effective measure is useful for evaluating a slide page's content, since pages include characteristics such as layout and design. We adopted an evaluation measure based on "willingness to pay (WTP)" and "experience utility (EU)[4]." As in a previous study[1], we prepared a text box for entering a payment value and a slider bar for indicating an evaluation value. Users evaluated the content of a slide page using these methods to indicate WTP and EU, as shown in Figure 3. The WTP values of each evaluator were standardized using the following fomula after all slide pages were evaluated to reduce differences between payment values for the evaluators.

$$WTP_{(i)} \mapsto \frac{WTP_{(i)} - \overline{WTP_{(topicA)}}}{s_{(topicA)}}. \qquad (1)$$

Here, $WTP_{(i)}$, $\bar{WTP}_{(topicA)}$, and $s_{(topicA)}$ indicate WTP value of slide-page i, average WTP value of slide-pages of $topicA$, and WTP standard deviation of slide-pages of $topicA$, respectively. The EU values used a slider labeled between -100 and 100.

Table 1. Correlations between the slide evaluation values of each presentation method and those of the original slide

		Simple text presentation	Text presentation based on indention	Presentation of proposed method
CORREL*	WTP	0.46	0.50	0.56
	EU	0.34	0.35	0.42

∗ Correlation with the original slide evaluation value.

Slides were evaluated by sixteen graduate students who often search for information using the Internet. To choose slides for the evaluation, we collected from the Internet 20 slide files each on the topics of grid computing, crowd computing, electronic books, and YouTube. Twenty slide pages for each topic were selected from among the slide files based on the following characteristics: differences of positions within a slide page that includes a retrieval key, inclusion of a figure or table, inclusion of an indented layout, and amount of textual information within a slide page. Before they evaluated the slide data, we explained to the evaluators that there was no time restriction and that each representation of a slide page should be evaluated for its value as material regarding the four topics. The evaluators used the system interfaces for each presentation method shown in Figure 4 to evaluate the content of each slide page. We also included a presentation method using only a text layout based on indentations to confirm the usefulness of presentation layout characteristics. Each evaluator provided data regarding each topic using the three presentation methods and the original slide page, in such a way that topics and presentation methods were counterbalanced as far as possible. The evaluation of each slide-page for each method was averaged values which some evaluators provided. The size of the presentation region included in each interface was set to four lines, which is standard for typical search services.

4.2 Results

The correlations between the slide evaluation values for each presentation method and those for the original slide were calculated as shown in Table 1. The presentation of the proposed method yielded the largest correlations for both WTP and EU among the three presentation methods. Therefore, it is useful to present both text information with a specific layout and a reduced-size image of a slide page to provide a compact presentation that accurately conveys the content of the page.

To investigate effects of the proposal method in the detail, we analyzed slide type with large difference between evaluation of the original presentation and each presentation method. The numbers of a slide page which have positive and negative difference between the original presentation and each presentation method in the WTP are shown in Table 2.

In the experiment, slide pages including figure/table and those including more amount of information than the extracted information were more correctly evaluated by presentation of the proposed method than the other two. And also,

Table 2. Number of a slide page which have positive and negative difference between the original presentation and each presentation method in the WTP

Slide type	STP	TPI	PPM
A) Including figure/table.	9	8	1
B) Including layout-indent.	14	10	10
C) Including more amount of information than the extracted information.	15	14	9
D) Including information, being matched by retrieval key phrase, with attributes except "Title".	13	13	10
E) C and D	10	8	8

STP: Simple text presentation, TPI: Text presentation based on indention, PPM:Presentation of propsed method

slide pages including layout-indent were more correctly evaluated by presentation of the proposed method and text presentation based on indention than the simple text presentation. Thus, it is useful of our approach to provide data on the content of each slide page accurately and compactly. On the other hand, it is not such effective to apply the proposed method to slide pages that include more amount of information than the extracted information and include information, being matched by retrieval key phrase, with attributes except "Title". We should be going to develop a novel approach for improving that.

5 Conclusion

We proposed a method that extracts relevant pieces of information from slide pages to promote efficient searching of information on slides. The proposed method provides a presentation interface that consists of two regions: one for presenting a reduced-size image of a slide page and one for presenting indented text information selected by our information selection method. The experiment showed that our approach provides a compact presentation that facilitated more accurate understanding of the content of slide pages.

In our future work, we will investigate the effects of the proposed method in more detail by gathering more experimental data. Also, we will develop a ranking algorithm for slide information searching based on the experimental data.

Acknowledgment. This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research(B), 20300046, and the Telecommunication Advanced Foundation.

References

1. Arapakis, I., Jose, J.M., Gray, P.D.: Affective feedback: an investigation into the role of emotions in the information seeking process. In: Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 20–24 (2008)

2. Liew, G.M., Kan, M.Y.: Slide image retrieval: a preliminary study. In: Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDLf 2008), pp. 359–362 (2008)
3. Hayama, T., Nanba, H., Kunifuji, S.: Structure Extraction from Presentation Slide Information. In: Ho, T.-B., Zhou, Z.-H. (eds.) PRICAI 2008. LNCS (LNAI), vol. 5351, pp. 678–687. Springer, Heidelberg (2008)
4. Lopatovska, I., Mokros, H.B.: Willingness to pay and experienced utility as measures of affective value of information objects: Users' accounts. Information Processing and Management: an International Journal 44(1), 92–104 (2008)
5. Kan, M.Y.: SlideSeer: a digital library of aligned document and presentation pairs. In: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDLf 2007), pp. 81–90 (2007)
6. Yokota, H., Kobayashi, T., Muraki, T., Naoi, S.: UPRISE: Unified Presentation Slide Retrieval by Impression Search Engine. IEICE Transactions on Information and Systems E87-D(2), 397–406 (2004)
7. Okamoto, H., Kobayashi, T., Yokota, H.: Presentation Retrieval Method Considering the Scope of Targets and Outputs. In: Proceedings of International Workshop on Challenges in Web Information Retrieval and Integration (WIRI 2005), pp. 47–52 (2005)

Semantics of a Graphical Model for Representing Hypotheses and a System Supporting the Management of Hypotheses

Madori Ikeda[1], Masaaki Nishino[2], Koichiro Doi[3],
Akihiro Yamamoto[1], and Susumu Hayashi[4]

[1] Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan
m.ikeda@iip.ist.i.kyoto-u.ac.jp, akihiro@i.kyoto-u.ac.jp
[2] NTT Cyber Solutions Laboratories, NTT Corporation, Kanagawa 239-0847, Japan
[3] Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8561, Japan
[4] Graduate School of Letters, Kyoto University, Kyoto 606-8501, Japan

Abstract. Constructing hypotheses is a principal work for academic researchers to constitute knowledge. Hypotheses often contain uncertainty, might conflict with each others, and might be revised by discovery of new witnesses. Therefore researchers are heavily loaded with the management of hypotheses. In order to support such activities we propose a graphical model for representing hypotheses, under the recent situation that materials for academic researches are distributed in digital manners, e.g. digital libraries and WWW. We also propose how to construct and manage hypothesis graphs based on our model. It is well-known that graphs are suitable to represent knowledge and many works treating knowledge as graphs have been presented. Previous researches proposed methods how to construct and manage hypotheses automatically, while our present work is aimed at supporting researchers so that they can construct and manage graphs based on semantics with uncertainty. In this paper, we define the syntax and the semantics of hypothesis graphs in our model and present methods to support researchers in the construction and management of hypothesis graphs with a system implemented the methods.

Keywords: Graphical Representation of Hypotheses, Construction of Hypotheses, Supporting System, Knowledge Representation.

1 Introduction

Constructing hypotheses is one of the central works for academic researchers, in particular, researchers in Humanities. Constructed hypotheses constitute knowledge in academic researches. Materials of academic researches are documents and objects, and hypotheses have been represented in the form of documents in natural languages. Recently materials of academic researches are distributed in digital manners, for example, in digital libraries and on WWW. The final aim of our work is to provide a new form for representing hypotheses which is suitable to use materials represented in digital manners and is also suitable to be exchange in some digital materials, e.g. homepages in WWW.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 32–43, 2011.

In this paper we propose a model of representing hypotheses in graphs (in the sense of Computer Science), and call it the *Reasoning Web Model* (the *RW model* for short). We employ the form of graphs because it is known that graphs are suitable to represent knowledge in Computer Science. For example, knowledge represented in the Semantic Web model can be regarded as graphs and are useful for representing and exchanging knowledge in WWW. The RW model is similar to Semantic Web but is for representing reasoning while Semantic Web is for representing meaning of documents.

We also propose how to support researchers in constructing hypotheses and managing them with demonstrating a system called *Hypothesis Management System*. In the Artificial Intelligence area many researches have been proposed for treating knowledge represented in graphs, e.g. Truth Maintenance System[1] and Bayesian Networks[4]. Compared with them, the RW model is proposed with aiming at revising hypotheses based on both semantics and parameters representing confidence. In this paper, we give the syntax and the semantics of graphs based on the RW model, and treat how to support constructing and managing the graphs, but do not treat how to automatically construct the graphs from data like in Bayesian Networks.

This paper is organized as follows: Section 2 gives a running example of a hypothesis and the graph representing it. The system to support constructing and managing the graphs is also shown. Section 3 gives the formal syntax of hypothesis graphs, and their semantics. Section 4 shows how to support constructing and managing the graphs by using some previous works in the Machine Learning area. We explain our conclusion and future works in Section 5.

2 Hypothesis Graphs and Hypothesis Management System

In the RW model every hypothesis is represented as a graph, and called a *hypothesis graph*. We start with giving a simple example of a hypothesis and its graphical representation.

Example 1. Let us consider a case that a researcher has obtained digital photos of a handwritten notebook written in around the end of the 19th century. He knows that it was written by a famous mathematician, and would like to decide when it was written. He constructed the following hypothesis:

> I found letters in the notebook and guessed it denotes "ABC" which is the title of a book. By searching in a digital library, I know a book titled "ABC" was published in 1870. So, I conjecture that the notebook was written after 1870. In addition, I know the author of the notebook died in 1890. I conclude the notebook was written in years from 1870 to 1890.

This hypothesis is represented as a hypothesis graph in Fig. 1. In the figure, every rectangle represents a *node*, every round rectangle represents a set of nodes which we call a *form*, every arrow represents a *link* and every text drawn in a node represents a *label*. Note that the picture named "A digital library" is just for

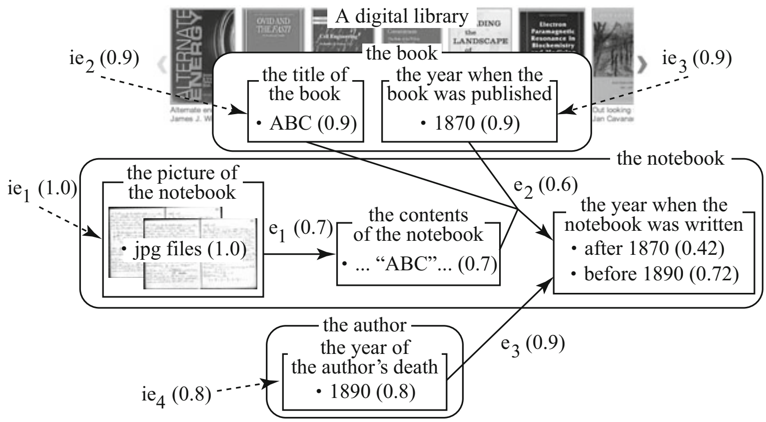


Fig. 1. A hypothesis graph representing a hypothesis in Example 1

representing how he obtained the information about the book "ABC" and is not an element of hypothesis graphs.

In hypothesis graphs, each form represents some object and each node in the form represents some attribute of the object. The form and the node are assumed to be named for the object and the attribute respectively. In this paper, we do not distinguish a form or a node and its name. Each label is a text in some natural language and the set of labels attached to a node represents a value of the attribute which the node represents. Each link represents reasoning which derives some property of the attribute of the *target node* based on *grounds* as the values of *source nodes*. In addition, each real number in parentheses represents a degree of *confidence* or *certainty* of a label or a link. In Fig. 1, the node named "the contents of the notebook" represents that the notebook object contains letters "ABC" and the link e_1 represents a reasoning which derives the contents based on the picture of the notebook object.

Our *Hypothesis Management System* which illustrated in Fig. 2 implements the method of managing hypotheses based on the semantics of hypothesis graphs. Users of this system can construct and manage their own hypothesis graphs through interactive operation on the system.

3 Syntax and Semantics of Hypothesis Graphs

3.1 Syntax

We define the syntax of hypothesis graphs in the RW model[1] as follows.

Definition 1. Let V be a set of *nodes*. A *link* is a pair $e = (U, v)$ where $U \subseteq V$ and $v \in V$. The set U is denoted by $\mathrm{src}(e)$ and the node v is denoted by $\mathrm{tgt}(e)$. Every element of $\mathrm{src}(e)$ is called a *source node* of e and $\mathrm{tgt}(e)$ is called the *target node* of e. For a set E of links, a sequence $v_0, v_1, ..., v_n$ of nodes is a *path* if

[1] The definition here is a modification of our previous work[3].

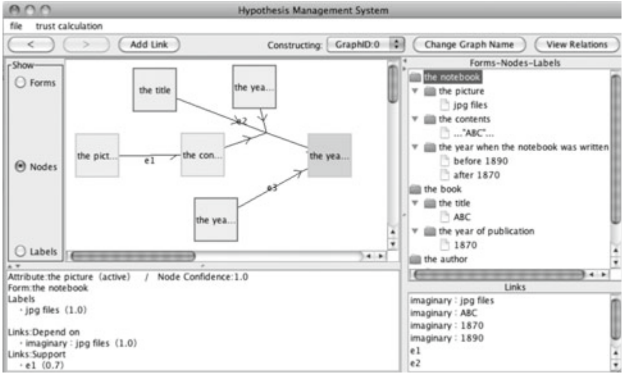


Fig. 2. A view of Hypothesis Management System

for every $i (1 \leq i \leq n)$ there exists a link $(U, v_i) \in E$ such that $v_{i-1} \in U$. If $v_0 = v_n, n \neq 0$, the path is called a *cycle*. A *hypothesis graph* is a pair $\langle V, E \rangle$ where E is a set of links which has no cycle.

Definition 2. For a hypothesis graph $\langle V, E \rangle$, let $\mathcal{F}$ be a family of subsets of V satisfying that $\bigcup_{f \in \mathcal{F}} f = V$ and that $f \cap f' = \emptyset$ for $f, f' \in \mathcal{F}, f \neq f'$. Each element of $\mathcal{F}$ is called a *form*.

Definition 3. Let Λ be a finite set of labels. We assume that $V \cap \Lambda = \emptyset$. Each node $v \in V$ of a hypothesis graph $\langle V, E \rangle$ is given some labels and the set of labels of v is denoted by $L(v) \in 2^{\Lambda}$. We assume that $L(v) \cap L(v') = \emptyset$ if $v \neq v'$.

Definition 4. In a hypothesis graph $\langle V, E \rangle$, every label $l \in \bigcup_{v \in V} L(v)$ and every link $e \in E$ are given a real number in the range $(0, 1]$ and we call the number *confidence*. The confidence of a label l is denoted by $\text{conf}(l)$ and the confidence of a link e is denoted by $\text{conf}(e)$.

3.2 Semantics of Nodes

In a hypothesis graph $\langle V, E \rangle$, for every node $v \in V$ we assume that there exists the *domain* $D(v)$ which any value of the attribute of v must be contained. We also assume that for every node $v \in V$, if a subset $X \subseteq L(v)$ satisfies some property, then any subset $Y \subseteq X$ also satisfies the same property. For every label $l \in L(v)$, we assign a subset $I(l)$ of $D(v)$.

Definition 5. Let $\text{val}(v)$ be the subset of $D(v)$ defined as $\text{val}(v) = \bigcap_{l \in L(v)} I(l)$. We call $\text{val}(v)$ the *value* of the attribute which v represents.

This definition means that the value of the attribute of v is decided with $L(v)$ and that each $l \in L(v)$ represents a certain kind of property of the attribute. Entire property of an object is represented by labels of each node in the form representing the object.

Confidence of the property represented by a label l is represented by $\text{conf}(l)$ and we intend that the higher $\text{conf}(l)$ means the more certain on the property.

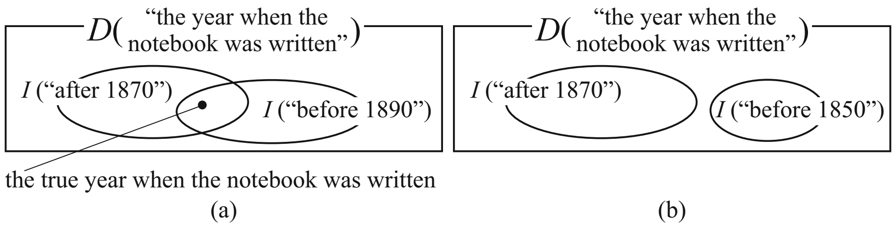


Fig. 3. Interpretation of a value of node

For every node $v \in V$, we assume that there exists a *fact* in $\mathrm{val}(v)$. Based on this assumption, we interpret a condition $\mathrm{val}(v) = \emptyset$ as an occurrence of contradiction on the node v.

Example 2. In the hypothesis graph shown in Fig. 1, there is a notebook object and the node named "the year when the notebook was written" represents the year when the notebook object was written. Since L("the year when the notebook was written") = {"after 1870","before 1890"}, the value of the node can be interpreted as years from 1870 to 1890 as shown in Fig. 3(a) and is not empty. Therefore we can conjecture that the fact, the true year when the notebook was written, is in the value. Moreover, from the confidences of the labels, the property represented by "before 1890" is more certain than the property represented by "after 1870". However, if we assume that a label "before 1850" takes the place of the label "before 1890", the value is empty because it should be interpreted as years from 1870 to 1850 as shown in Fig. 3(b). We cannot obtain the true year assumed to be in the value, and this is contradiction.

Based on the semantics of nodes, we define contradictions of nodes as sets of labels. For the definition, we introduce an *upper family* and a *lower family.* A family $\mathcal{S}$ of subsets of S is an upper family on S if $X \subseteq Y$ and $X \in \mathcal{S} \Longrightarrow Y \in \mathcal{S}$. The family $\mathcal{S}$ is a lower family on S if $X \supseteq Y$ and $X \in \mathcal{S} \Longrightarrow Y \in \mathcal{S}$.

Definition 6. In a hypothesis graph $\langle V, E\rangle$, each node $v \in V$ is given an upper family $\mathcal{N}(v)$ on $L(v)$ and we define $\mathcal{P}(v)$ as $2^{L(v)} \setminus \mathcal{N}(v)$. Note that $\mathcal{P}(v)$ is a lower family on $L(v)$. We call each element of $\mathcal{N}(v)$ a *contradiction* of v and a hypothesis graph $\langle V, E\rangle$ is *valid* iff $\mathcal{N}(v) = \emptyset$ for every node $v \in V$.

Example 3. In the assumption given in Example 2 that replacing "before 1890", a set { "after 1870", "before 1850" } is a contradiction of "the year when the notebook was written". This contradiction continues even if this node has more labels. So a set of contradictions is an upper family.

3.3 Semantics of Links

We define relation between labels and links. In a hypothesis graph $\langle V, E\rangle$, we assume a mapping $cnsq : \{\, e \in E \mid \mathrm{tgt}(e) = v \,\} \to L(v)$ for each link $e \in E$ and call $cnsq(e)$ the *consequence* of e.

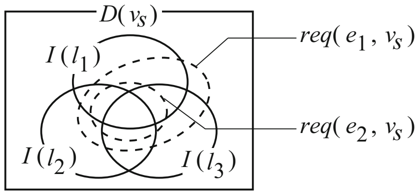


Fig. 4. Interpretation for validity of links

Definition 7. For each label $l \in \bigcup_{v \in V} L(v)$, a set $\mathrm{Sup}(l)$ is defined as follows:

$$\mathrm{Sup}(l) = \begin{cases} \{\, e \in E \mid cnsq(e) = l \,\} & \text{if } \{\, e \in E \mid cnsq(e) = l \,\} \neq \emptyset, \\ \{\, ie \,\} & \text{otherwise,} \end{cases}$$

where ie is an *imaginary link* not in E and $cnsq(ie) = l$. For any imaginary link ie, we define that $\mathrm{src}(ie) = \emptyset$, $\mathrm{tgt}(ie) \in V$, $cnsq(ie) \in L(\mathrm{tgt}(ie))$ and $\mathrm{conf}(ie) \in (0, 1]$. E^+ denotes the union of E and the set of imaginary links.

This definition means that every label l is derived as a consequence of a reasoning $e \in \mathrm{Sup}(l)$. Imaginary links represented as arrows with broken lines in Fig. 1 are defined to make this semantics well-defined for all labels. In the following of this paper, a hypothesis graph means $\langle V, E^+ \rangle$.

The confidence of a reasoning which e represents is represented by $\mathrm{conf}(e)$ and we intend that the higher $\mathrm{conf}(e)$ means the more certain on the reasoning.

Example 4. In the hypothesis graph shown in Fig. 1, e_3 represents the reasoning that the notebook must have been written before the author's death. This reasoning is represented as follows: $\mathrm{src}(\mathrm{e}_3) = \{$"the year of the author's death"$\}$, $\mathrm{tgt}(\mathrm{e}_3) =$ "the year when the notebook was written" and $cnsq(\mathrm{e}_3) =$"before 1890". Moreover, the confidence 0.9 of e_3 represents that the researcher who made the reasoning represented by e_3 thinks the reasoning is almost certain.

Now, we give an assumption relating to e_3 that the author of the notebook died in 1850. Under this assumption, the labels "1890" and "before 1890" are replaced by "1850" and "before 1850" respectively in the hypothesis graph in Fig. 1. The assumption is called *contradicting assumption*, and intends the contradiction on the node "the year when the notebook was written" as shown in Fig. 3(b).

We define the validity of a link as the relation with each value of source nodes.

Definition 8. Every link $e \in E$ is assumed to have a subset $req(e, v) \subseteq D(v)$ for each $v \in \mathrm{src}(e)$ and e is *valid* iff $\mathrm{val}(v) \subseteq req(e, v)$ for every $v \in \mathrm{src}(e)$.

Example 5. Let e_1 and e_2 be links and a node v_s having labels l_1, l_2 and l_3 be a source node of both. Suppose that a situation of subsets on $D(v_s)$ is shown in Fig. 4. This situation means that removing l_1 or l_3 makes e_1 invalid and that removing any label makes e_2 invalid.

Similar to the definition of contradictions, we define grounds of links as sets of labels attached to source nodes based on the definition of the validity.

Definition 9. In a hypothesis graph $\langle V, E^+ \rangle$, every $e \in E$ is given upper families $\mathcal{G}(e, v)$ on $L(v)$ for each source node $v \in \mathrm{src}(e)$. We call an element of $\mathcal{G}(e, v)$ a *ground* of e on v.

Using this definition, we define a set of labels which is essential to the validity of a link. For the definition, we introduce some notations. For a family $\mathcal{S}$ of sets, the set of minimal sets in $\mathcal{S}$ is denoted by $\min\mathcal{S}$ and the set of maximal sets in $\mathcal{S}$ is denoted by $\max\mathcal{S}$. We also introduce a definition from Satoh et al.[5].

Definition 10 ([5]). For a family $\mathcal{S}$ of subsets of a finite set, a *hitting set* H of $\mathcal{S}$ is a set such that for each $X \in \mathcal{S}, X \cap H \neq \emptyset$. A *hitting set* M of $\mathcal{S}$ is minimal if there exists no other hitting set H of $\mathcal{S}$ such that $H \subset M$ and the family of minimal hitting sets of $\mathcal{S}$ is denoted by $\mathcal{MHS}(\mathcal{S})$.

Now we define a set which is essential to the validity of a link.

Definition 11. For every $e \in E$, a subset $G \subseteq L(v)$ is an *essential ground* if $G \in \mathcal{MHS}(\min\mathcal{G}(e, v))$ where $v \in \mathrm{src}(e)$.

Example 6. In Fig. 4, both $\{ l_1, l_3 \}$ and $\{ l_1, l_2, l_3 \}$ are grounds of the link e_1, and $\mathcal{MHS}(\min\mathcal{G}(e_1, v_s)) = \{ \{ l_1 \}, \{ l_3 \} \}$. Let us go back to Example 5. Removing one essential ground of a link makes it invalid.

3.4 Causes and Contradictors

Based on the definitions above, we define a *cause of deriving a consequence* as a set of links.

Definition 12. For every label $l \in \bigcup_{v \in V} L(v)$ in a hypothesis graph $\langle V, E^+ \rangle$, a family $\mathcal{C}(l)$ of subsets of E^+ is recursively defined as follows:

1. $\mathrm{Sup}(l)$ is an element of $\mathcal{C}(l)$.
2. For every $e \in \mathrm{Sup}(l)$ such that $e \in E$, a set $\mathrm{Sup}(l) \cup \bigcup_{l' \in G} \mathcal{C}(l')$ is also an element of $\mathcal{C}(l)$ where G is an essential ground of e.

We call an element of $\mathcal{C}(l)$ a *cause* of l.

This definition means that each cause of a label is a set of links which is a set of reasonings needed to derive a consequence represented by the label.

Example 7. In Fig. 1, causes of the label “after 1870” are $\{\mathrm{e}_2\}$, $\{\mathrm{e}_1, \mathrm{e}_2\}$, $\{\mathrm{ie}_2, \mathrm{e}_2\}$, $\{\mathrm{ie}_3, \mathrm{e}_2\}$ and $\{\mathrm{ie}_1, \mathrm{e}_1, \mathrm{e}_2\}$.

In our definitions, a contradiction is a set of labels and causes of labels are sets of links. So we define a *cause of a contradiction* as a set of links and we call the cause a minimal contradictor.

Definition 13. Let a set $\{ l_1, l_2, ..., l_n \}$ be an element of $\mathcal{MHS}(\min\mathcal{N}(v))$ and M be an element of $\mathcal{MHS}(\{ \mathcal{C}(l_1), \mathcal{C}(l_2), ..., \mathcal{C}(l_n) \})$. A *minimal contradictor* of v is a subset of E^+ such that $\bigcup_{X \in M} X$. The set of minimal contradictors of v is denoted by $\mathcal{MC}(v)$.

Example 8. In Fig. 1 with the contradicting assumption following Example 4, minimal contradictors of the node "the year when the notebook was written" are $\{ \mathrm{e}_2 \}$, $\{ \mathrm{e}_1, \mathrm{e}_2 \}$, $\{ \mathrm{ie}_2, \mathrm{e}_2 \}$, $\{ \mathrm{ie}_3, \mathrm{e}_2 \}$, $\{ \mathrm{ie}_1, \mathrm{e}_1, \mathrm{e}_2 \}$, $\{ \mathrm{e}_3 \}$ and $\{ \mathrm{ie}_4, \mathrm{e}_3 \}$.

We extend this definition for nodes as follows.

Definition 14. Let V be a set of nodes $\{ v_1, v_2, ..., v_n \}$ and M be an element of $\mathcal{MHS}(\{ \mathcal{MC}(v_1), \mathcal{MC}(v_2), ..., \mathcal{MC}(v_n) \})$. A *minimal contradictor* of V is a subset of E^+ such that $\bigcup_{X \in M} X$. The set of minimal contradictors of V is denoted by $\mathcal{MC}(V)$.

4 Management of Hypothesis Graphs

4.1 Construction of Hypothesis Graphs

We consider a situation that a *user* constructs a hypothesis graph on *Hypothesis Management System* (*HMS* for short).

Any hypothesis graph $\langle V, E^+ \rangle$ starts with $V = \emptyset$ and $E^+ = \emptyset$. For constructing a hypothesis graph, the user is allowed to do the following actions:

Adding a new link: The user can add a new link $e \notin E^+$ to $\langle V, E^+ \rangle$. Before adding e, if $\mathrm{tgt}(e) \notin V$, the user adds $\mathrm{tgt}(e)$ to V. If $cnsq(e) \notin L(\mathrm{tgt}(e))$, the user adds $cnsq(e)$ to $L(\mathrm{tgt}(e))$. After adding e, the user updates $\mathrm{Sup}(cnsq(e))$ by adding e to $\mathrm{Sup}(cnsq(e))$ for HMS and if e is not imaginary, inputs $\min\mathcal{G}(e, v)$ where $v \in \mathrm{src}(e)$ to the system.

Removing an existing link: The user can remove any link $e \in E^+$ from $\langle V, E^+ \rangle$.

In addition, we assume that for any $v \in V$, the user can judge whether a subset $X \subseteq L(v)$ is an element of $\mathcal{N}(v)$ or not.

Then, HMS responds to the actions by executing the following processes:

When a link is added, HMS checks whether or not any contradiction occurs on any node, and if any contradiction occurs, the system revises the hypothesis graph by removing all links of a minimal contradictor to remove the contradiction.

When a link is removed, If $\mathrm{Sup}(cnsq(e)) = \{ e \}$ for the removed e, HMS removes $cnsq(e)$ from $L(\mathrm{tgt}(e))$ and removes every link e' such that $\{ cnsq(e) \}$ is an essential ground of e'. Note that this process is executed when e was removed by either the user or HMS, so this is recursive.

4.2 Enumerating Minimal Contradictions

We formulate the method how to check whether or not any contradiction occurs as enumerating minimal contradictions. Suppose that, for a hypothesis graph $\langle V, E^+ \rangle$, HMS is given $L(v)$ of every node $v \in V$ and can ask queries whether or not any subset $X \subseteq L(v)$ is an element of $\mathcal{N}(v)$ to the user. Enumerating minimal contradictions is enumerating all elements of $\min\mathcal{N}(v)$ for given $v \in V$ and $\min\mathcal{N}(v) \neq \emptyset$ means occurrences of contradictions on v.

In order to enumerate them, we adapt an algorithm developed in [5]. Let Π be a set and $\mathcal{S}^+$ be a lower family on Π. We define $\mathcal{S}^- = 2^\Pi \setminus \mathcal{S}^+$, then $\mathcal{S}^-$ is an upper family on Π and assume the *oracle* (in the sense in Machine Learning, and the user of HMS here) knows $\mathcal{S}^+$. Then, the algorithm is given Π and enumerates elements of $\max\mathcal{S}^+$ by asking queries whether or not any subset $X \subseteq \Pi$ is an element of $\mathcal{S}^+$ to the oracle.

Let $\overline{X}$ be the complement of a subset X of S. For a family $\mathcal{S}$ of subsets of S, $\overline{\mathcal{S}}$ denotes $\{\, \overline{X} \mid X \in \mathcal{S} \,\}$. Since $\mathcal{N}(v)$ is an upper family on $L(v)$, $\overline{\mathcal{N}(v)}$ is a lower family on $L(v)$ and $\min\mathcal{N}(v)$ equals to $\overline{\max\overline{\mathcal{N}(v)}}$. If the algorithm is given $L(v)$ and the user knows $\mathcal{N}(v)$, the algorithm can enumerate all elements of $\max\overline{\mathcal{N}(v)}$ by asking queries whether or not any subset $X \subseteq L(v)$ is an element of $\overline{\mathcal{N}(v)}$ to the user. Then, HMS can obtain elements of $\min\mathcal{N}(v)$ by calculating $\overline{\max\overline{\mathcal{N}(v)}}$.

4.3 Enumerating Minimal Contradictors

We formulate the revision of a hypothesis graph as enumerating minimal contradictors. In a hypothesis graph, a cause of contradictions is a subset of links defined as a minimal contradictor. Therefore, HMS can remove the contradictions by removing all links in the minimal contradictor and their consequences.

Suppose that for a hypothesis graph $\langle V, E^+ \rangle$, HMS is given $L(v)$ and $\min\mathcal{N}(v)$ of every $v \in V$, $\min\mathcal{G}(e, v)$ of every $e \in E$ and $v \in \mathrm{src}(e)$, $\mathrm{Sup}(l)$ of every $l \in \bigcup_{v \in V} L(v)$. Enumerating minimal contradictors is enumerating all elements of $\mathcal{MC}(V)$. In order to enumerate them, we adapt an algorithm developed by Uno[6] to HMS which enumerates elements of $\mathcal{MHS}(\mathcal{S})$ for a given family $\mathcal{S}$.

Computing $\mathcal{C}(l)$ is achieved by a process based on the definition of $\mathcal{C}(l)$. An element of $\mathcal{C}(l)$ is already given as $\mathrm{Sup}(l)$. Other elements of $\mathcal{C}(l)$ can be obtained in a recursive manner based on the definition. Because every hypothesis graph is acyclic, this recursive process stops.

In order to calculate $\mathcal{MC}(v)$, HMS needs elements of $\mathcal{MHS}(\min\mathcal{N}(v))$ and elements of $\mathcal{MHS}(\{\, \mathcal{C}(l) \mid l \in M \,\})$, where M is an element of $\mathcal{MHS}(\min\mathcal{N}(v))$. Given $\min\mathcal{N}(v)$, HMS can enumerate minimal hitting sets by Uno's algorithm. Then HMS can obtain all elements of $\mathcal{MC}(v)$ by calculating $\bigcup_{X \in M'} X$ where M' is an element of $\mathcal{MHS}(\{\, \mathcal{C}(l) \mid l \in M \,\})$. Similarly, HMS can enumerate minimal contradictors of V by calculations.

4.4 Management of Confidences

Certainty of a hypothesis should be related to the structure of its graphical representation. So confidences in a hypothesis graph should be managed when the graph is changed.

On HMS, every confidence of a link is given by the user when the link is added but every confidence of a label is automatically given by calculation. A confidence of a label l derived by a link e is calculated as $\mathrm{conf}(l) = \mathrm{conf}(e) \times g(\mathrm{conf}(l_1), ..., \mathrm{conf}(l_n))$ where l_i $(1 \leq i \leq n)$ is an element of $\bigcup_{v \in \mathrm{src}(e)} \min\mathcal{G}(e, v)$ and g is a function which combines confidences of the labels. In this paper,

Table 1. The relation of hypothesis graphs

	the notebook	the book	the author
$\langle V, \{ ie_1, ie_2, ie_3, ie_4, e_1, e_2 \} \rangle$	notebook 1	book 1	author 1
$\langle V, \{ ie_1, ie_2, ie_3, e_1, e_2 \} \rangle$	notebook 1	book 1	author 2
$\langle V, \{ ie_1, ie_2, ie_3, ie_4, e_1, e_3 \} \rangle$	notebook 2	book 1	author 1
$\langle V, \{ ie_1, ie_2, ie_4, e_1, e_3 \} \rangle$	notebook 2	book 2	author 1
$\langle V, \{ ie_1, ie_3, ie_4, e_1, e_3 \} \rangle$	notebook 2	book 3	author 1
$\langle V, \{ ie_1, ie_2, ie_3, ie_4, e_3 \} \rangle$	notebook 3	book 1	author 1
$\langle V, \{ ie_2, ie_3, ie_4, e_3 \} \rangle$	notebook 4	book 1	author 1

Table 2. The relation of the notebook

	the picture of the notebook	the contents of the notebook	the year when the notebook was written
notebook 1	jpg files (1.0)	...“ABC”... (0.7)	after 1870 (0.42)
notebook 2	jpg files (1.0)	...“ABC”... (0.7)	before 1850 (0.72)
notebook 3	jpg files (1.0)	null	before 1850 (0.72)
notebook 4	null	null	before 1850 (0.72)

Table 3. The relation of the book

	the title of the book	the year when the book was published
book 1	ABC (0.9)	1870 (0.9)
book 2	ABC (0.9)	null
book 3	null	1870 (0.9)

Table 4. The relation of the author

	the year of the author's death
author 1	1850 (0.8)
author 2	null

we do not describe calculation manner of g in detail[2]. Roughly speaking, the calculation is like calculation of joint probability (but precisely not). Anyway, the calculation of confidences is executed one by one along in directions of links and a consequence is more certain if a link deriving it is more certain or is based on more certain labels.

4.5 Managing Hypothesis Graphs in Parallel

One of the minimal contradictors of V is sufficient to revise contradictions of a hypothesis graph $\langle V, E^+ \rangle$. So some different valid hypothesis graphs can emerge from the revision. HMS maintains all of these valid graphs. Note that differences among hypothesis graphs which HMS maintains are only sets of links. In other words, the differences of these hypothesis graphs can be represented as differences of objects which forms represent.

In order to maintain hypothesis graphs, we use a relational database. We prepare two types of relations, a relation of hypothesis graphs and several relations of forms. In the relation of hypothesis graphs, rows are for hypothesis graphs and columns are for objects and every element in the relation is the key for the tuple

[2] We describe the detailed calculation manners of confidences in [2].

global: hypothesis graph $\langle V, E_1^+\rangle, ..., \langle V, E_n^+\rangle$; set $check_1, ..., check_n$;
add_link$(e, \langle V, E_i^+\rangle)$
  for each $\langle V, E_j^+\rangle (1 \leq j \leq n)$
    if $\text{ver}(f, E_i^+) = \text{ver}(f, E_j^+)$ for all $f \ni v$ where $v \in \text{src}(e)$ **then**
      add e to $\langle V, E_j^+\rangle$;
      add $\text{tgt}(e)$ to $check_j$;
  return;
end

Fig. 5. An algorithm adding a link to hypothesis graphs

in a relation of a form. In a relation of a form, rows are for objects, columns are for nodes and every element is a set of labels of a node in the form.

Example 9. The invalid hypothesis graph in Fig. 1 with the contradicting assumption has seven minimal contradictors, $\{\, \text{e}_2 \,\}$, $\{\, \text{e}_1, \text{e}_2 \,\}$, $\{\, \text{ie}_2, \text{e}_2 \,\}$, $\{\, \text{ie}_3, \text{e}_2 \,\}$, $\{\, \text{ie}_1, \text{e}_1, \text{e}_2 \,\}$, $\{\, \text{e}_3 \,\}$ and $\{\, \text{ie}_4, \text{e}_3 \,\}$. So there are seven valid hypothesis graphs after revision by removing each of them. We show the relational database for these valid hypothesis graphs in Table 1-4.

Maintaining some graphs on HMS, the user assumed to add or to remove only one link to one hypothesis graph at a time, and might have to do the same action to different hypothesis graphs. HMS supports the user in constructing and managing hypothesis graphs in parallel.

Definition 15. In a hypothesis graph $\langle V, E^+\rangle$, a subset $X \subseteq E^+$ is a *version* of a form f, if $X = \{\, e \in E^+ \mid \text{tgt}(e) \in f \,\}$. A version of f in $\langle V, E^+\rangle$ is denoted by $\text{ver}(f, E^+)$.

Definition 16. A form f is *common* to hypothesis graphs $\langle V, E_i^+\rangle$ and $\langle V, E_j^+\rangle$ if $\text{ver}(f, E_i^+) = \text{ver}(f, E_j^+)$.

Because the property of an object is based on labels derived by links, if a form f is common to $\langle V, E_i^+\rangle$ and $\langle V, E_j^+\rangle$, then the object represented by f in $\langle V, E_i^+\rangle$ and f in $\langle V, E_j^+\rangle$ are same. Moreover, if a new link will be valid in $\langle V, E_i^+\rangle$ and for every source node v, $\text{ver}(f, E_i^+) = \text{ver}(f, E_j^+)$ where $f \ni v$, then the link will be valid also in $\langle V, E_j^+\rangle$. Based on this observation, HMS automatically adds a link to some hypothesis graphs if the link is valid in the graphs. In removing a link from a hypothesis graph, HMS does not anything to other graphs based on our policy of removing no links as long as any contradiction occurs.

We show an algorithm making the user add a link to some hypothesis graphs at a time in Fig. 5. This algorithm is given a new link e and a hypothesis graph $\langle V, E_i^+\rangle$ which the user is constructing and adds the link to some hypothesis graphs. In the algorithm, $check_j$ is a subset of V which is related to hypothesis graph $\langle V, E_j^+\rangle$ and $check_j$ memorizes nodes which need to be checked whether or not any contradiction occurs. When the user changes a hypothesis graph under construction from $\langle V, E_i^+\rangle$ to $\langle V, E_j^+\rangle$, at first HMS enumerates minimal contradictions of every $v \in check_j$. If any contradiction occurs, HMS enumerates minimal contradictors and revises $\langle V, E_j^+\rangle$. Note that if revision is done for $\langle V, E_j^+\rangle$, HMS newly maintains revised graphs instead of $\langle V, E_j^+\rangle$.

To support the user in constructing a valid hypothesis, HMS shows the relational database to the user. With the relational database, the user can easily compare the hypothesis graphs about values of nodes or confidences of labels.

5 Conclusion and Future Works

We have introduced a graphical model by defining hypothesis graphs and their semantics. We have also proposed how to construct and manage hypothesis graphs. With hypothesis graphs, researchers can represent relations between grounds and consequences of reasoning. The semantics of hypothesis graphs enables us to provide a method for revising hypothesis graphs by removing contradiction. We have implemented HMS to support constructing and managing hypothesis graphs. Users of HMS can obtain valid hypothesis graphs if the users correctly recognize contradictions occurring on their own hypothesis graphs and input the contradictions to the system.

In academic researches, researchers often collaborate with each other for constructing a hypothesis. In such cases, researchers not only agree but also ague with each other about the hypothesis. In our future work, we should improve our system so that it may support the argument between researchers. For this improvement we should reveal points of the argument based on analysis of difference between hypothesis graphs conflicting with each other. Another subject would be constructing a framework for the argument. It would be also a subject correcting differences of tendency concerning confidences among researchers in the case a researcher observes a hypothesis graph constructed by another researcher. The differences of the tendency come from that the degree of confidences depends on subjectivity of each researcher. To this subject, it may be a solution normalizing confidences based on statistical analysis of the tendency.

Acknowledgment. This work was partly supported by Kyoto University Global COE "Information Education and Research Center for Knowledge-Circulating Society", and Grant-in-Aid from JSPS.

References

1. Doyle, J.: A Truth Maintenance System. Artificial Intelligence 12(3), 251–272 (1979)
2. Ikeda, M., Doi, K., Yamamoto, A., Hayashi, S.: Reformalization of Reasoning Web Model for Graphical Representation of Hypotheses. In: SIG-FPAI-A804-13, JSAI, pp. 69–76 (2009) (in Japanese)
3. Nishino, M., Yamamoto, A. and Hayashi, S.: Graphical Representation of Valid Conjectures and a System for Supperting Construction of Hypotheses. Technical Report of IEICE 107(523), AI2007-44; IEICE, pp. 1–6 (2008) (in Japanese)
4. Pearl, J.: Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning. Proceedings, Cognitive Science Society (UCLA Computer Science Department Technical Report 850021 (R-43)), UC Irvine, CA, pp. 329–334, (1985)
5. Satoh, K., Uno, T.: Enumerating Maximal Frequent Sets using Irredundant Dualization. In: Grieser, G., Tanaka, Y., Yamamoto, A. (eds.) DS 2003. LNCS (LNAI), vol. 2843, pp. 256–268. Springer, Heidelberg (2003)
6. Uno, T.: A Practical Fast Algorithm for Enumerating Minimal Set Coverings. In: SIGAL 83, IPSJ, pp. 9–16 (2002) (in Japanese)

An Approach for Improving Thai Text Entry on Touch Screen Mobile Phones Based on Distance and Statistical Language Model

Siwacha Janpinijrut[1], Cholwich Nattee[1], and Prakasith Kayasith[2]

[1] School of ICT, Sirindhorn International Institute of Technology, Thammasat University, Thailand
[2] National Electronics and Computer Technology Center
National Science and Technology Development Agency (NSTDA), Thailand Science Park
siwacha.janpinijrut@student.siit.tu.ac.th
cholwich@siit.tu.ac.th
prakasith.kayasith@nectec.or.th

Abstract. This paper introduces an approach for improving Thai text inputting via virtual keyboard on touch screen mobile phones. We propose a technique for character candidate selection to choose the most proper character sequence. The proposed approach consists of two main parts i.e. candidate character generation based on touch area, and trigram model for candidate selection. The candidate generation is used to acquire more than one character per one touch. We define the area nearby the touch point as a touch area. The area is defined based on the statistics collected from various users. The characters within the touch area are considered as a set of candidate characters. The candidates are weighted based on their distances from the touch point. Then, the character trigram model is applied to select the combination of characters with the highest probability suggested to user. In order to evaluate the performance of the proposed technique, we evaluate the word-level accuracy of the words generated by the proposed technique and the words generated by combining the characters with the shortest distances from the touch point. Our technique yields better performance. It produces the results with the accuracy of 93.29% in character level and 66.67% in word level, while the accuracy is 83.48% in character level and 48.33% in word level when the nearest characters are selected.

Keywords: virtual keyboard, touch screen mobile phone, word selection, candidate selection, distance-based candidate generation, trigram model.

1 Introduction

Nowadays, a number of mobile phones equipped with touch screen is significantly increased. Most of them are equipped with two widely-used types of touch screens i.e. resistive and capacitive touch screens. In the former type, its panel composes

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 44–55, 2011.

of two metallic, electrically conductive layers separated by a narrow gap. When an object presses down, two layers become connected and shot circuit at that point. The controller then calculates resistance so as to know which point is touched. This type allows any kind of object to touch, but it does not allow multi-touch. While the capacitive touch screen panel consists of insulator coats with transparent conductor. When part of human bodies directly touch on the screen, distortions of electrostatic field occurs and its controller will calculate the touch points. This type allows multi-touch, but it restricts only human bodies touch.

Virtual keyboard is an important input method of the touch phones. Focusing on Thai language, it has more alphabets than English. This causes more problems when we display Thai keyboard on the limited touch screen area. Keys are displayed very close to each other. Even if the standard keyboard layout for Thai language splits all 88 letters into two layers, it still has a lot of buttons per layer. Each button becomes a very small area. Especially on capacitive touch screen type, mostly human finger always larger than the button. This makes it hard to accurately touch on the intended position. This problem causes a lot of errors, and decreases efficiency of Thai text entry on mobile phones.

This paper focuses on reducing the typing error. With very close distance between each character position on small keyboard area, users may not be able to touch the exact area of the character that they want to type. However, based on an assumption that user tries to touch on the exact area. Even they miss the character; the wrong position should be close to the right area. This characteristic is applied in our approach that the characters nearby the touch point are selected as the candidate characters. This should provide a high possibility that the right character is one of the candidates. Another assumption in this paper is that most of words that users type are common words widely used in various documents. So, we grade the candidates with a statistical language model training from a Thai corpus in order to find out the best candidate with the highest probability and suggest to user.

We scope only the performance in term of correctness of our algorithm by ignoring the processing time between each keystroke. So, we use mobile phones only for collecting input data. Then, input data are fetched to process on computer.

This paper composes of five sections. Next, we explain the background of this paper and other related works. Distance-based candidate generation and candidate selection are described. Then, Section 5 shows the experiment method and the results. Finally, we conclude the paper and show our future direction in Section 6.

2 Background

Designing new layout for virtual keyboard is one way to improve performance of text entry. For example, Bi et al. [1] and Zhai et al. [9] introduced a new layout for English language while Sarcar et al. [7] proposed new Indian keyboards. The problem of new layout is that it is hard to make most of users become familiar with.

We intend to improve performance of Thai text entry based on the standard keyboard, Thai Kedmanee keyboard layout, the most well-known Thai keyboard layout, rather than create new layout.

There are many researches proposed for improving text input methods on standard layout of each language. Potipiti et al.[6] addressed an approach to reduce the number of keystrokes on physical keyboard. For Thai users that have to type bi-lingual document i.e. English and Thai, they introduced intelligent keyboard with 2 main properties. The first one is a technique for language identification that automatically switches language without pressing language-switch button. The other approach is key prediction without pressing shift key for Thai language. Both functions use character trigram probabilistic model compare probability between English word and Thai word for Language Identification and compare probability between Thai character with shift key and without shift key for Key Prediction. Brewster and Hughes [2] used pressure of finger touch on touch phone to separate between lowercase and uppercase of English language instead of using shift key. Soft press means a lowercase and hard press means an uppercase letter.

Another way to reduce keystroke is to make a short form of words. Shieber and Baker [8] reduced keystrokes of typing in English by abbreviating input word to short form. They abbreviated by dropping all vowels and consecutive duplicate consonants. Kristensson and Zhai [4] used geometric pattern as a model to do word prediction on soft keyboard using stylus. They trained model by drawing lines between letter key center points in lexicon word, and doing the same thing but with input point sequence to create input pattern. Then, match input pattern against legitimate pattern of word to find out what the most match word.

MacKenzie and Zhang [5] introduced eye typing technique in English language. User typed by gazing at on-screen keyboard on computer screen via eye tracking equipment. Fixation algorithm used to determine which button receives eye-over. They considered character button nearby eye-focus position as candidates. Combining with knowledge of previous character and language model, they can predict which button user should type. This technique suitable for low accuracy input method. Trying to touch on button that smaller size than user finger is one kind of low accuracy method. So, this paper applies this technique to typing on touch screen.

3 Distance-Based Candidate Generation

Keyboard buttons on virtual keyboard have smaller size than users finger so it is hard to touch on the right position. But, characters that users want to type should nearby their touch point. Bases on this hypothesis, we consider all characters close to touch point as candidates. With too small area, it may not cover the right character that user wants to type. Whereas, with too large area, it gets a lot of candidate characters that affect the processing performance with insufficient resources in mobile phone devices. Although this paper does not consider processing time as the factor yet. But, in our future work, we plan to implement our algorithm running on device. Suitable touch area should be the area that most people touch on. So, we find out which area those users mostly touch on each character. We use the distribution of touch points to identify the suitable area of each key. We create character sequence by shuffling all of Thai alphabets and some special characters that exist on standard Thai keyboard. It composes of 88 characters. All the 88 characters are split into two keyboard layers. Each layer consists of 44 characters. Both layers have the same layout therefore it has only 44 unique positions. We let 20 users typing that sequence in their style, no specific hand of fingers that user have to use. After that, we get 40 coordinates of each character position.

Fig. 1 shows standard keyboard layout with distribution of 44 distinct character positions.

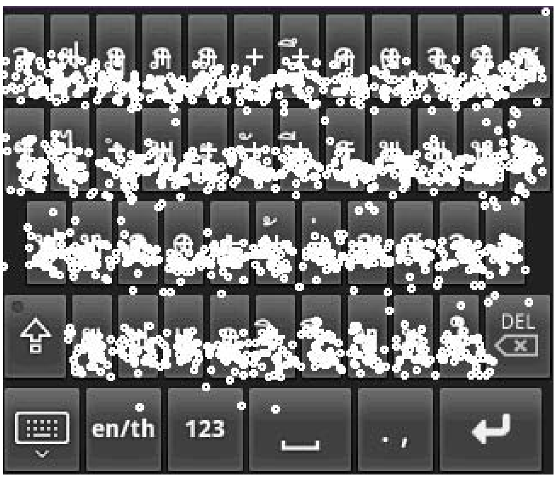


Fig. 1. Distribution of touch point of each character position

3.1 Touch Area

We can find the mean distance and standard deviation for each position.

$$\bar{d_c} = \frac{\sum_{i=1}^{n_c} d_{c_i}}{n_c}, \tag{1}$$

$$SD(d_c) = \frac{\sqrt{\sum_{i=1}^{n_c} (d_{c_i} - \bar{d_c})^2}}{n_c - 1} \tag{2}$$

where
d_{c_i} is distance between touch point and center point position c,
$\bar{d_c}$ is average distance of position c,
n_c is number of touch point of position c.

After we obtain the average distance and SD of each position, we find the average distance and the average SD for all key positions with the following equations.

$$\bar{d} = \frac{\sum_{c=1}^{44} \bar{d_c}}{44}, \tag{3}$$

$$\overline{SD(d)} = \frac{\sum_{c=1}^{44} SD(d_c)}{44} \tag{4}$$

The suitable radius should cover average distance and most of its distribution. Therefore, we set the radius as the following equation.

$$\bar{r} = \bar{d} + 2\overline{SD(d)} \tag{5}$$

However, the radius should not be larger than maximum distance. This equation shows maximum radius it should be.

$$r_{max} = \max_c(\bar{d_c} + 2SD(d_c)) \tag{6}$$

In Section 5, we conduct experiments with various sizes of the radius i.e. 25, 27, 29 ($\bar{r}$), 31, ..., 43 (r_{max}) to find out the most suitable radius..Furthermore, if we consider the distribution of touch points in x and y axes separately, we will find another interesting thing. The distribution on x axis is not biased to the left or right side for most character position. But, most of the distribution on y axis is biased down side as shown in Table 1.

Table 1. Average radius and SD on y axis reference with center point of each row

Keyboard row	$\bar{d_y}$	$SD(d_y)$
1^{st}	15.15	7.71
2^{nd}	10.64	7.29
3^{rd}	7.03	7.12
4^{th}	7.95	7.64

The top row is the most shifted down while the lower rows are slightly shifted. Based on this distribution, the center point of each key button may not be a good reference point. So, we propose a new reference point by averaging the y values of each row. Then, 16, 11, 7 and 8 pixels are the shift down distance from center point of button of the top to bottom rows, respectively. We call it as center of distribution keyboard layout and call the ordinary layout as center of button keyboard layout.

3.2 Candidate Weight

Each candidate character covered by touch area have a different weight depending on its distance from the touch point.

$$w_c = 1 - \frac{d_c}{r} \tag{7}$$

where
w_c is distance weight of candidate character c,
d_c is distance of character c from touch point,
r is radius of touch area.

This weighting scheme is applied based on the concept that user intends to touch the point as close as possible. The character with smaller distance from the touch point, gains higher weight. This value is used to weight the probability of the character obtained by the language model. The character with lower probability may win the higher one if it has higher distance weight.

4 Candidate Selection

N-gram model is widely-used technique in language processing. The more number of n, the more correctness of statistic model is produced and the more system resources are used to process, too. The high number of n cannot apply in this research because of the limited resource on mobile device. Then, trigram which has n equal to three is selected to create character level language model by obtaining the statistics from a Thai corpus named BEST2010 [3] created by NECTEC. This corpus consists of more than 5 million words.

From a sequence of keystrokes, we generate a set of candidate sequences. Each candidate concatenates all possible candidates to create all possible character sequences. The character sequence that has higher probability than any other sequence is presented to user as the most appropriate word. The probability for each character sequence is calculated based on the character trigram model with the distance weight as shown in the following equation.

$$P(c_1 c_2 ... c_n) = \prod_{i=1}^{n} P(c_i | c_{i-2} c_{i-1}) w_i \tag{8}$$

Where
$P(c_1 c_2 ... c_n)$ is probability of each character sequence,
$P(c_i | c_{i-2} c_{i-1})$ is probability of character c_i follow by character $c_{i-2} c_{i-1}$,
w_i is distance weight of character c_i.

The flowchart of proposed methods is shown in Fig. 2.

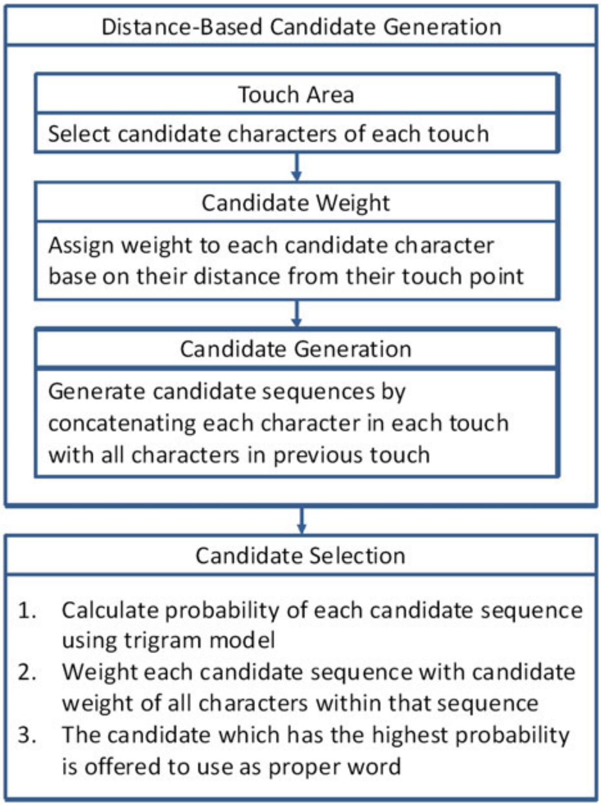


Fig. 2. Flowchart of proposed methods

5 Experiments

There are many touch screen phones available in the market. Android phone becomes a hot new trend and is growing up faster. HTC Magic is Android phone that we use in our work. Its resolution is 320×480 pixels. Its default operating system is Android 1.6. CN Thai keyboard layout is chosen to be experiment keyboard.

Most applications on touch screen phone allow rotatable; change orientation of use, including text input method. But, we focus only on portrait orientation layout that has smaller keyboard area than landscape orientation.

Each character on keyboard is represented by the center point of its button as shown in Fig. 3. The distance between the center points of two horizontally adjacent characters is 27 pixels. While the distance between the center points of two adjacent rows is 55 pixels.

We randomly choose 80 words and 20 words from the corpus to be a training set and a test set respectively. Both sets consist of words with high, medium

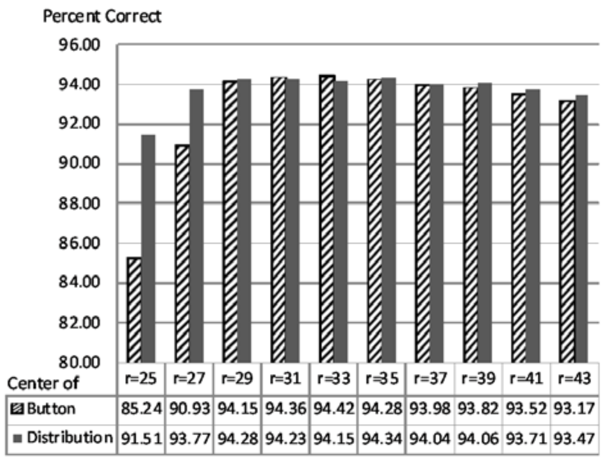
(a) Shift-off layout

(b) Shift-on layout

Fig. 3. Center point of keyboard button

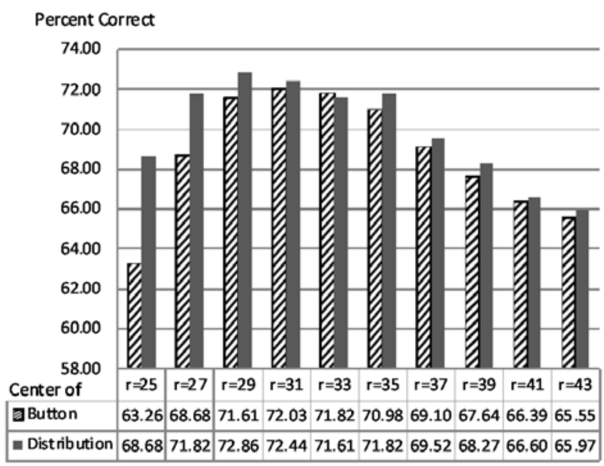
and low frequency as well as long, medium and short word length in the same average number of each set.

With the training set, we conduct the experiments using various radiuses and two kinds of keyboard layouts mentioned in the previous section. We let six users type all training set words in their style on the mobile phone, word by word, to get 480 coordinate sequences. Every coordinate sequences are processed by our algorithm to get all possible character sequences with its probability. Then, we select the character sequences that have most probability as the output words. We match 480 the obtained words against the correct words to find out the accuracy in character and word levels. In character level, we match and count the number of characters in words, and we match word by word in word level.



(a)

Fig. 4. Percent correctness in character level of various radius and layout



(b)

Fig. 5. Percent correctness in word level of various radius and layout

Fig. 4 and 5 show the accuracies at the different radiuses in character and word levels respectively.

In character level, the proposed method works well. It yields around 93 – 95% of correctness for most of the radiuses. The performance in word level drops to 63 – 73% since we applied only the character trigram model, it makes sense that its mechanism work well in character level. However, the method did not apply any statistics or information related to words. We plan to apply the language model in word level to improve correctness in the future works.

Table 2. The correctness of different word frequency

Frequency	Method	Character level (%)	Word level(%)
Low	Proposed	90.51	47.62
	Ordinary	88.72	52.38
Medium	Proposed	95.08	73.81
	Ordinary	81.97	38.10
High	Proposed	95.46	80.56
	Ordinary	85.35	55.56

Table 3. The correctness of different word length

Length	Method	Character level (%)	Word level(%)
Short	Proposed	81.61	52.38
	Ordinary	89.66	69.05
Medium	Proposed	93.00	66.67
	Ordinary	83.00	40.48
Long	Proposed	97.71	83.33
	Ordinary	85.42	33.33

Considering the center of button and the center of distribution keyboard layers, both yields not much difference in the result. But in overall, the center of distribution is slightly better than the ordinary center.

For the size of radius, the radius that is shorter than $\bar{r}$ yields a quite low accuracy when it is compared to the others. Both figures show that the suitable radius should be between 29 and 35 pixels. We then select the radius of 35 pixels because it has the highest percent of correctness of all radiuses in character level.

Finally, we conduct the experiment on the rest 20 random words in the test set to find out the actual performance of our algorithm against the ordinary typing method. The correctness of the ordinary method in character level is 54.43% and 48.33% in word level. While our algorithm perform 93.29% and 66.67% in character and word level respectively. This result shows that our algorithm provides the improvement in the text entry accuracy and the performance may increase if we introduce the language model in word level.

We also have more detail about these results. We separately consider the test set in two viewpoints i.e. word frequency aspect and word length aspect. These results show in table 2 and 3 respectively.

Words appear in corpus more than 1,000 times, between 100 to 1,000 times and less than 100 times are considered high, medium and low frequency word group, respectively. The proposed method provides more accuracy while the frequency is increase. On the other hand, the frequency does not affect the performance of the ordinary method because it only bases on the distance, it does not use any frequency information.

For the word length aspect, the short word length group has two to six characters per word while the medium word length group has seven to eleven characters per word and the long word length group has twelve to sixteen characters per word. The proposed and ordinary methods both base on the distance so the main different of these two methods is the language model. More character in word makes more information we can obtain from the model thus longer word length yields higher performance of the proposed method on both character and word level.

6 Conclusion and Future Works

This paper uses distance-based candidate generation and character trigram model to reduce error from text input method via virtual keyboard on mobile phones by word selection. We compare two kinds of keyboard layout. One is center of button keyboard layout. The other is center of distribution keyboard layout. The results show that the later layout is better. We found that the best radius is 35 pixels. Our algorithm performs better than the ordinary method about 1.7 times in character level and 1.38 times in word level.

For the future works, we plan to implement our algorithm on a real device in order to evaluate the processing time. We also plan to improve performance of candidate selection by including the language model in the word level into candidate selection method. By introducing the concept of word, it would yield a better performance. Furthermore, we aim to do word completion so as to reduce the number of keystrokes which is the main problem of using Thai keyboard.

In market, there are many brands and models of touch-screen mobile phones. Each model has difference screen size. Thus we plan to apply these proposed methods on various screen size of mobile phone.

Acknowledgments. This research is financially supported by Thailand Advanced Institute of Science and Technology Tokyo Institute of Technology (TAIST&TokyoTech), National Science and Technology Development Agency (NSTDA) Tokyo Institute of Technology (Tokyo Tech) and Sirindhorn International Institute of Technology (SIIT), Thammasat University (TU). The equipment used in this research is supported by the Telecommunication Research and Industrial Development Institute(TRIDI).

References

1. Bi, X., Smith, B., Zhai, S.: Quasi-qwerty soft keyboard optimization. In: CHI 2010: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 283–286. ACM, New York (2010)
2. Brewster, S., Hughes, M.: Pressure-based text entry for mobile devices. In: MobileHCI 2009: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 1–4. ACM, New York (2009)
3. Electronics, T..N., (NECTEC), C.T.C.: Benchmark for Enhancing the Standard of Thai language processing 2010 (BEST 2010), http://www.hlt.nectec.or.th/best/?q=node/10
4. Kristensson, P., Zhai, S.: Relaxing stylus typing precision by geometric pattern matching. In: IUI 2005: Proceedings of the 10th International Conference on Intelligent User Interfaces, pp. 151–158. ACM, New York (2005)

5. MacKenzie, I.S., Zhang, X.: Eye typing using word and letter prediction and a fixation algorithm. In: ETRA 2008: Proceedings of the 2008 Symposium on Eye Tracking Research & Applications, pp. 55–58. ACM, New York (2008)
6. Potipiti, T., Sornlertlamvanich, V., Thanadkran, K.: Towards an intelligent multilingual keyboard system. In: HLT 2001: Proceedings of the First International Conference on Human Language Technology Research, pp. 1–4. Association for Computational Linguistics, Morristown (2001)
7. Sarcar, S., Ghosh, S., Saha, P., Samanta, D.: Virtual keyboard design: State of the arts and research issues. In: 2010 IEEE, Students' Technology Symposium (TechSym), pp. 289–299 (April 2010)
8. Shieber, S.M., Baker, E.: Abbreviated text input. In: IUI 2003: Proceedings of the 8th International Conference on Intelligent User Interfaces, pp. 293–296. ACM, New York (2003)
9. Zhai, S., Hunter, M., Smith, B.: The metropolis keyboard - an exploration of quantitative techniques for virtual keyboard design. In: UIST 2000: Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology, pp. 119–128. ACM, New York (2000)

Baseline Detection for On-line Cursive Thai Handwriting Recognition Based on PCA

Pimolluck Jirakunkanok and Cholwich Nattee

School of Information, Computer, and Communication Technology,
Sirindhorn International Institute of Technology, Thammasat University
pimolluck.jirakunkanok@student.siit.tu.ac.th, cholwich@siit.tu.ac.th

Abstract. In the handwriting recognition, the baseline detection is one of the preprocessing processes. It is important for the efficiency of the handwriting recognition. For Thai language, a Thai word composes of vowels and tones located on above or below the consonants. This is different from other languages. Therefore, detecting the baseline of Thai handwritten words becomes a challenging task. In this paper, we propose a new approach to detect the baseline of Thai handwritten words using PCA algorithm. The principle of this method is to remove the outliers from words before conducting PCA algorithm to find the baseline direction. For finding the outliers, the proposed method finds the centroid of the handwritten word using three different methods including the mean method, the median method and the middle method. We evaluate the performance of the proposed technique using a real world dataset collected from various Thai natives. From the experiments, the mean method is the best among the three methods. Then, we compare the performance of baseline detection between the mean method and the ordinary PCA approach. The result shows that the proposed approach outperforms the ordinary PCA approach. The number of words which the proposed approach is able to detect baseline more accurately is 21 from 30 words.

Keywords: Baseline detection, on-line, Thai handwriting recognition, PCA.

1 Introduction

Writing is one of basic communication skills of human which is used generally to communicate or record information.

Nowadays, pen-based input devices such as tablets, personal digital assistants (PDAs), etc play an important role in human life. Using a pen is more convenient than using a keyboard or a mouse in some situations. For example, in the class, the student may take note of the lecture; therefore using a pen is faster and easier than using a keyboard. Moreover, a pen can be used everywhere because it is small-sized and portable.

For this reason, there are many research works about pen-based input devices. Due to the development of these devices, the handwriting recognition is one of basic software components which becomes important.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 56–67, 2011.

Handwriting recognition is an ability of computer to translate handwritten text corresponding to the printed characters. It can be divided into on-line and off-line recognition. For the on-line recognition, the input data is a sequence of the (x,y)-coordinates which represent the movement of pen tip, while the input data of the off-line recognition is an image of handwritten text. Palmondon and Srihari [6] describe the process of both on-line and off-line handwriting recognition.

In most of handwriting recognition systems, the preprocessing is an important process because handwritten words have many variations such as baselines, size, etc. Jaeger et al. [4] describe the preprocessing process of the on-line handwriting recognition which consists of several steps, i.e., computing baselines, normalizing size, smoothing, etc.

Since the baseline is utilized for the other steps in the preprocessing process such as normalizing size or deriving features, the baseline computation becomes an essential step. It is the main technique implemented in both on-line and off-line recognition.

In other languages, many methods have been proposed for baseline computation. For example, the baseline detection methods for the on-line handwritten text [1], [2], [3], [8] are used in English and Arabic languages. These methods are not suitable for Thai language because the structure of Thai words is more complicated. In Thai words, there are consonants, vowels, tones and special symbols.

For this reason, we propose a new baseline detection approach for on-line cursive Thai handwritten word based on PCA algorithm.

In this paper, we organize as follows: Section 2 describes background and related works including baselines of Thai word and baseline computation approaches. The details of the proposed approach are described in Section 3. The experiments and results are presented in Section 4. Section 5 discusses the results and the performance between the proposed approach and the basic PCA approach. The conclusion and future work are presented in Section 6.

2 Background and Related Works

2.1 Baselines of Thai Handwritten Word

In other languages, the baselines of word are different from Thai language. For example, in English language [4], the baselines have four lines as shown in Fig. 1.

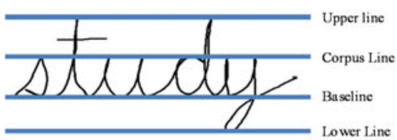


Fig. 1. Baselines of English handwritten word

In Thai language, there are 44 consonants, 17 vowels, four tonal symbols and two punctuation marks [5]. Thus, the baselines of Thai handwritten words can be separated into four-line levels [7] or five lines as shown in Fig. 2.

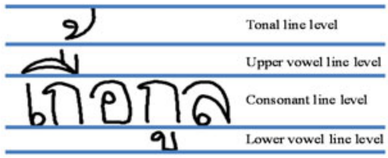


Fig. 2. Baselines of Thai word

Because Thai language has the tonal symbols which are located on the upper line above the upper vowel line level, the baselines of Thai word are different from other languages.

2.2 Baseline Computation Approaches

Several methods of baseline detection for on-line and off-line handwriting recognition have been proposed for many languages. In this paper, we focus on methods which are suitable for on-line handwriting recognition.

The Histogram-Based Method

This method is to find the baseline borders of words using the vertical density histogram [1]. The baselines which are detected by this method are four horizontal border lines as shown in Fig. 3.

The advantage of this method is robust and easy to implement. But it is suitable for words which have straight lines and long words. Moreover, the histogram method is not useful for pure cursive handwritten text.

The Local Minima Selection-Based Method

The purpose of this method [8] is to detect the minima points which are used to compute the linear regression formula in order to identify the direction of the baseline. The process of this method can be shown in Fig. 4.

This method has two advantages: it is easy to implement and it is suitable to be used for on-line handwritten text. However, this method has a limitation because it is suitable for handwritten text which has a lot of curve lines.

The PCA-Based Method

Principal Component Analysis (PCA) is a technique to find directions with greatest variation. PCA can be used to determine the baseline direction [3],[2]. The process of this method consists of four steps:

1) Perform the PCA for pixel distribution.
2) Determine baseline estimate direction.
3) Rotate data according to the direction and calculate the rotation angles.
4) Apply horizontal projection to detect the vertical position of the baseline.

This method is affected by the slope or the skewed, that is, before performing the PCA algorithm it needs to remove the slope or the skewed of the handwritten word. In addition, this method cannot detect baseline of short words or sub words. However, this method is easy to implement.

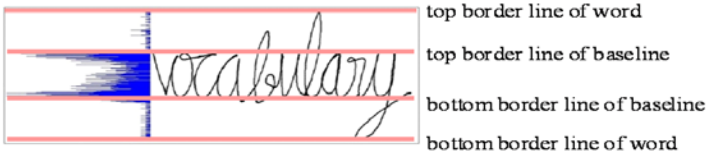


Fig. 3. Baseline of English word

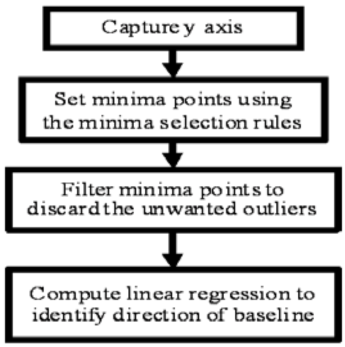


Fig. 4. Overall process of the local minima selection method

From the three methods, they are basically not suitable for Thai handwritten words. In the first method, the histogram-based method requires handwritten text which has a lot of straight lines but Thai words do not have. Thus, this method is not suitable for Thai words. In this paper, we will find the baseline detection method for on-line cursive handwritten word since the histogram-based method does not work well with cursive handwritten text.

The local minima selection-based method needs handwritten text that has a lot of curve lines at the top and bottom. This does not match to the shape of Thai characters. It therefore does not work well with Thai handwritten words.

In the PCA-based method, it is easy to implement but it is sensitive to the outliers, so it cannot work with Thai handwritten text. Since Thai word has many outliers such as vowel and tone that are affected baseline detection of the PCA method so that we propose a new method of baseline detection based on PCA algorithm.

3 The Proposed Approach

The proposed approach consists of six main steps as shown in Fig. 5. From all steps, some steps can be grouped into one process. Thus, the algorithm of baseline detection can be divided into three main processes, i.e., data acquisition, PCA computation, and finding and removing the outlier groups.

3.1 Data Acquisition

The data of Thai handwritten word is captured in form of (x,y)-coordinates representing the movement of the pen tip on the tablet. A stroke is a sequence

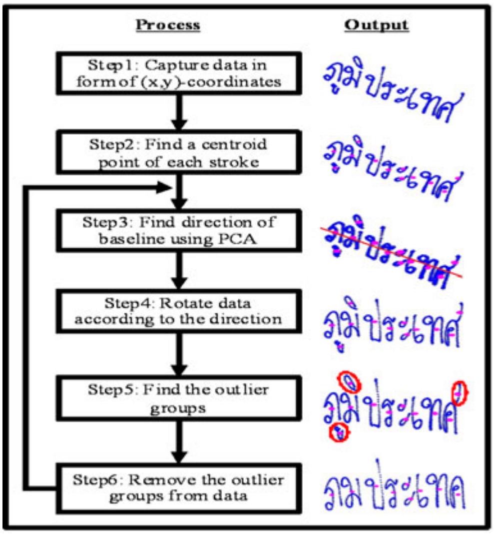


Fig. 5. Overall baseline detection

of (x,y)-coordinates while the pen tip is touching the tablet surface. In order to detect the baseline of the word, we first use centroid of each stroke (c_x, c_y). It is calculated based on a bounding box of stroke using the following equation.

$$(c_x, c_y) = \left(\frac{x_{max} - x_{min}}{2} + x_{min}, \frac{y_{max} - y_{min}}{2} + y_{min}\right) \quad (1)$$

where (c_x, c_y) is a coordinate of a centroid point, x_{max} is a max value of x-coordinates, x_{min} is a min value of x-coordinates, y_{max} is a max value of y-coordinates, and y_{min} is a min value of y-coordinates.

These centroid points are used in PCA computation to find a direction and rotate data.

3.2 PCA Computation

PCA is used to find the direction of the baseline. In the proposed approach, the centroid points of all strokes are used in PCA computation that is different from the existing technique for baseline detection that uses all points of data for PCA computation. There are three steps in PCA computation:

1) Perform PCA on these centroid points.
2) Define a direction of baseline by PCA eigenvector.
3) Calculate the rotation angles and rotate data according to the direction.

3.3 Finding and Removing the Outlier Groups

The outlier group is a set of outlier points which may interfere with the baseline detection using PCA technique, such as the centroids of Thai vowel strokes located on above and below the consonants. We assume that the outlier point is a point which is away from the centroid point of the stroke sequences. In this approach, we first remove all outlier points from data before conducting PCA computation, so that we can find the direction of baseline more accurately.

For computing the distance, we will consider only the y value on y axis because we assume that the outlier points are the vowels or the tones that are located upper or lower of the consonants. Thus, the distance between centroid points of all strokes and the centroid point of data can be denoted as follows.

$$D_i = |C - C_i| \tag{2}$$

where D_i is the distance between the centroid point of all strokes and the centroid point of the stroke sequences, C is the y-coordinate value of the centroid point of the stroke sequences, and C_i is the y-coordinate value of the centroid point of each stroke.

To compute the centroid point of the stroke sequences, we propose three methods that are, (1) the mean method, (2) the median method and (3) the middle method. We use these methods in order to find the centroid point of the stroke sequences that can provide removing the outliers accurately. The mean method is to calculate the centroid point of the stroke sequences from the mean value of all points. The mean is a way to summarize many data and ease to use but it is sensitive to the outliers. For the median method, the centroid point of the stroke sequences is computed from the median value of all points which is the middle number of a set of data. The median is more robust in the presence of the outliers than the mean. However, it has some disadvantages, i.e., it does not take all the data into account. In the last method, the middle method, we use a bounding box to calculate the centroid point of the stroke sequences as the equation (1). In this method, it does not consider all the data that is, it considers only the min and the max value of whole data. For this reason, it gives the middle value which is between the min and the max value. When the centroid point of the stroke sequences is defined, the distance is calculates as the equation (2).

To find the outlier groups, we will find the median of height which is calculated from height of all strokes and compare with D_i. In this paper, we will assume that if D_i is greater than α of the median of height, this point is the outlier point (the α is a value obtained from the experiment for Thai handwritten words). The process of finding the outlier groups is shown as Fig. 6.

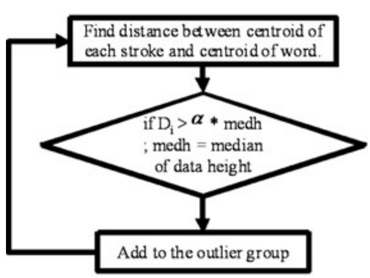


Fig. 6. Process of finding the outlier groups

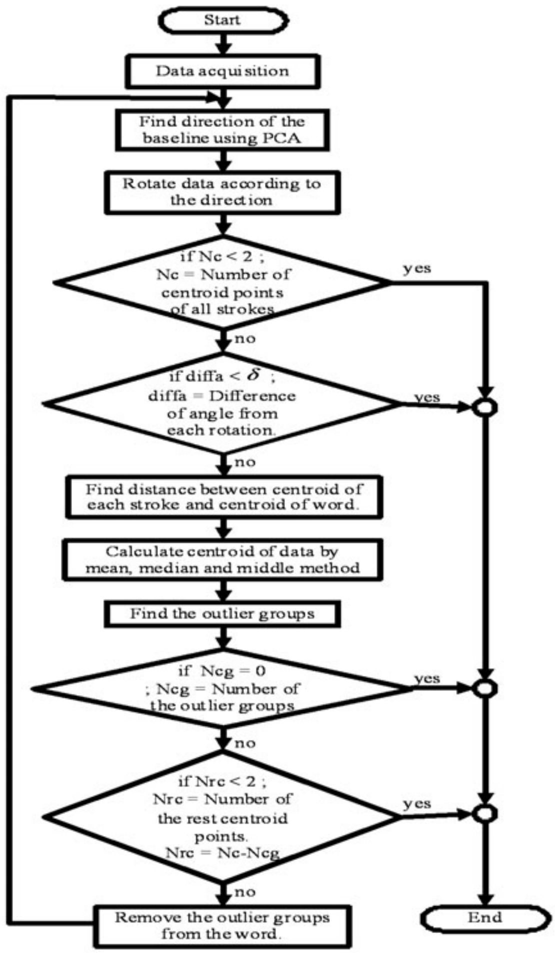


Fig. 7. Baseline detection process flow

From all above processes, the process flow is shown in Fig. 7.

There are three main processes including data acquisition, PCA computation, and finding and removing the outlier groups.

Firstly, the data acquisition is processed then the second process and the third process will be repeated until one of the four conditions is true. The four conditions can be described as follows:

1) If the number of centroid points of all strokes is less than two points ($Nc < 2$): this condition is used to check the number of centroid points before the next PCA computation because the input data for PCA computation should be greater than one point.
2) If the difference of angle in each rotation is less than δ degree ($diffa < \delta$): this condition is used to check the difference between the baseline direction which is just detected and the previous baseline direction (the δ is a value obtained from the experiment for Thai handwritten words).

3) If the number of the outlier groups is equal to zero ($Ncg = 0$): this condition is used to check there are no outlier groups to remove.
4) If the number of the rest centroid points of all strokes is less than two points ($Nrc < 2$): this condition is used to check if the outlier groups are removed, the number of the rest of centroid points must be greater than one point in order to use in the next PCA computation.

The first two conditions are checked after the PCA computation finished. After the process of finding the outlier groups finished, the condition 3) and 4) are checked.

4 Experiments and Results

In order to evaluate the performance of the proposed approach, we conduct experiments based on the collected dataset and compare the performance between the proposed approach and the basic PCA method.

The dataset is created by collecting Thai words from a Thai corpus named BEST2010 [9] provided by NECTEC. We select only words with length between 7–14 characters. We randomly choose 30 words. After that, the data is written by 10 Thai natives. Each writer writes 30 words. Firstly, the data is recorded in form of the straight word. Then, we rotate data using 10 different angles including 5, 10, 15, 20, 25, -5, -10, -15, -20 and -25 degrees. Finally, we get data having different start angles. The start angle of the straight word is assumed to be 0 degree and the rotated data have the start angle according to the angle which is rotated. Here, the number of all data is 3,300 words.

For the experiments, we first set the value of α in the process of finding and removing the outlier groups as 55% and the value of δ in the process of PCA computation as 0.1. These values come from the preliminary experiment. Then, we run the proposed technique varying the methods for finding the centroid point of the word into three methods: mean method, median method and middle method. We compare the performance of our technique with the basic PCA method on these data in order to detect the baseline. After the baseline is detected, the error angle is calculated from the following equation.

$$ErrorAng = |StartAng + RotateAng| \tag{3}$$

where $ErrorAng$ is the error angle of each word, $StartAng$ is the start angle of data and $RotateAng$ is the summation of rotation angles that comes from PCA computation.

For example, if the start angle of data is 15 degrees then the data is rotated two times into two angles that are, -14.04 and -0.6 degrees respectively. Thus, the summation of rotation angles is -14.64 degrees and the error angle is equal to 0.36 degrees that is calculated as the equation (3).

Table 1 shows the average error angles of each word from 11 angles and 10 writers using four baseline detection methods. In Table 1, the star (*) marks the least error angles of each word.

Fig. 8 shows the example results of baseline detection. The original word which is rotated into 10 degrees is shown as Fig. 8(a). Fig. 8(b), 8(c), 8(d) and 8(e)

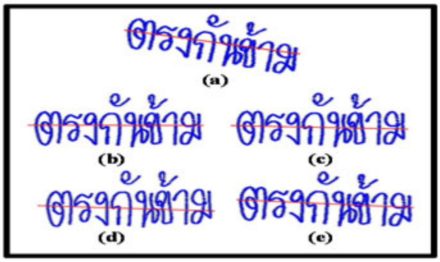


Fig. 8. Example results of baseline direction from four baseline detection methods

Table 1. The error angle of each word using four baseline detection methods

Word No.	The Proposed methods			Basic PCA
	Mean	Medain	Middle	
1	1.898	1.989	2.328	1.193*
2	1.175*	1.180	1.514	1.381
3	1.764	1.789	1.694	1.641*
4	2.299	2.188*	2.337	3.526
5	1.739	1.707	2.744	1.322*
6	7.748	7.775	9.665	4.623*
7	1.047*	1.145	1.957	1.234
8	1.756*	1.925	3.401	1.813
9	2.248*	2.609	2.886	3.173
10	1.537	1.481*	1.565	2.194
11	1.842*	1.962	3.694	1.954
12	2.718	2.576	4.432	2.101*
13	1.765	1.731	1.794	1.591*
14	1.956*	2.160	2.296	3.035
15	0.928	0.920	0.903*	1.116
16	1.361	1.319*	2.734	1.915
17	2.501	2.461*	3.730	2.590
18	1.214	1.193*	1.194	1.229
19	6.275	5.679	7.375	3.515*
20	1.825	1.849	1.548*	1.839
21	0.927	0.925	0.916*	0.981
22	1.123*	1.264	1.731	1.250
23	1.347	1.208*	2.163	1.308
24	2.060*	2.236	2.331	2.712
25	1.806*	1.957	4.440	2.204
26	1.867	1.821*	2.398	2.405
27	1.298*	1.365	1.578	1.723
28	1.110	1.040*	1.676	1.565
29	1.046*	1.055	1.166	1.445
30	2.652	2.831	2.767	1.878*

are the results of baseline detection from using the mean method, the median method, the middle method and the basic PCA method respectively.

5 Discussion

After we got the average error angles of each word using four baseline detection methods, we use the average error angles to measure the performance of each method. To measure the performance of four methods, we assume that the angle of the real baseline is 0 degree. The error angle will be 0 degree when the data can be rotated into the straight word that is the baseline can be detected precisely. Thus, the performance can be measure from the error angle. If the error angle is less, the performance of baseline detection will be good. In each word, we assume that the method which has the least error angle is given one efficient score.

From Table 1, we count the number of words which has least error angles of each method. The number of words is the efficient scores of each method.

Firstly, we compare the efficient scores of three proposed methods in order to find the best method for detection the baseline. Table 2 shows the efficient scores of three proposed methods from 30 words. From Table 2, the best method for detection the baseline is the mean method which has 14 efficient scores from 30 words. Then, the mean method which is the best method will be compared to the basic PCA method. The efficient scores of the mean method and the basic PCA method from 30 words are shown as Table 3.

In Table 3, the performance of the mean method is better than the basic PCA method because it has 21 efficient scores which are higher. From Table 4, the median method has 18 efficient scores which are higher than the basic PCA method so that its performance is better. From Table 5, the efficient scores of the basic PCA method are higher than the middle method. Thus, the performance of the basic PCA method is better.

From all words, we can classify into three groups including one-level, two-level and three-level. With the mean method, the average error angles of each level from all words can be shown as Table 6. In Table 6, the number of levels in a word affects to the performance of the baseline detection. That is, if the number of levels in a word increases, the performance of baseline detection will decrease.

Although the mean method is the best method among the three proposed methods, it may not detect the baseline of all words correctly. In some words, the baseline can be detected by the basic PCA method more accurately. In Fig. 9, we will show the example word from 30 words that the mean method can detect the baseline more accurately than the basic PCA method. In Fig. 9, the straight word is shown in Fig. 9(a) and the results with baseline using the mean method and the basic PCA method are shown in Fig. 9(b) and 9(c) respectively. In Fig. 10, we will show the example word that the basic PCA method can detect the baseline better than the mean method. From Fig. 10, the straight word is shown as (a) and the results of baseline detection using the mean method and the basic PCA method are shown in (b) and (c) respectively.

In this case, the first baseline direction of word may affect to the next baseline detection that is, if the first baseline has more slope, the word will be rotated according to that baseline that is the word is rotated in wrong direction. Thus, the next baseline detection may be wrong too.

Table 2. The efficient scores of three proposed methods

	Mean Method	Median Method	Middle Method
Score	14	12	4

Table 3. The efficient scores of the mean method and the basic PCA method

	Mean Method	Basic PCA Method
Score	21	9

Table 4. The efficient scores of the median method and the basic PCA method

	Median Method	Basic PCA Method
Score	18	12

Table 5. The efficient scores of the middle method and the basic PCA method

	Middle Method	Basic PCA Method
Score	12	18

Table 6. The error angles of each level in all words from the mean method

	One-level (1 word)	Two-level (21 words)	Three-level (8 words)
Error angle	0.927	1.674	3.093

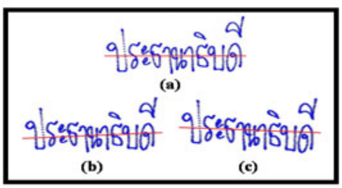


Fig. 9. The example word that the mean method provides better performance than the basic PCA method

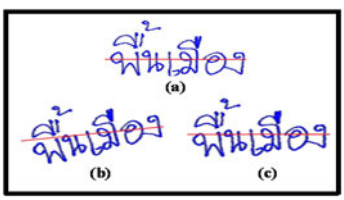


Fig. 10. The example word that the basic PCA method provides better performance than the mean method

6 Conclusion and Future Work

This paper has proposed a new approach to detect the baseline of Thai handwritten word based on PCA algorithm. For recording data, we use 30 Thai words which are selected randomly from Thai corpus name BEST2010 from NECTEC and these words are written by 10 Thai natives. Then these words are rotated into 10 angles. Thus, the number of all data is 3,300 words.

In the experiment, it is to find the best baseline detection method of three proposed methods, i.e., the mean method, the median method and the middle method. Then, we compare the performance of the proposed method with the basic PCA method.

From the experimental results, the mean method is the best baseline detection method among the three proposed methods. It has 14 efficient scores which are greater than two methods. Then, the performance of the mean method is compared with the basic PCA method. The result is the mean method can detect the baseline more accurately than the basic PCA method.

However, the proposed method cannot detect the baseline of all words correctly because it is affected by the first baseline detection. Moreover, this method cannot used to detect the short word such as one word because it is the limitation of the PCA algorithm. We plan to improve this in our future works.

Acknowledgment. This research is financially supported by Thailand Advanced Institute of Science and Technology-Tokyo Institute of Technology (TAIST&Tokyo Tech), National Science and Technology Development Agency (NSTDA) Tokyo Institute of Technology (Tokyo Tech) and Sirindhorn International Institute of Technology (SIIT), Thammasat University (TU).

References

1. Aida-zade, K.R., Hasanov, J.Z.: Word base line detection in handwritten text recognition systems. International Journal of Computer Systems Science and Engineering 4, 49–53 (2009)
2. AL-Shatnawi, A., Omar, K.: A comparative study between methods of arabic baseline detection. In: International Conference on Electrical Engineering and Informatics 2009 (2009)
3. Burrow, P.: Arabic handwriting recognition. Master's thesis, Master of Science, School of Informatics, University of Edinburgh (2004)
4. Jaeger, S., Manke, S., Reichert, J., Waibel, A.: On-line handwriting recognition: The npen++ recognizer. International Journal on Document Analysis and Recognition 3, 169–181 (2001)
5. Methasate, I., Sae-tang, S.: The clustering technique for thai handwritten recognition. In: Proceedings of the 9th Int'l Workshop on Frontiers in Handwriting Recognition (IWFHR-9 2004) (2004)
6. Plamondon, R., Srihari, S.N.: On-line and off-line handwriting recognition: A comprehensive survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 63–84 (2000)
7. Pornpanomchai, C., Batanov, D.N., Dimmitt, N.: Recognizing thai handwritten characters and words for human-computer interaction. The International Journal of Human-Computer Studies 55, 259–279 (2001)
8. Shamsuddin, M.R., Mohamed, A.: Baseline image classification approach using local minima selection. In: Badioze Zaman, H., Robinson, P., Petrou, M., Olivier, P., Schröder, H., Shih, T.K. (eds.) IVIC 2009. LNCS, vol. 5857, pp. 727–737. Springer, Heidelberg (2009)
9. Thailand : National Electronics and Computer Technology Center (NECTEC): Benchmark for Enhancing the Standard of Thai language processing 2010 (BEST 2010), http://www.hlt.nectec.or.th/best/?q=node/10

Ontology-Driven Accounting Information System: Case Study of the Stock Exchange of Thailand

Jira Yammeesri[1], Seung Hwan Kang[2], and Sim Kim Lau[3]

[1] School of Accountancy, University of the Thai Chamber of Commerce, Bangkok, 10400, Thailand
[2] Computer Information Systems, Payap University, Chiang Mai, 50000, Thailand
[3] School of Information Systems and Technology, University of Wollongong, NSW, 2500, Australia
jira_yam@utcc.ac.th, seung_h@payap.ac.th, simlau@uow.edu.au

Abstract. Accounting information systems have improved conventional information control. Due to corporate meltdowns, corporate governance has come under greater scrutiny to improve practices of managing, retrieving and extracting accounting information within organizations. This paper introduces a framework for an ontology-driven accounting information system as a mean to share information in line of common agreement. The proposed approach contributes to weakening of asymmetric information and information transparency problems.

Keywords: AIS, Ontology, Corporate Governance, Semantic Web.

1 Introduction

Corporate governance practices have come under increasing scrutiny for more than a decade. It is broadly defined as rules and incentives by which the management of a firm is directed and controlled so as to maximize profitability and long-term value of the firm to shareholders, while taking into account the interests of other stakeholders [3]. Corporate governance has been called for to raise the level of shareholders protection since several corporate meltdowns and frauds such as Enron, Tyco, and WorldCom, which has led to obliteration of billions of dollars of shareholder wealth, loss of thousand of jobs, and record-breaking bankruptcy filings. As a result, the Sarbanes-Oxley (SOX) Act of 2002 was quickly passed by the Congress in the United States [18]. The New York Stock Exchange and the NASDAQ (National Association of Securities Dealers Automated Quotations) stock market have proposed new listing standards that require companies to improve their corporate governance [18]. SOX Section 404 (known as SOX 404) requires public companies to report adequate company's internal control over financial reporting. Auditors need to verify management's reports and provide reports on the effectiveness of the internal control system. SOX 404 addresses the need of investors to have confidence in the financial reports issued by a company and also the process and control that are integral part of producing these reports.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 68–78, 2011.

Information Technology (IT) plays an important role in the financial reporting and it has become an integral part of companies' internal control systems. It provides day-to-day operations and is used to perform more functions from input devices to sophisticated financial management planning and processing systems. Business transactions of most companies are routinely processed electronically [15]. Companies have faced with greater expectation to provide more accurate, visible, and timely information while ensuring the privacy and security of their corporate information assets are protected. Sonde [24] argues that Accounting Information Systems (AIS) have not met expectations of organizations because information is not easily accessible.

Modern AIS use the World Wide Web (WWW) technologies. The Web has become the channel of information disclosure and making accounting information widely available to external users. In particular, from corporate governance perspective, this leads to asymmetric information problems where one party has more or better information than the other. Moreover, the diverse range of Web users may result in users not having a common understanding of accounting and financial terminologies used.

Furthermore, financial Web resources in the existing Web environment are mainly published for mass media. This is generally not a context-sensitive approach. For example, when a researcher uses keywords to search in an online financial database (generally the keywords are considered to be "valid" keyword in the domain of search), the results of the query often do not match as intended. The current Web contains vast amount of financial information resources, such as trading summary, financial reports and stock prices. Making these resources available in a more structured way is one of the goals of our long term project efforts

This study aims to support and improve quality of accounting information by integrating ontology in the AIS. Ontology approach provides a common domain to enable users to have further access to financial and accounting information of their targeted corporations with a clear understanding of the domain among all the parties concerned. Thus it opens the alternative information channel to weakening the asymmetric information problem and information transparency problems.

2 Literature Review

2.1 Corporate Governance

The concept of corporate governance is difficult to define because it potentially covers a large number of different economic phenomena. Traditionally, corporate governance is associated with a principal (owner or shareholders or sometimes investor) and an agent (manager) relationship; shareholders employ managers to manage firms to maximize shareholders' benefits [23, 17]. Corporate governance depends on corporate governance system as well as its associated mechanisms such as information transparency and disclosure requirements. The quality of transparency and information disclosures can contribute to the effectiveness of shareholders' protection system

[16]. Information of the firm disclosed to shareholders should be complete, adequate, reliable, creditable and timely. The information usually includes financial situation, list of major shareholders and board of directors, governance structure, and the firm's objectives and its policies. Positive and negative information should be disclosed regularly. However, such information should not confuse or mislead users because shareholders as well as investors require such information for decision making and evaluation of performance of management.

In addition, quality of transparency and information disclosure significantly depends on accounting and auditing standards as well as financial systems. It has been documented that international accounting standard is needed to contribute to higher quality of transparency and information disclosure in the firm. In addition, independence of auditing and good financial reporting systems should be called for. This is because these systems can ensure information disclosed to shareholders and other investors is reliable and adequately credible [19, 2, 16]. For example, in the United States and United Kingdom the accounting standards are set by professional bodies of accountants, who ultimately understand the rules. In some countries, such as Thailand, there is an inadequate number of well-qualified accountants and auditors as well as professional self-regulatory body [1, 21]. In 1999, the Stock Exchange of Thailand (SET) required all listed firms to establish audit committee to be responsible for examining the quality and reliability of the firm's financial reports.

Ontology defines the domain of interest and domain of conceptualization. It works as a standardized list of vocabularies to enable users in the domain, in this case users of accounting information system, to communicate with each other with the same language. This way we propose that ontology can minimize asymmetric information problems. For corporate information disclosure, the ontology can be used by users to navigate the information repository of an organization in an effective manner.

2.2 Accounting Information Systems

AIS provide financial information to monitor and manage organizational resources together with conventional accounting controls: inputs, processes and outputs [13]. Thus study of AIS is important because it covers not only the principle but also practice of accounting with the design, implementation and maintenance of AIS. Enterprise resource planning (ERP) is an example of IT system that uses IT to perform business functions from input devices to sophisticated financial management planning and processing systems. It focuses on how information is exploited in AISs to produce value added supports in business.

Financial data have been represented in computing systems such as AISs and ERP, operated by automated procedures like eXtensible Business Reporting Language (XBRL), and then integrated into business settings. This is an important feature that can provide answers to questions such as “Why did the revenue fall very quickly in the last two weeks?”, ‘What will happen to the stock prices?” To support the above examples of business queries, Artificial Intelligence (AI) techniques can be applied to formalize many interesting financial data to analyze information such as using technique of data mining and supporting decision making using decision support system in AISs. However, this way of utilizing financial data is still not flexible enough to accommodate the need of AIS.

Kabak and Beam [14] explain that there is a need of two dimensions of accounting. The first dimension is a traditional way of recording and presenting information for financial reporting purposes. The second dimension is a way of relating to management's question that may need a different way of presenting information. For instance XBRL is introduced to answer electronic communication between business and financial data reporting. The initial idea was to propose a mechanism that can tackle non-standardization of controls, continuation of controls, fraud detection, and lack of transparency of transaction changes. With XBRL taxonomies, it can incorporate business rules and add data layer validation. Data validation is the process of checking if the data satisfy certain criteria. For fraud detection, validation takes place at the data layer instead of at the application layer. As XBRL supports continuous auditing, interfaces between different systems can be standardized. This way an Embedded Audit Module (EAM) can also be standardized. EAM is an integrated part of auditing tools to monitor specific transactions. A study conducted by Premuroso and Bhattacharya [20] shows that there is greater transparency to corporate governance and operating performance by adopting the XBRL format. Thus XBRL endorses standardized data format to exchange data. In turn, it promotes standardized controls, standardized testing, and improves audit processes. However, XBRL is mainly used as a means to exchange data between applications and maps between current systems. It is still limited as a business reporting language not as a financial analytical language.

One of the challenges in the development of AIS is the difficulty in representing knowledge domain explicitly. A domain expert often finds it is very difficult to expressively describe his or her domain knowledge to the system engineer. Thus the need to build better and interoperable interpretation approach is essential to solve the ambiguity problem in knowledge representation. There is also a need to share meaning of terms in a given domain by an individual or the community. In order to effectively deploy a shared term or terminology, a clear understanding of the domain needs to be agreed by the community of practice. In other words, a way to conceptualize the given domain by the community has to be published and broadly agreed among the interest groups. This approach ensures situations where misunderstanding and misinterpretation of conceptual modeling in terms of taxonomy will not occur. This is known as an ontological approach.

2.3 Ontology

The term ontology is defined as "a formal explicit specification of a shared conceptualization" [11, p.2], or "a formal specification of a shared conceptualization" [6, p.28]. Primarily, ontology works like a standardized list of vocabularies that allows everyone to communicate with each other in a common language and common understanding of terminologies [11]. A key idea of constructing ontology is to share common understating of domain knowledge. For example, *a/c* is used to indicate an *account*, and this terminology is well-understood to everyone in the accounting and finance discipline. However there are other concepts or representations that can be associated with the term *a/c* such as bank or financial credit account. In this case *a/c* has a similar meaning it is associated with, we call this the synonym. Ontology seeks to demonstrate a clear understanding of the domain of interest for well structured Web resources in the Semantic Web environment. The Semantic Web is described as an extension of the current Web in which information is given well-defined meanings (known as Web ontology), and enable computers (in particular software agents that behave like personal assistants) and people to work in a cooperative manner [4].

3 OntoAIS

OntoAIS (Ontology-driven Accounting Information System) aims to develop a well-formed conceptual model of the financial data obtained from the SET. Then we have applied an ontological re-engineering approach to build our own ontology. With the well-formed conceptual model of the financial data, we produce a reliable query processing module which can be integrated into a Web server through simple n-tier client-server architecture. The end-user query can be widened or narrow-downed to specific concepts such as daily transactions by a particular investor of the trading summary. The former information is already available in the SET and the latter is also available within the SET that maintains details in a semi-structured format.

In brief, the client accesses ontological information available on an application server that is built with the SPARQL query. It uses its own ontology that has a common semantic source. SPARQL is the query language for RDF (Resource Description Framework) as well as OWL (Web Ontology Language) [26]. The semantic source maps the concepts and their relationships about the SET. A database server stores additional information such as instances of the concepts (known as classes of ontology) and value of their attributes (known as properties of the classes).

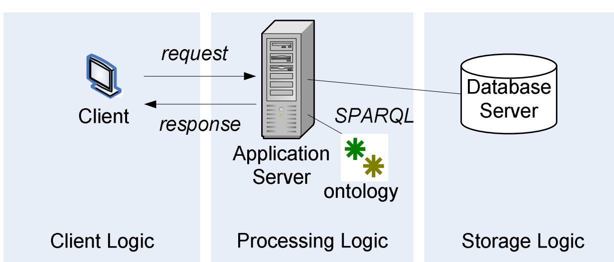


Fig. 1. Ontology-driven accounting information system on the Semantic Web environment

3.1 Ontology Development

Consider a trading summary ontology that we have developed for the SET. Figure 2 shows tabulated information about the trading summary on Monday, 12 April 2010 from the SET. Firstly, it can be noticed that there is an ambiguous term called ‘Type’ as it can refer to a type of investors or investment or investment sources. Next, we assume that the term ‘Net’ refers to the value obtained by subtracting the value in ‘Buy’ from ‘Sell’. The term ‘Unit’ refers to million Thai currency (Baht). The term ‘more>>’ refers to forwarding current page to other page for additional available information. In the following page, our first question is answered as ‘Type’ instead refers to ‘Investor Type’ however there are still more unanswered questions such as the meanings of ‘Buy’, ‘Sell’ and ‘Net’. It is also unclear whether % means an increase in a percentage or decrease in a percentage based on daily, weekly or monthly. In the ‘Net’ column, it shows - with a red color and + with a green color. We assumed that it means - for loss and + for profit. In other words, it may be easy for human users to understand these terms of *Buy*, *Sell* and *Net*. However it is not clear in situation where it is machine processable.

Trading Summary

As of 12 Apr 2010 Unit: M.Baht

Type	Buy	Sell	Net
Institution	2,437.43	4,864.22	-2,426.79
Proprietary	3,478.65	3,647.55	-168.90
Foreign	6,755.98	8,865.95	-2,109.97
Individual	22,537.95	17,832.29	4,705.66
Total Trading Value		35,210.01 M.Baht	

more>>

Daily as of 12 Apr 2010

Unit: M.Baht

Investor Type	Buy		Sell		Net	
	Value	%	Value	%	Value	%
Local Institutions	2,437.43	6.92	4,864.22	13.81	-2,426.79	-
Proprietary Trading	3,478.65	9.88	3,647.55	10.36	-168.90	-
Foreign Investors	6,755.98	19.19	8,865.95	25.18	-2,109.97	-
Local Individuals	22,537.95	64.01	17,832.29	50.65	4,705.66	-

Fig. 2. The trading summary on Monday, 12 April 2010 from the SET

In summary, early finding shows that financial information is presented in the form of the current WWW that may mislead the machine users. In particular, such information can be referred asymmetric information problem in corporate governance as it confuses or not useable for non-financial experts.

3.2 Ontological Engineering

A large number of ontologies have been developed by different researchers using different methods, methodologies, languages and tools. This set of efforts is referred to as ontological engineering [10, p.5]. In particular, we will discuss the Uschold and King' method that consists of the following four phases: 1) identify purpose and scope, 2) building the ontology, 3) evaluation, and 4) documentation [8].

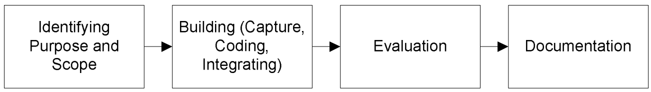


Fig. 3. Illustrated Uschold and King's method (Source: Gómez-Pérez et al. 2004 p.115)

The first phase of *identifying purpose and scope* includes clear understanding of possible uses and intended uses of the ontology. It may include identifying the relevant terms on the domain. Our finding shows that the SET has six categories on its website: *About SET, Companies / Securities Info., Prices & Statistics, Products & Services, SET News* and *Supervision & Regulations*. Each category has several subcategories. For example, *About SET* has *Mission & Vision*, *SET Overview*, *Corporate Reports*, *SET Holidays*, *Job Opportunities* and *Map to SET*. Our current approach focuses on a self-adaptive development to deal with the dynamic requirements and continual user involvement. Once we have an idea of what we want to know, we start to list the important terms that we need. These include basic concepts, properties (characteristics of a concept) and relationships. It is a repeated process when there are inconsistent, duplicated or missing concepts.

The second phase of *building* the ontology consists of three activities: capture, coding, and integrating existing ontologies. Ontology capture includes identifying key concepts and relationships in the domain of interest, producing precise unambiguous text definitions for such concepts and relationships, identifying terminology that refers to such concepts and relationships, and reaching an agreement. Ontology coding involves specifying the ontology by placing the underlying representational terms, choosing a presentation language, and writing the code. Integrating existing ontologies refers to joining all or part of existing relevant ontologies together. Thus, we construct concepts by starting from categorizing different terms according to their relations in the ontology. A key idea of constructing such an ontology is to share common understating of domain knowledge, in this case the SET. Figure 5 shows an example of the SET ontology constructed for this research. To construct the Stock Exchange Thailand (SET) ontology, Protégé editor is used. Concepts (also known as classes) are built by categorizing different terms according to their relations in the ontology. Figure 4 shows a graphical representation of the SET ontology that specifies the concepts of SET related information, services and products for decision support in financial practices, and sharing of accounting data within the organization. There are over 100 defined concepts of the SET ontology. For instance, the concept of Stock has 11 sub-concepts. Each sub-concept of Stock is disjoint. This means it must belong to an industry only. That is, company α that is classified as the *Technology* industry cannot be classified as a *Resources* industry at the same time.

Fig. 4. SET ontology with partial information

The third step of *evaluation* refers to "the correct building of the content of ontology, ensuring definitions… in the real world" [8, p.3]. In this step, it includes evaluation of each individual definition and axiom, a collection of definitions and axioms imported from other ontologies, and definitions that can be inferred using other definitions and axioms [9]. For illustration purposes, we will use the trading summary as shown in Figure 2.

Figure 5 shows ontological information of the trading summary in the SET ontology. In Figure 5 we can infer several facts from these relationships. All classes are subclasses of *Thing*. For example, there are three subclasses: SET, By Date and Investor. Each subclass has several subclasses of its own. For example Investor Group consists of Buy, Sell and Net. Investor consists of Local Institutions, Proprietary Trading, Foreign Investors and Local Individuals. In Web Ontology Language (OWL), we say that the subclasses of Investor are disjoint; that is there is no instance that belongs to another subclass. These relationships can help us to do reasoning. For example the property hasValue of Buy tells us that if we want to prove 2,426.79 million Baht, we know that it has value 2,437.43 million Baht. The property hasValue of Sell is 4,864.22 million Baht. Thus, we can conclude that the property hasValue of Net. A simple SPARQL query can return the result. The query consists of two parts: the SELECT clause identifies the variables to appear in the query results, and the WHERE clause provides the basic graph pattern to match against the data graph (See Figure 6).

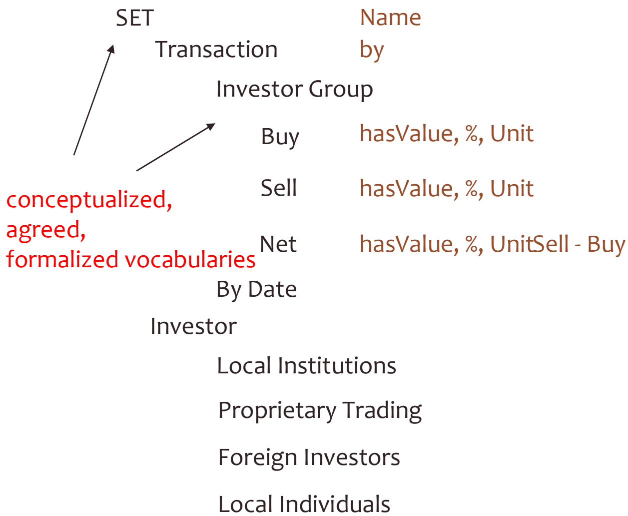


Fig. 5. Ontological information of the Trading summary in the SET ontology

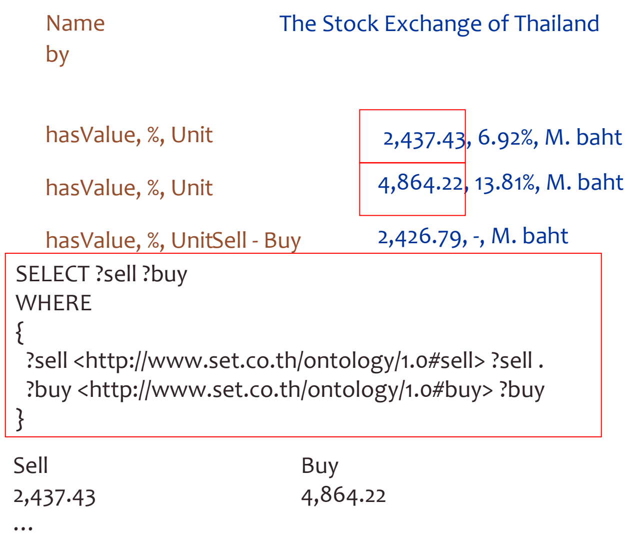


Fig. 6. Example ontological information of the trading summary in the SET ontology

The fourth step of *documentation* includes setting up guidelines for documenting ontologies so that it addresses problems of inadequately documented existing knowledge bases and ontologies such that the fundamental concepts are defined in the ontology. For instance, we had not considered the term 'company' in our initial examination as it is not a part of the SET. However, we have to define it in our ontology to allow contact details of the company to be asserted, which are not available from the SET. Therefore, we have decided to include company as a term of interest as well as properties associated with company.

4 Conclusion

It is essential to improve disclosure of financial information such as managing, retrieving and extracting accounting information within organizations, as a mean of sharing information in line of common agreement in practice. This way, the system not only provides clear understanding of terms but also unambiguous definitions of complex concepts and classifications. Moreover, users have access to financial and accounting information of their targeted corporations for corporate performance and trends. This study demonstrated improved quality of financial information by integrating ontology in the AIS. Thus it contributes to weakening the asymmetric information and information transparency problems in corporate governance practices.

In future research, we will investigate OntoAIS as a way to monitor fraud in the business.

References

[1] Alba, P., Claessens, S., Djankov, S.: Thailand's Corporate Financing and Governance Structures: Impacts on Firm's Competitiveness. In: Conference on Thailand's Dynamic Economic Recovery and Competitiveness, Thailand, May 20-21 (1998)

[2] Asian Development Bank. Corporate Governance and Finance in the Five Affected Countries. In: Capulong, V., Edwards, D., Webb, D., Zhuang, J.(eds.): Corporate Governance and Finance in East Asia: A Study of Indonesia, Philippines, and Thailand, Asian Development Bank, Manila, vol. 2 (2001)

[3] Berle, A., Means, G.: The Modern Corporation and Private Property. Macmillan Publishing, New York (1932)

[4] Berners-Lee, T., Hendler, J. and Lassila, O.: The Semantic Web. Scientific American (2001), http://www.sciam.com/2001/0501issue/0501berners-lee.html (accessed June 1, 2010)

[5] Bonsón, E., Cortijo, V., Escobar, T.: Towards the global adoption of XBRL using International Financial Reporting Standards (IFRS). International Journal of Accounting Information Systems 10(1), 46–60 (2009)

[6] Borst, W.: Construction of Engineering Ontologies for Knowledge Sharing and Reuse, 243 pages. University of Twente (1997)

[7] Cormier, D., Ledoux, M.-J., Magnan, M.: The use of Web sites as a disclosure platform for corporate performance. International Journal of Accounting Information Systems 10(1), 1–24 (2009)

[8] Gómez-Pérez, A., Juristo, N., et al.: Evaluation and Assessment of Knowledge Sharing Technology. In: Mars, N. (ed.) Towards Very Large Knowledge Bases - Knowledge Building and Knowledge Sharing, pp. 289–296. IOS Press, Amsterdam (1995)

[9] Gómez-Pérez, A.: Ontological Engineering: A State of The Art. Expert Update: Knowledge Based Systems and Applied Artificial Intelligence 2(3), 33–44 (1999)

[10] Gómez-Pérez, A., Fernández-López, M., et al.: Ontological Engineering: with examples from the areas of Knowledge Management. e-Commerce and Semantic Web. Springer, Heidelberg (2004)
[11] Gruber, T. (ed.): Toward Principles for the Design for Ontologies used for Knowledge Sharing. Kluwer Academic Publishers, Padova (1993)
[12] Hendler, J.: Agents and the Semantic Web. The IEEE Intelligent Systems, 30–37 (March/April 2001)
[13] Jones, F., Rama, D.: Accounting Information Systems. South-Western Publisher (2005)
[14] Kabak, I., Beam, T.: Two-Dimensional Accounting: A New Management Demand, Industrial Management (November 1, 1990), `http://www.entrepreneur.com/tradejournals/article/print/9295980.html` (accessed March 10, 2009)
[15] Li, C., Lim, J.-H., Wang, Q.: Internal and External Influence on IT Control Governance. International Journal of Accounting Information Systems 8, 225–239 (2007)
[16] Limpaphayom, P.: Thailand. In: Capulong, V., Edwards, D., Webb, D., Zhuang, J. (eds.) Corporate Governance and Finance East Asia: A Study of Indonesia, Republic of Korea, Malaysia, Philippines, and Thailand, Manila, Philippines, vol. II (2001)
[17] Mayer, C.: Financial System and Corporate Governance: A Review of the International Evidence. Journal of Institutional and Theoretical Economics 154(1), 144–164 (1998)
[18] Monks, A.G.R., Minow, N.: Corporate Covernance, 4th edn. Wiley-Blackwell (2008)
[19] Noy, N.F., Rubin, D.L., Musen, M.A.: Making Biomedical Ontologies and Ontology Repositories Work. IEEE Intelligent Systems 19(6), 78–81 (2004)
[20] OECD. OECD Principles of Corporate Governance. OECD Publications, France (1999)
[21] Phruithithada, U.: The Stock Exchange of Thailand: Consolidated and Entity Financial Statements, December 31, 48 pages. PricewaterhouseCoopoers ABAS Limited (2009)
[22] Prowse, S.: Corporate Governance Emerging Issues and Lessons from East Asia, The World Bank (1998)
[23] Robertson, S.: A tale of two knowledge-sharing systems. Journal of Knowledge Management 6(3), 295–308 (2002)
[24] Shleifer, A., Vishny, R.: A Survey of Corporate Governance. The Journal of Finance LII(2), 737–782 (1997)
[25] Sonde, T.: Accounting and Finance Systems: Unlocking the Hidden Value. The CPA Journal LXXVIII(3) (2008), `http://www.nysscpa.org/printversions/cpaj/2008/308/p70.htm` (accessed March 14, 2009)
[26] World Wide Web Consortium. Semantic Web (2001), `http://www.w3.org/2001/sw/` (accessed June 13, 2010)
[27] World Wide Web Consortium. SPARQL Query Language for RDF (2008), `http://www.w3.org/TR/rdf-sparql-query/` (accessed June 13, 2010)

Discovering Relationship between Hepatitis C Virus NS5A Protein and Interferon/Ribavirin Therapy

Saori Kawasaki[1], Tu Bao Ho[1], Tatsuo Kanda[2], Osamu Yokosuka[2], Katsuhiro Takabayashi[3], and Nhan Le[1]

[1] Japan Advanced Institute of Science and Technology
[2] Graduate School of Medicine Chiba University
[3] Chiba University Hospital, Japan

Abstract. Less than a half of hepatitis C patients respond to the current therapy by peg-interferon combined with ribavirin, and the genomic basis of those drug resistance remains unknown. It is recognized that the emerging challenge in the development of therapies for hepatitis C is new functions and mysteries of hepatitis C virus (HCV) non structural 5A (NS5A) protein. Different from current studies using small experimental samples to analyze relations between the HCV NS5A region and the response to interferon/ribavirin therapy, we introduce a data mining-based framework that exploits the largest available HCV database. This paper focuses on the methodology and early results of analyzing NS5A data.

1 Introduction

Hepatitis C is an infectious disease affecting the liver, caused by the hepatitis C virus (HCV) [1]. The worldwide prevalence of HCV infection is estimated to be approaching 200 million people. HVC persists in the liver of about 85% of infected people and the persistent infection can be treated with medication. In the evolution of HCV treatment, most early studies were naturally based on using standard interferon (IFNα) and later pegylated interferon (PEG IFNα). Whereas treatment with IFNα alone achieved only modest success, the addition of the antiviral agent ribavirin (RBV) greatly improved responses. However, more than a half of the infected individuals with chronic disease do not achieve sustained clearance of hepatitis C virus [2], [3], and the genetic basis of resistance to antiviral therapy is unknown [4].

To optimize current therapeutic strategies and to develop new therapies, a deep understanding of the mechanism of resistance to IFN and ribavirin will be essential [6], [7]. It is also well recognized that "the clinical success of HCV-targeted drugs will depend on their ability to suppress all viral variants as well as prevent the emergence of resistant viruses" [8].

HCV has length of approximately 9,600 nucleic acids that encodes a single polyprotein (Figure 1), consisting of structural proteins (Core, E_1, E_2, p7) and nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A, NS5B). In the development of therapies for hepatitis C, the research has focused predominantly

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 79–90, 2011.

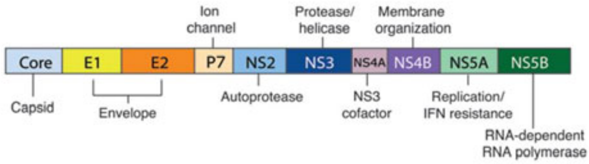


Fig. 1. The HCV Open Reading Frame [5]

on inhibitors of the viral enzymes NS3 protease and the RNA-dependent RNA polymerase NS5B, but recently on new functions and new mysteries of NS5A protein [2], [3]. NS5A is known as the protein most reported to be implicated in the interferon resistance [9]. Indeed, NS5A has emergently attracted tremendous attention in the HCV field and presents a very promising target for anti-HCV therapy.

In an effort to understand the relation between NS5A and IFN/RBV therapy, numerous studies have been performed, typically [3], [7], [4], [6], [2], [5], [10]. Some general mechanisms of NS5A resistance to IFN and ribavirin have been described and surveyed in [9], [11], [12]. However, this field is controversial and deep understanding of inhibitions remains unknown [11].

Different from most work on this problem, which is usually based on a small number of experimental samples, we introduce a data mining-based approach to discovery of the relationship between NS5A and the response to IFN/RBV therapy. By exploiting a large amount of unlabeled NS5A sequences and link them with available labeled NS5A sequences, we target to detect new and significant relationships between NS5A and IFN/RBV therapy. This paper presents our proposed framework and early results on data encoding and analysis.

2 Background

2.1 NS5A Protein

The non structural protein 5A (NS5A) has RNA-binding activities. Figure 2 shows the linear representation of the NS5A proteins where the grey zones are sites of interest, having name and amino acid positions. The typical functions of NS5A include:

1. NS5A interacts with IFN cellular antiviral pathways and thus inhibits the IFNα activity, typically either binding and inhibiting the IFNα-inducible double-stranded RNA-activated PKR (Protein Kinase R) or interacting with the IFN pathway in a PKR-independent manner [9], [11].
2. The mutations in the NS5A resist the IFN therapy. The 40 first amino acids of the PKR binding domain are of high level of variability, called Interferon Sensitivity Determining Region (ISDR), involve in sensitive/resistance to IFNα in HCV genotype 1b. However, recent studies have questioned the existence of an ISDR in NS5A and identified another domain of 27 amino

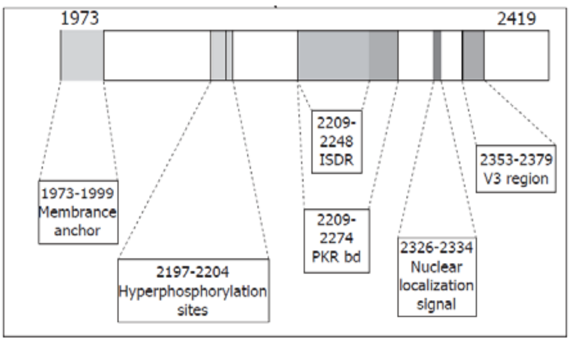


Fig. 2. Linear representation and the main interesting domains of NS5A protein [9]

acids in length, called V3 domain, that shows a high quasispecies diversity relating to the success of the IFN-based therapy [9].

Among the viral factors examined, viral genotype seems to be the most important predicting factor [9].

The HCV NS5A protein is also viewed in three domains, designated domains I, II and III, which are separated by low complexity sequence blocks of repetitive amino acids (LCS I and II). Domains II and III are more variable among HCV genotypes. Recent studies showed that domain II contains the PKR and the HCV NS5B binding domains and domain III could influence RNA replication [13]. Some open research questions can be raised:

- Concerning the role of NS5A in RNA replication, most replication phenotypes mapping to the domain I of NS5A. What is the remained enigmatic role of the domains II and III (DI, DII and DIII) in the HCV lifecycle [3]?
- Can V3 region be a more accurate biomarker than the ISDR region [10]?
- The larger size of NS5A samples may be necessary for understanding its role in the therapy outcome of HCV 1a/1b infection [4]. How to deal with it when NS5A data related to IFN/RBV therapy responders/non-responders are limited?

2.2 Interferon and Ribavirin

Interferons are a group of small proteins made by the human body in response to viral infections. IFN therapy is thought to work by stimulating processes within cells which help to slow down the reproduction and growth of the virus. These processes can increase the body's immune response to the virus so that you can fight the infection more effectively.

Most of early clinical trials used conventional interferon IFNα, and more recent clinical trials used pegylated interferon alpha (PEG-IFNα) [14]. PEG-IFNα is able to stay in the circulation for longer as it is less likely to leak out into other tissues and it is also filtered and removed by the kidneys at a slower rate.

Ribavirin is a drug that has activity against some viruses. The way it works in hepatitis C is not completely understood. Taking ribavirin alone is not effective against HCV infection in the long term. When ribavirin is used in combination with the drug interferon, researchers have found that about twice as many people as those using interferon alone show a long term clearance of detectable HCV from the blood.

The general mechanisms of resistance to IFNα and ribavirin and in particular NS5A and IFNα resistance and many questions are open to uncover are described in [12]:

– No significantly different virologic response rates between HCV subtypes were reported.
– It must be assumed that additional mutations outside the NS5A gene but within other genes of the HCV ORF are important for determinations of IFN sensitivity.
– Nothing is known about the frequency of UA and UU dinucleotides before and during IFN based therapy in correlation with HCV RNA clearance.

3 Discovery of Relationships by Data Mining

3.1 The Project Framework

Figure 3 describes our project framework to achieve the research target. The key idea is to exploit the large amount of unlabeled data to find interesting patterns in the NS5A sequences then detect their relations to IFN/RBV therapy by labeled data. To this end the framework consists of three steps. First is to encode the original NS5A sequences into the forms that can be exploited by various data mining methods. Second is to find interesting sequential patterns in both detected clusters or individual sequences and from different views, specially from the domain knowledge on NS5A and its resistance to interferon and ribavirin. Third is to link those patterns to labeled sequences for detecting expected relations. The essence of this step is to develop new semi-supervised learning methods that can exploit limited volume of labeled data by the connection to the findings in unlabeled data. The steps are not necessarily sequential but can be mixed in appropriate schemes.

3.2 The Datasets

Two kinds of data related to IFN/RBV therapy:

– *Unlabeled dataset* (in terms of two groups of response and non-response patients): About 5000 NS5A sequences belonging to 6 genotypes of which the genotype 1 divided into subtypes 1a, 1b and and 1c, and the genotype 2 divided into subtypes 2a and 2a. These data were taken from the database of Nagoya City University and GenBank.

Given:

1. *Un-labeled data: About 5000 sequences of NS5A proteins each belongs to one genotype or known subgenotype.*
2. *Labeled data: About 270 NS5A sequences each from one of either IFN/RBV responder or non-responder.*
3. *Domain knowledge about mechanism of resistance to IFN/RBV.*

Find: *Relations between NS5A and IFN/RBV therapy responders or non-responders.*

Method:

1. *Encoding the NS5A sequences:*
   (a) *Non-transformed sequences.*
   (b) *Transformed sequences, by*
      i. *k-gram with different k.*
      ii. *active motifs.*
      iii. *descriminant patterns*
2. *Find the relations in terms of interesting subsequences in NS5A sequences of given genotypes or in groups of unlabeled sequences obtained by clustering. The interestingness is investigated in different views such as:*
   (a) *Criteria in sequence analysis such as common longest/shortest subsequences, closed and maximal frequent subsequences.*
   (b) *High discriminant and covering ability patterns.*
   (c) *Relations to the resistance mechanisms of IFN/RBV or domain knowledge on NS5A.*
3. *Find the relations in terms of connection analysis results of unlabeled data (step 2) and labeled NS5A sequences by appropriate semi-supervised learning methods.*

Fig. 3. Framework for discovering knowledge on NS5A and IFN/RBV therapy

– *Labeled datasets*: One is NS5A sequences of non-response (134) and response patients (93) to IFN/RBV therapy from Los Amalos HCV database [15] and the other is 40 NS5A response and non-response sequences of subtype 1b under preparation by the physicians at Graduate School of Medicine, Chiba University.

The followings are examples of two NS5A V3 regions and two NS5A ISDR regions in HVC sequences of subtypes 1a and 1b, respectively:

STSGITGDNTTTSSEPAPSGCPPD (1a)
ESSAADSGTATAPPDQPSSDGDAG (1b)

PSLKATCTANHDSPDAELIEANLLWRQEMGGNITRVESEN (1a)
PSLKATCTTHHDSPDADLIEANLLWRQEMGGNITRVESEN (1b)

It can be easily observe that while sequences in two subtypes are mostly different at each position in the V3 region, they are almost identical in the ISDR region.

3.3 Encoding the NS5A Sequences

Targeting to fully exploit the NS5A sequences we tend to use different analysis methods which requires different input formats. Generally, two kinds of data formats are needed. First is the original sequences without any processing that is appropriate for sequence analysis methods. Second is transformed sequences by different methods that convert the original sequences into other representation spaces, mostly vector spaces.

Denote by $S = \{S_1, S_2, \ldots, S_n\}$ the set of n sequences. A *pattern* P is a subsequence of one or more sequences in S. Denote by $m(P)$ the set of all sequences in S each contains P as a subsequence. For a set of patterns $PS = \{P_1, P_2, \ldots, P_q\}$, a sequence S_j is matched by PS if it is in $m(P_i)$ for at least one $P_i \in PS$.

The frequent occurrence of the same pattern in biological sequences usually indicates that the sequences biologically related (i.e., they contain similar motifs), and it is assumed that these important regions are better conserved in evolution. These patterns are possibly related to an important function of a set of sequences, and are important factors in sequences classification.

Extracting Different Datasets. In order to analyze and compare the unlabeled data and labeled data as well as response and non-response to IFN/RBV therapy, according to the analysis purpose we extract different datasets for the labeled and unlabeled data, accordingly:

- The ISDR region of 40 amino acids in length,
- The V3 region of 27 amino acids in length,
- The domains I, II and III have 213, 93 and 111 amino acids in length,
- The full NS5A sequences.

k-Gram. The k-grams extraction can be done in a $\mathcal{O}(m \times n \times k)$ time, where m is the maximum length of a sequence, n is the number of sequences and k is the motif length. The k-grams are used in two cases: one is to transform the sequences into vectors of k-grams (with their frequency), and the other is to use k-grams to detect significant patterns in the NS5A sequences.

Active Motifs. This method enables to extract the commonly occurring motifs whose lengths are longer than a specified length, called *active motifs*, in a set of biological sequences. The activity of a motif is the number of sequences which match it within an allowed number of mutations. The motifs extraction is based on the construction of a Generalized Suffix Tree (GST).

Coverage and Discriminant Patterns. Given n sequences each belongs to one of K classes $C_1, C_2, \ldots, C_K$. Denote by $m_{C_i}(P)$ the subset of sequences in class C_i that match P.

Given:

1. *Un-labeled data: NS5A sequences in the V3, ISDR regions and domains I, II, III of HCV subtypes 1ac (1559 sequences) and 1b (1191 sequences).*
2. *Labeled data: NS5A sequences in the V3, ISDR regions and domains I, II, III of HCV subtypes 1ac (87 sequences) and 1b (79 sequences) each is from either IFN/RBV response or non-response patients.*

Find:

1. *α-coverage patterns characterizing subtypes 1ac and 1b, given $\alpha (0 < \alpha < 1)$.*
2. *β-discriminant patterns characterizing response and non-response sequences in subtypes 1ac and 1b, given $\beta (0 < \beta < 1)$.*

Procedure:

For each dataset from the V3, ISDR regions and domains I, II, III (L denotes the length of each region):

1. *For $k = 1, 2, \ldots, L$, find all:*
   - *α-coverage patterns for subtype 1ac.*
   - *α-coverage patterns for subtype 1b.*
2. *For $k = 1, 2, \ldots, L$, find all:*
   - *Shortest β-discriminant patterns for response and non-response sequences in subtype 1ac.*
   - *Shortest β-discriminant patterns for response and non-response sequences in subtype 1b.*

Fig. 4. Algorithm for finding α-coverage and β-discriminant patterns

Definition 1. A pattern P is a *α-coverage* for class C_i if $\frac{|m_{C_i}(P)|}{|C_i|} \geq \alpha$ for a given parameter α $(0 < \alpha < 1)$.

Definition 2. A pattern P is a *β-discriminant* for class C_i if $\frac{|m_{C_i}(P)|}{|m(P)|} \geq \beta$ for a given parameter β $(0 < \beta < 1)$,

Note that the coverage for a class of a pattern measures its frequent occurrence in the region, or in other words, its conservation in evolution. If pattern P_1 is a subsequence of a pattern P_2, then we have $m(P_2) \subseteq m(P_1)$, i.e., the coverage of P_1 is larger and the discrimination ability of P_1 is smaller than those of P_2, respectively. Given an α-coverage pattern P, the most informative pattern related to P in terms of coverage is the longest α-coverage pattern containing P. Alternatively, given a β-discriminant pattern P, the most informative pattern related to P in terms of discrimination is the shortest β-discriminant pattern contained in P.

3.4 Find Interesting Subsequences in NS5A Sequences

Figure 4 describes our algorithm to find interesting subsequences in terms of α-coverage patterns for HCV subtypes 1ac and 1b and β-discriminant patterns

Table 1. α-coverage patterns with high values of α found in V3 and ISDR regions

V3 region							ISDR region						
Type	1ac R	1ac NR	1ac UL	1b R	1b NR	1b UL	Type	1ac R	1ac NR	1ac UL	1b R	1b NR	1b UL
# Seq.	14	73	1559	30	49	1159	# Seq.	14	73	1559	30	49	1159
TGD	14	63	1232	0	0	3	AEL	14	73	1482	1	0	5
STS	14	72	1519	0	0	2	DAE	14	73	1520	1	0	5
TSG	14	73	1532	0	0	4	ELI	14	73	1466	0	0	6
SGIT	10	36	1146	0	0	1	NHD	11	73	1440	0	0	11
GSSS	12	72	1457	0	0	2	PSLK	14	73	1513	20	43	2
SSST	12	72	1465	0	0	3	DSPD	11	73	1476	23	43	2
TSSE	11	40	955	0	0	2	HDSP	11	73	1471	23	42	2
CPPD	10	69	970	0	0	2	IEAN	11	48	1274	21	44	3
TSGI	10	46	1230	0	0	2	ADL	0	0	19	22	42	1111
GTA	0	0	1	23	41	1045	DAD	0	0	17	22	41	1112
PDQ	0	0	1	18	39	917	TRH	0	0	0	17	23	654
PDQ	0	0	1	18	39	917							
SGTA	0	0	1	23	39	1037							
GTAT	0	0	1	22	40	1010							
ATAP	0	0	2	18	39	1060							
TAPP	0	0	2	18	39	1055							
TATA	0	0	1	18	39	992							
SDDG	0	0	0	15	13	992							
DDGD	0	0	0	13	9	992							
DGDA	0	0	0	12	15	992							

for response and non-response sequences in these subtypes. The algorithm is applied to extracted data of V3 and ISDR regions, domains I, II, and III.

Some results are reported and analyzed as below where many motifs found with $\alpha = 0.98$ and $\beta = 0.95$ or $\beta = 0.85$.

α-Coverage Patterns in Subtypes 1ac and 1b. Table 1 shows many 0.98-coverage patterns in subtypes 1ac and 1b found from the V3 and ISDR regions, respectively. In these tables, the cells in the first three rows at columns 2-3 and columns 5-6 indicate the number of response sequences (with label "R") and non response (with label "NR") of subtypes 1ac and 1b, respectively. Those in columns 4 and 7 indicate the number of available sequences without information of IFN/RBV therapy (with "UL" stands for unlabel). Each row from the fourth row shows information of one pattern found. The first column shows the subsequence found. The columns 4 and 6 show the number of sequences in unlabeled data (1ac UL and 1b UL) that match the patterns (key information). The columns 2, 3, and 5, 6 show the number of R and NR sequences in groups 1ac and 1b, respectively. For example, the pattern "AEL" in Table **??** matches 1482 out of 1559 1ac sequences, and 5 out of 1159 1b sequences. Among 1482 1ac sequences that contain "AEL", 14 are response ones and 73 are non response

Table 2. β-discriminant patterns with high value of β in the V3 and ISDR regions

V3 region							ISDR region						
Type	1ac R	1ac NR	1ac UL	1b R	1b NR	1b UL	Type	1ac R	1ac NR	1ac UL	1b R	1b NR	1b UL
# Seq.	14	73	1559	30	49	1159	# Seq.	14	73	1559	30	49	1159
PSG	2	57	455	0	0	5	ANH	1	61	649	0	0	3
GCP	2	63	592	0	0	1	CTA	1	61	673	4	3	47
SGC	2	66	678	0	0	1	TAN	1	61	653	0	0	3
APSG	1	57	392	2	0	2	AAN	0	25	87	0	0	0
NTTA	0	29	69	0	0	0	IAA	0	25	84	0	0	0
EPAS	0	9	300	0	0	0	LIA	0	25	85	0	0	0
PDC	7	0	28	0	0	0	WNQ	0	25	0	0	0	0
DDT	5	0	204	0	0	2	ESES	0	24	6	0	0	0
GDD	5	0	249	0	0	7	DAN	0	0	1	3	0	3
PPDC	7	0	9	0	0	0	LVD	0	0	0	3	0	1
SEPT	4	0	67	0	0	0	VDA	0	0	0	3	0	1
EPTP	4	0	35	0	0	0	CTTH	0	0	10	2	15	317
DQA	0	0	0	2	12	264							
QSA	0	0	0	1	11	277							

ones (all of 87 available labeled 1ac sequences). Among 79 available labeled 1b sequences, only 1 response one contains "AEL".

β-Discriminant Patterns in Subtypes 1ac and 1b. Table 2 shows the 0.95-discriminant patterns in subtypes 1ac and 1b found from the V3 and ISDR regions, respectively. The table cells have the same meaning of those in Table 1. In these tables, the key information is on the columns 2, 3 or 5, 6 depending on the pattern is for 1ac or 1b. For example, in Table 2, the pattern "PSG" discriminates NR sequences from N sequences of 1ac with estimated accuracy ratio is $57/59 = 0.96$ (thus a 0.95- discriminant pattern). It does not match any labeled sequences of subtype 1b. Moreover, the pattern appears in 455 out of 1559 1ac UL sequences, and only in 5 out of 1150 1b UL sequences. The subsequences "PSG", "GCP", "SGC","APSG", "NTTA" and "EPAS" are typical for NR sequences in 1ac, while "PDC", "DDT", "GDD", "GDD", PPDC", "SEPT' and "EPTP" are that typical for R sequences in 1ac. We can find only two 0.85-discriminant patterns "DQA" and "QSA" for NR sequences in 1b.

The last row in Table 2 (in the right part) shows the shortest pattern containing subsequence "TTH" that can distinguish the NR from R sequences with accuracy of 0.88. Note that the ISDR regions of 1ac and 1b are different at only 4 positions, "TAN" in 1ac versus "TTH" in 1b and "E" versus "D" at the position of 17th. The subsequence "TTH" is a 0.98-discriminant pattern for 1ac and also with large coverage (0.71-coverage pattern for NR 1ac sequences) while subsequence "TTH" is the core of a 0.88-discriminant and 0.21-coverage pattern for R 1b sequences).

Table 3. The 0.95-discriminant patterns in the domain I and domains II-III

Domain I							Domains II and III						
Type	1ac R	1ac NR	1ac UL	1b R	1b NR	1b UL	Type	1ac R	1ac NR	1ac UL	1b R	1b NR	1b UL
# Seq.	14	73	1559	30	49	1159	# Seq.	14	73	1559	30	49	1159
EYP	1	58	846	0	0	1	AAN	0	25	86	0	0	0
LHE	1	58	848	0	0	39	LWN	0	25	0	0	0	0
RVD	1	18	20	0	0	3	NQE	0	25	0	0	0	0
CDF	0	20	4	0	0	0	WNQ	0	25	0	0	0	0
MWD	0	14	1	0	0	0	SES	0	24	6	0	0	7
CDF	0	20	4	0	0	0	SKV	0	24	6	0	0	15
APP	0	0	2	0	7	9	KNP	0	0	0	8	0	31
DYA	0	0	0	0	7	7	NPD	0	0	0	8	0	40
KNP	0	0	0	8	0	31	WKN	0	0	0	8	0	31
NPD	0	0	0	8	0	40	GEI	0	0	1	7	0	11
WKN	0	0	0	8	0	31							
SGT	11	59	1241	6	0	8							
WSG	11	59	1247	6	0	3							
DSH	0	0	42	6	0	71							
GDS	0	0	42	6	0	72							
SNM	0	0	2	0	7	213							

It can be observed from these two tables:

- There are more discriminant patterns with larger coverage for NR sequences in 1ac than those for R sequences.
- Some discriminant patterns can be found for R sequences in the V3 regions but with smaller coverage.
- There are less discriminant patterns found in the regions V3 and ISDR for R or NR sequences in the subtype 1b, and they also appear less frequent (small coverage).

When applying the method to datasets extracted from the domains I, II and III we could find a number of 0.95-discriminant patterns for subtype 1ac, and more importantly a number of 0.95-discriminant patterns for subtype 1b as shown in Table 3.

3.5 Linking Subsequences Found and Resistance Mechanisms to IFN/RBV

We have introduced some simple ways of exploiting the large volume of unlabeled data to investigate the labeled sequences. Based on various interesting patterns found in subsection 3.4, the three directions can be pursued:

– Investigate the partnership of detected motifs in NS5A and IFN cellular antiviral pathways (especially JAK-STAT and PKR pathways).
– Investigate the partnership of detected motifs and mutations in NS5A to IFN resistance.
– Develop new semi-supervised learning methods to exploit better unlabeled sequences in linking with labeled data.

3.6 Discussion

The obtained results in this work is primitive but already allow us to have some observations for further investigation.

– It is more difficult to find discriminant patterns for R sequences than NR sequences.
– The V3 region is shorter (24 amino acids) and has more different amino acid pairs than the ISDR region (40 amino acids). It was also suggested that the V3 regions can be a more accurate biomarker [10]. However, the results in Table 2 (right part) provide many significant patterns in the ISDR region.
– Discriminant patterns found in domains II and III show that they can also play a significant role in the relation between NS5A and IFN/RBV therapy.

This work on resistance mechanisms of NS5A to interferon/ribavirin therapy will be combined with other two ones in our project aims at providing molecular basis for reaching higher performance of hepatitis therapy, says, higher than current 50%. The other two include the interplay between epigenetic factors in HCC progression and the RNA interference in silencing hepatitis viruses.

4 Conclusion

We have presented the framework and early results obtained in our project. The target is to detect the relationship between HCV NS5A protein and the interferon/ribavirin therapy effectiveness. These results are promising as they contain many patterns that were not known previously. To our best knowledge, this work is the first one making the link between unlabeled sequences to labeled sequences to solve the problem, and also analyzing the NS5A sequences with advanced data mining methods.

Much work to be continued. One is to complete the work described in the framework and used the discovered patterns to link to the IFN pathways and mutations to make clear the NS5A mechanisms of IFN/RBV resistance. The other is to establish semi-supervised methods that allow exploiting unlabeled data to solve problem often done with label data.

Acknowledgments. This work is partially supported by MEXT project “Advanced computation methods for analyzing scientific data”, JSPS bilateral Japan-Vietnam project on “Computational methods in Biomedicine”, JSPS’s kakenhi project No. 21500135, and NAFOSTED (National Foundation for Science and Technology Development).

References

1. Murray, P.R., Rosenthal, K.S., Pfaller, M.A.: Medical Microbiology, Ch. 66. Mosby Philadelphia (2009)
2. Gao, M., Nettles, R.E., et al.: Chemical genetics strategy identifies an hcv ns5a inhibitor with a potent clinical effect. Nature 465, 953–960 (2010)
3. Lemon, S.M., Keating, J.A., Pietschmannc, T., Frickd, D.N., Glenne, J.S., Tellinghuisenf, T.L., Symonsg, J., Furman, P.A.: Development of novel therapies for hepatitis c. Antiviral Research 86, 79–92 (2010)
4. Jardim, A.C.G., Yamasaki, L.H.T., Queiro, A.T.L., Bittar, C., Rahal, R.P.R., de Carvalho Mello, I.M.V.: Quasispecies of hepatitis c virus genotype 1 and treatment outcome with peginterferon and ribavirin. Infection, Genetics and Evolution 9, 689–698 (2009)
5. Aurora, R., Donlin, M.J., Cannon, N.A.: Genome-wide hepatitis c virus amino acid covariance networks can predict response to antiviral therapy in humans. The Journal of Clinical Investigation 119, 225–236 (2009)
6. Feld, J.J., Hoofnagle, J.H.: Mechanism of action of interferon and ribavirin in treatment of hepatitis c. Nature 436, 967–972 (2005)
7. Enomoto, N., Maekawa, S.: Hcv genetic elements determining the early response to peginterferon and ribavirin therapy. Intervirology 334, 77–81 (2010)
8. Francessco, R., Migliaccio, G.: Challenges and successes in developing new therapies for hepatitis c. Nature 436, 96–100 (2005)
9. Guillou-Guillemette, H.L., Vallet, S., Gaudy-Graffin, C., Payan, C., Pivert, A., Goudeau, A., Lunel-Fabiani, F.: Genetic diversity of the hepatitis c virus: Impact and issues in the antiviral therapy. World Journal of Gastroenterology 13(17), 2416–2426 (2007)
10. AlHefnawi, M.M., Zada, S., El-Azab, I.A.: Prediction of prognostic biomarker for interferon-based therapy to hepatitis c virus patients: a metaanalysis of the ns5a protein in subtypes 1a, 1b, and 3a. Virology Journal 7, 1–9 (2010)
11. Macdonaldt, A., Harris, M.: Hepatitis c virus ns5a: Tales of a promiscuous protein. Journal of General Virology 85, 2485–2502 (2004)
12. Wohnsland, A., Hofmann, W.P., Sarrazin, C.: Viral determinants of resistance to treatment in patients with hepatitis c. Clinical Microbiology Reviews, 20 (2007)
13. Sharma, S.D.: Hepatitis c virus: Molecular biology & current therapetic options. Indian Journal of Medical Research 131, 17–34 (2010)
14. Kamal, S.M.: Acute hepatitis c: a systematic review. The American Journal of Gastroenterology 103(5), 1283–1297 (2008)
15. Kuiken, C., Yusim, K., Boykin, L., Richardson, R.: The los alamos hepatitis c sequence database. Bioinformatics 21(3), 379–384 (2005)

Inclusion-Based and Exclusion-Based Approaches in Graph-Based Multiple News Summarization

Nongnuch Ketui and Thanaruk Theeramunkong

School of Information, Computer and Communication Technology,
Sirindhorn International Institute of Technology, Thammasat University
nongnuch.ketui@studentmail.siit.tu.ac.th,
thanaruk@siit.tu.ac.th
http://www.siit.tu.ac.th

Abstract. As combination of information extraction and relation analysis, constructing a comprehensive summary from multiple documents is a challenging task. Towards summarization of multiple news articles related to a specific event, an ideal summary should include only important common descriptions of these articles, together with some dominant differences among them. This paper presents a graph-based summarization method which is composed of text preprocessing, text-portion segmentation, weight assignment of text portions, and relation analysis among text portions, text-portion graph construction, and significant portion selection. In the process of portion selection, this paper proposes two alternative methods; inclusion-based and exclusion-based approach. To evaluate these approaches, a set of experiments are conducted on fifteen sets of Thai political news articles. Measured with ROUGE-N, the result shows that the inclusion-based approach outperforms the exclusion-based one with approximately 2% performance gap (80.59 to 78.21%).

Keywords: Thai Text Summarization, Multiple News Summarization, Graph-based Summarization.

1 Introduction

Nowadays a gigantic number of news articles are produced in any language all over the world. Not an exception, we can also find a large pile of Thai news articles online. The situation of having too much information make us difficult to find needed information and requests us a long time to read many related news issues before catching the occurring events. Especially it is usual to find news articles relating to an event, from several sources with similar and different facts. To solve this problem, it is necessary to study an efficient and effective method to make a summary from multiple written Thai news articles. The expected summarization gives a short comprehensive description of the similarity and the difference among related news.

In this paper, we present a graph-based summarization method which is composed of text preprocessing, text-portion segmentation, weight assignment of

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 91–102, 2011.

text portions, relation analysis among text portions, text-portion graph construction, and significant portion selection. In the process of portion selection, this paper proposes two alternative methods; inclusion-based and exclusion-based approach. Both approaches are evaluated with a set of Thai political news articles and measured with ROUGE-N. Section 2 describes related works on multi-document summarization and two alternative methods are presented in Section 3. In Section 4, we show experimental settings and results. Finally, conclusion and future work are given in Section 5.

2 Related Works

In past decades, summarization has been recognized as an interesting application in the field of natural language processing (NLP). In an earlier period of summarization research, several works placed interest on scientific documents and proposed various paradigms to extract salient sentences from texts by making use of features like word and phrase frequency [1], position in the text [2] and key phrases [3]. In general, two main summarization approaches are extraction-based and abstraction-based methods. The extraction-based methods [4], [5] and [6] usually involve a process to assign weights for each processing unit, mostly sentence unit or phrase unit. In fact, many approaches differ on the manner of their problem formulation. Various researchers published their work concentrated on newswire data. Many approaches are developed under different problem setting for summarization. One of the important issues in summarization is how to evaluate the obtained summary. As an early stage of multi-document summarization, Radev and his colleagues [7], [8] and [9] proposed centroid-based techniques to generate a composite sentence from each cluster. Radev and Barzilay [7] and that of Barzilay et al. [9] formulated summarization as a clustering problem. To compute a similarity measure between text units, they are mapped to feature vectors that are represented by a set of single words weighted by some weighting system such as TF-IDF. Goldstein and Carbonell [10] combined query relevance with information novelty as the topic and made a major contribution to topic-driven summarization by introducing the maximal marginal relevance (MMR) measure. Mani [11] presented an information extraction framework for summarization as well as a graph-based method to find similarities and dissimilarities in pairs of documents. In this approach, it is possible to specify on maximal number of common and different sentences to control the output. This method uses these structures to actually compose abstractive summaries, rather than to extract sentences from the text.

3 Inclusion- and Exclusion-Based Summarization

This section illustrates four main parts involved in the summarizing process, (1) text preprocessing, (2) node weight assignment, (3) edge weight assignment and (4) candidate paragraphs selecting. Their details are gave in subsections as follow in Fig. 1.

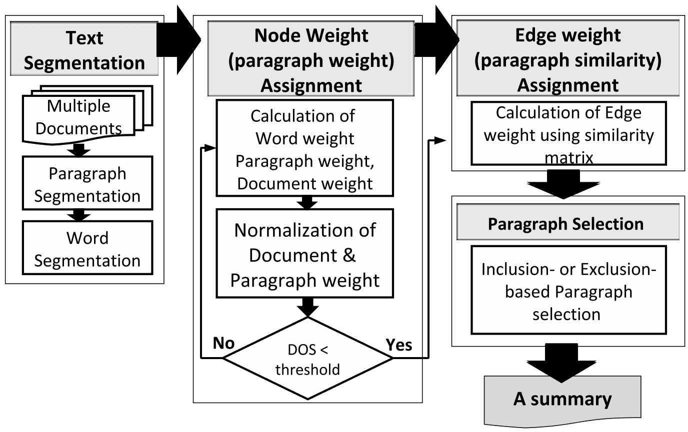


Fig. 1. The process of inclusion- and exclusion-based summarization. Thus, DOS is defined the different of summation of document weight.

3.1 Text Segmentation

As a text segmentation, the document related discovery proposed in [12] is applied to find a set of related news. For this purpose, three main types of news relations are classified based on the relevance, of new events: (1) "completely related" (CR), (2) "somehow related" (SH) and (3) "unrelated" (UR). A CR relation is detected when two new documents mention a same story. For SH relation, it is a kind of relation which has only somewhat closely related. The relation of UR is defined as a relationship of having absolutely unrelated in their events between news documents. In this work, we consider only a CR relation on Thai news articles and solve the problem on multi-document summarization. The problem of Thai language is even worse since the Thai language has no explicit sentence boundary and no clear definition. Therefore, paragraph segmentation only uses the HTML tags <p> (a new paragraph) and
 (line-break) tags. Finally, words are segmented with the maximal matching algorithm [13]. This algorithm first generates all possible segmentations for a sentence and then selects the one that contains the fewest words, which can be done efficiently by using dynamic programming technique.

3.2 Node Weight Assignment

Before assigning the edge weight, the weight of text-portion is measured by a statistical method. This method evaluates the importance of words, paragraphs and documents. Generally, TFIDF (Term Frequency times Inverse Document Frequency) is a well-known as weighting words in document. By this work we slightly modified TFIDF to find important words in a paragraph later called TFIPF (Term Frequency times Inverse Paragraph Frequency) for calculating weights. Formulations of node weight assignment can be summarized as follows.

Let D be a set of news articles where $D = \{d_1, d_2, d_3, ..., d_i\}$, and $CRD = \{d_1, d_2, d_3, ..., d_m\} \subset D$ with m news articles is called a completely related news articles set. Given d_i represents the i-th order document, the document d_i has a set of paragraph p_{ij} where $\{p_{i1}, p_{i2}, p_{i3}, ..., p_{ij}\} \in d_i$ and p_{ij} is a j-th order paragraph in the document d_i. According to the paragraph p_{ij} in the document d_i has a set of word w_{ijk} where $\{w_{ij1}, w_{ij2}, w_{ij3}, ..., w_{ijk}\} \in p_{ij}$. The word weight, $W^{(t)}(w_{ik})$ is used to focus on important words appearing in both documents and paragraphs. Let t be the member of iterations and w_{ik} represents word w_k in the document d_i. Before calculating the weight of a particular word w_k, the initial word weight and paragraph weight is assigned as $W^{(0)}(w_{ik}) = \frac{N(w_{ik})}{N(w_i)}$ and $W^{(0)}(p_{ij}) = \frac{1}{P_i}$ respectively. Let $N(w_{ik})$ is the total number of a word w_k that occurs in the document d_i and $N(w_i)$ is the total number of all words in the document d_i. While the total number of paragraphs appears in the document d_i is given to P_i.

$$W_w^{(t+1)}(w_{ik}) = N(d_i) \times W^{(t)}(w_{ik}). \tag{1}$$

The frequency of a particular word w_k appears in the document d_i as $W_w^{(t+1)}(w_{ik})$. Given $N(d_i)$ is the total number of a particular word appearing in the i-th document. While the weight of a particular word w_k appears in the paragraph p_j in the document d_i as $W_p^{(t+1)}(w_{ik})$ that will be assigned as

$$W_p^{(t+1)}(w_{ik}) = P(d_i) \times \sum_{p_{ij} \in d_i} N(w_{ijk}) \times W^{(t)}(p_{ij}), \tag{2}$$

where $P(d_i)$ is the total number of paragraphs appears in i-th order document, $N(w_{ijk})$ is the total number of a particular word w_k appearing in the paragraph p_j of the document d_i.

$$TF^{(t+1)}(w_{ik}) = \sqrt{W_w^{(t+1)}(w_{ik}) \times W_p^{(t+1)}(w_{ik})}, \tag{3}$$

where $TF^{(t+1)}(w_{ik})$ is the expected frequency of a particular word w_k appearing in both of the document and paragraph as Equation 1 and 2. Let $IDF^{(t+1)}(w_k)$ be inversed document frequency represented by the following formula:

$$IDF^{(t+1)}(w_k) = \log\left(1 + \frac{\sum_{d \in D} W^{(t)}(d)}{\sum_{d_i \in D} W^{(t)}(d_i)}\right), \tag{4}$$

such that $w_k \in d_i$. Given $IDF^{(t+1)}(w_k)$ represents by the summation of document weight $W^{(t)}(d)$ in corpus divided by the expected summation of the document weight where word w_k occurs in the document d_i and in corpus D. Here, the initial document weight $W^{(0)}(d)$ is assigned as the total of documents in the corpus where $W^{(0)}(d) = \frac{1}{|D|}$. Then the word weight will be assigned with Equation 3 and 4 using the following formula:

$$W^{(t+1)}(w_{ik}) = \frac{TF^{(t+1)}(w_{ik})}{\sum_{w \in d_i} TF^{(t+1)}(w)} \times IDF^{t+1}(w_k). \tag{5}$$

The word weight $W^{(t+1)}(w_{ik})$ has the value between 0 and IDF. After assigning a word weight in the document d_i, an inverse paragraph frequency IPF is calculated as follows.

$$IPF^{(t)}(w_{ik}) = \log\left(1 + \frac{\sum_{p \in d_i} W^{(t)}(p)}{\sum_{p_j \in d_i} W^{(t)}(p_j)}\right), \tag{6}$$

such that $w_{ik} \in p_j$, where $IPF^{(t)}(w_{ik})$ is assigned with the summation of the paragraph weight $W^{(t)}(p)$ in the document d_i and the summation of the paragraph weight $W^{(t)}(p_j)$ that has a word w_i occurring in the paragraph p_j and in the document d_i. Then the paragraph weight is calculated by using the following formula:

$$W^{(t)}(p_{ij}) = \sum_{w_{ik} \in p_j} W^{(t)}(w_{ik}) \times IPF^{(t)}(w_{ik}). \tag{7}$$

All words in this paragraph p_{ij} are summed up. The paragraph weight $W^{(t)}(p_{ij})$ is calculated by the summation of the multiplication of the word weight in this paragraph $W^{(t)}(w_{ik})$ and an inverse paragraph frequency $IPF^{(t)}(w_{ik})$ using Equations 5 and 6. Before calculating a document weight, we need to find the location weight W_{loc} by linear formula, $W_{loc}(p_{ij}) = -0.1 \times j + 1.1$;. Let j is the j-th paragraph. This formula is firstly generated on the random news articles and weighed by human, and then plotted graph. The location of paragraph indicates the important information in the new articles.

$$W^{(t)}(d_i) = \sum_{p_{ij} \in d_i} W^{(t)}(p_{ij}) \times W_{loc}(p_{ij}). \tag{8}$$

All paragraphs in the document d_i are summed up. The document weight $W^{(t)}(d_i)$ is calculated by the multiplication of the paragraph weight $W^{(t)}(p_{ij})$ in the paragraph p_j and the location weight $W_{loc}(p_{ij})$ that paragraph p_j occurs.

Finally, the document and paragraph weight is normalized by the two following formulas:

$$W^{(t+1)}(d_i) = \frac{W^{(t)}(d_i)}{\sum_{d \in D} W^{(t)}(d)}, \tag{9}$$

$$W^{(t+1)}(p_{ij}) = \frac{W^{(t)}(p_{ij})}{\sum_{p \in d_i} W^{(t)}(p)}. \tag{10}$$

The node weight assignment will be iterated until the difference of summation of the document weight (DOS) is less than threshold 0.5.

3.3 Edge Weight Assignment

Conceptually selecting candidate paragraph with Graph-based summarization in multiple news articles from a similarity relation graph, where a node corresponds to a paragraph in sets of news articles and an edge corresponds to an indirect

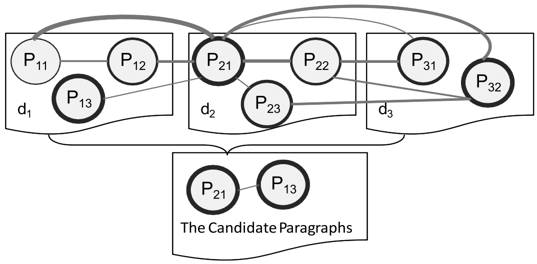


Fig. 2. An example of a similarity relation graph

relation of a paragraph to another paragraph. The edge weights display the similarity between two paragraphs in a set of completely related news articles. They are calculated by cosine similarity as the following formula:

$$sim(p_{ij}, q_{ij}) = \frac{\sum_{k=1}^{n} wp_{ijk} \times wq_{ijk}}{\sqrt{\sum_{k=1}^{n} wp_{ijk}^2}\sqrt{\sum_{k=1}^{n} wq_{ijk}^2}}, \tag{11}$$

where p_{ij} and q_{ij} are vectors of $< wp_{ij1}, wp_{ij2}, ..., wp_{ijk} >$ and $< wq_{ij1}, wq_{ij2}, ..., wq_{ijk} >$, respectively. Given wp_{ijk} and wq_{ijk} are the word weights of a word w_k occurring in the paragraph p_j in the document d_i. Then a term by paragraph matrix is generated with rows of the word weights and columns of the paragraphs.

3.4 Inclusion-Based Approach

The inclusion-based approach starts from selecting the most important text portion and then add the second most important text portion the semantics of which is not overlapped with the first selected text portion. By this approach selects the candidate paragraph from multiple news articles. Algorithm 1 shows a pseudo code of inclusion-based approach to analyze the similarity between paragraph and another paragraph. A set of node weight(paragraph weight) is given as S, and then we set the average paragraph for selecting candidate paragraphs. The average paragraph in each CR news is calculated by the ratio of number of paragraph and number of document and then multiply with 0.5. To initialize the node with a maximal weight w_{max} from the set S. We consider the neighbouring nodes which connect with the maximal node, and then find the maximum ratio of the neighbouring node weight and both of the included and non-included edge weight. Before this neighbouring node is added into the inclusion graph G, it is assigned to the maximal node. Then we iterate to consider the neighbouring nodes again until the total of nodes is equal the average of a paragraphs P_{avg}. Finally, the inclusion graph is returned to the candidate summaries.

Algorithm 1. Inclusion-based approach

Input: Number of paragraphs n_p, documents in related news n_d,
Node Weight w and Edge Weight e
Output:The candidate paragraphs in Graph G
create Graph S with all node weight w and connect with edge weight e
calculate the average of paragraphs per related news article $P_{avg} = \frac{n_p}{n_d} \times 0.5$
set the initial node w_{max} to the maximum node weight in S
do
  insert(G, w_{max})
  delete(S, w_{max})
  foreach all nodes $w_i \in S$ **do**
    foreach all nodes $w_j \in S$ are connected with w_{max} **do**
      calculate the summation of edge weights $e_{i,j}$
    calculate the average of edge weight $avg(e_{i,j})$
    foreach all nodes $w_k \in G$ are connected with w_{max} **do**
      calculate the summation of edge weights $e_{i,k}$
    calculate the average of edge weight $avg(e_{i,k})$
    $w_{inew} = \frac{w_i \times avg(e_{i,j})}{avg(e_{i,k})}$
    if w_{inew} is the maximum weight **then**
      $w_{max} = w_i$
until the total of nodes is equal P_{avg}
return G

3.5 Exclusion-Based Approach

Exclusion-based approach finds the maximal of common paragraphs content in the related news articles and omits the maximal of similarity between two paragraphs. The summary from multiple news articles can be constructed when we omit useless paragraphs. Algorithm 2 shows a pseudo code of exclusion-based approach. A set of node weight(paragraph weight) is given as G, and then we set the average paragraph for selecting candidate paragraphs. To initialize the node with a minimal weight from the set G. We consider the neighbouring nodes which connect with the minimal node, and then find the maximum ratio of the neighbouring node weight and both of the excluded and non-excluded edge weight. Before the edge of this neighbouring node is omitted from the exclusion graph G, the neighbouring node is assigned to the minimal node. Then we iterate to consider the neighbouring nodes again until the rest total of nodes is less than or equal the average of a paragraphs P_{avg}. Finally, the rest nodes in exclusion graph are returned to the candidate summaries.

4 Experiment

4.1 Experimental Setting and Evaluation Criteria

For significant portion selection, we investigate and compare the performance of both approaches when the selected candidate paragraphs are considered as the

Algorithm 2. Exclusion-based approach

Input: Number of paragraphs n_p, documents in related news n_d,
Node Weight w, Edge Weight e and Empty Graph S
Output:The candidate paragraphs in Graph G
create Graph G with all node weight w and connect with edge weight e
calculate the average of paragraphs per related news article $P_{avg} = \frac{n_p}{n_d} \times 0.5$
set the initial node w_{min} to the minimum node weight in G
do
  delete(G, w_{min})
  insert(S, w_{min})
  foreach all nodes $w_i \in G$ **do**
    foreach all nodes $w_j \in G$ are connected with w_{min} **do**
      calculate the summation of edge weights $e_{i,j}$
    calculate the average of edge weight $avg(e_{i,j})$
    foreach all nodes $w_k \in S$ are connected with w_{min} **do**
      calculate the summation of edge weights $e_{i,k}$
    calculate the average of edge weight $avg(e_{i,k})$
    $w_{inew} = \frac{e_{i,j}}{w_i \times e_{i,k}}$
    if w_{inew} is the maximum weight **then**
      $w_{min} = w_i$
until the rest nodes is equal P_{avg}
return G

correct answers. Precision, recall and F-measure are used in the evaluation are as follows:

$$Precision = \frac{|P_{ref} \cap P_{sys}|}{P_{ref}}. \tag{12}$$

$$Recall = \frac{|P_{ref} \cap P_{sys}|}{P_{sys}}. \tag{13}$$

$$F - measure = \frac{(\alpha + 1) \times recall \times precision}{recall + (\alpha \times precision)}. \tag{14}$$

where P_{ref} and P_{sys} denote the number of paragraphs appeared in the reference summary and in the system summary, respectively. For F-measure, the experiments use F1 (i.e., the value of α is 1).

In addition, Recall-Oriented Understudy for Gisting Evaluation (ROUGE) [14] is used for the evaluation. ROUGE calculation is based on various statistical metrics by counting overlapping units such as n-grams, word sequences, and word pairs between systems which generate summaries correlating with those extracted by human evaluations. ROUGE-N is included in the summarization evaluation package, distributed by National Institute of Standards and Technology (NIST). ROUGE-N is an N-gram recall between an automatic summary and a set of manual summaries, which is calculated as:

Table 1. The output paragraphs selection from human and system summaries

Set of CR News	Number of paragraph in document	Number of word in paragraph	Number of output paragraphs	Reference summaries	System summaries	
					Inc.-based approach	Exc.-based approach
1	3(3,2,3)	1(61,85,54) 2(132,85,54) 3(27,118,54)	2	(2-1, 2-2)	(2-1, 2-2)	(2-1, 3-2)
2	3(6,9,8)	4(67,167,187,160,138,105) 5(30,127,37,113, 70,126,29,146) 6(11,73,180,115,70, 160,144,104)	4	(6-3, 5-5, 4-3, 4-4)	(4-3, 6-3, 4-2, 4-4)	(4-3, 4-2, 6-3, 5-5)
3	3(7,4,5)	7(44,56,95,99,100,92,57) 8(218,149,59,62) 9(9,65,204,148,79)	3	(9-3, 8-2, 9-5)	(8-1, 9-4, 8-2)	(8-1, 8-2, 9-3)
4	3(4,5,2)	10(11,10,175,36) 11(25,50,75,72,45) 12(114,59)	2	(12-1, 12-2)	(12-1, 10-3)	(12-1, 10-3)
5	3(7,4,5)	13(44,56,95,99,100,91,57) 14(218,149,59,61) 15(9,65,204,148,79)	3	(15-3, 14-2, 15-5)	(14-1, 15-4, 14-2)	(14-1, 14-2, 15-3)
6	3(4,3,4)	16(12,22,141,37) 17(38,123,42) 18(30,95,47,39)	2	(16-3, 17-3)	(17-2, 16-3)	(17-2, 16-3)
7	3(8,18,14)	19(7,51,80,201,222 ,157,133,122) 20(37,56,15,72,32,44,57, 50,66,66,38,66,81,42, 44,46,56,65) 21(117,91,134,11,78,69, 58,122,75,134,12,86,95,33)	7	(19-4, 19-5, 19-6, 19-7, 19-8, 21-1, 19-2)	(19-4, 21-1, 19-5,19-8, 19-6,21-10, 21-3)	(21-1, 19-4, 19-5, 19-8, 19-6, 21-3, 19-3)
8	3(7,7,10)	22(11,150,59,312, 95,162,117) 23(72,143,32,229, 118,159,203) 24(40,88,84,85,101,103, 161,67,104,49)	4	(22-4, 23-5, 23-6, 23-7)	(22-4, 23-7, 22-2, 23-4)	(22-4, 23-7, 22-2, 23-4)
9	3(6,10,10)	25(34,94,95,79,156,86) 26(34,70,46,78,61, 47,45,83,73,35) 27(26,70,46,78,60, 47,45,83,73,35)	4	(25-2, 25-3, 25-5, 25-6)	(25-5, 26-6, 27-9, 25-2)	(25-5, 27-9, 25-2, 26-6)
10	3(5,4,4)	28(27,100,20,39,88) 29(12,133,39,88) 30(5,47,142,42)	2	(30-3, 28-5)	(30-3, 29-4)	(30-3, 29-2)
11	4(3,5,4,4)	31(144,68,78) 32(27,103,71,54,28) 33(44,128,71,53) 34(56,118,71,54)	3	(31-1, 31-2, 31-3)	(31-1, 32-3, 34-2)	(33-2, 31-1, 32-3)
12	3(5,6,7)	35(57,208,179,101,78) 36(9,36,146,197,98,88) 37(14,5,210,85, 101,129,115)	3	(37-3, 35-3, 35-4)	(35-2, 36-4, 37-3)	(36-4, 35-3, 37-3)
13	4(4,3,4,3)	38(29,76,67,57) 39(104,69,83) 40(27,72,63,92) 41(90,62,98)	2	(39-1, 40-3)	(39-3, 41-1)	(41-1, 41-3)
14	3(4,2,4)	42(29,64,52,63) 43(150,113) 44(29,64,52,63)	2	(43-1, 43-2)	(43-1, 42-4)	(43-1, 43-2)
15	3(4,3,2)	45(14,49,158,88) 46(36,160,87) 47(161,86)	2	(47-1, 45-4)	(47-1, 46-2)	(47-1, 46-2)

$$ROUGE-N=\frac{\sum\limits_{S\in\{ReferenceSummaries\}}\ \sum\limits_{gram_n\in S} Count_{match}\left(gram_n\right)}{\sum\limits_{S\in\{ReferenceSummaries\}}\ \sum\limits_{gram_n\in S} Count\left(gram_n\right)}. \quad (15)$$

4.2 Experimental Results

The experiment is to compare two alternative approaches; inclusion-based and exclusion-based approaches. Table 1 showed the selected paragraphs from manual and both approaches summaries. The detail of sets of CR news are also displayed the number of documents, paragraphs and words. For example, the first set of CR news has three documents; the document no.1 contains three paragraphs and each of them consists of 61,85 and 54 words, respectively.

From Table 1, the number of words in paragraph of the inclusion-based approach was less than the exclusion-based approach in five sets and greater than the exclusion-based approach in four sets. The number of output paragraphs are calculated from the average of paragraphs in related news and multiply with 0.5. Our approaches have to find the different and common of content, so that the output summaries are selected at least two paragraphs.

In addition, Table 2 showed the comparison of f-measure of both approaches, both approaches could achieve the similar performance in eights sets of CR news. The inclusion-based approach performed better than the exclusion-based

Table 2. F-measure Comparison between inclusion-based and exclusion-based approaches

Set of CR News	The number of output paragraphs	Inc.-based approach			Exc.-based approach		
		P(%)	R(%)	F(%)	P(%)	R(%)	F(%)
1	2	100.00	100.00	100.00	63.31	82.20	71.53
2	4	89.38	84.52	86.88	89.38	84.52	86.88
3	3	74.90	88.51	81.14	67.95	89.20	77.14
4	2	60.33	83.91	70.19	60.33	83.91	70.19
5	3	74.90	88.51	81.14	67.95	89.20	77.14
6	2	64.34	89.73	74.94	64.34	89.73	74.94
7	7	90.43	97.33	93.75	90.12	90.30	90.21
8	4	77.74	85.46	81.42	77.74	85.46	81.42
9	4	96.71	86.64	91.40	96.71	86.64	91.40
10	2	76.49	93.85	84.28	57.14	76.92	65.57
11	3	71.91	73.38	72.64	70.59	73.72	72.12
12	3	59.09	81.29	68.43	59.09	81.29	68.43
13	2	75.19	79.05	77.07	75.19	79.05	77.07
14	2	89.80	66.42	76.36	100.00	100.00	100.00
15	2	62.87	76.89	69.18	62.87	76.89	69.18
Average		77.60	85.03	80.59	73.51	84.60	78.21

approach in six sets while the exclusion-based approach was good measures in only one set. In the inclusion-based approach, we found that precision measure achieve 77.60% and recall measure achieve up 85.03%. While the exclusion-based approach has 73.51% in precision measure and 84.60% in recall measure. The inclusion-based approach can achieve up to 80.59% and the exclusion-based approach can achieve up to 78.21%. As a conclusion, the performance of the inclusion-based approach is better than the exclusion-based approach.

5 Conclusion and Future Work

This paper presented a graph-based summarization method to use the sets of Thai political news articles as resources for extracting the candidate summary. Four main steps of summarization in this work are as follows. First, preprocessing is performed to extract the sets of news article relations and remove all unnecessary significant symbols. Then, text-portion segmentation is used to separate an original article into paragraph and word unit as sources for assigning node weight. Second, node weight assignment is used to find an appropriate weight of a word, a paragraph and a document. Third, edge weight assignment find similarity between two paragraphs represented an edge of nodes. Finally, in the process of portion selection, we use two alternative methods; inclusion-based approach and exclusion-based approach. In the first approach, the most important contents are added. We considers the differential contents as finding the maximal of average weight between node weight and edge weight, and including the summary. In the second approach, exclusion-based approach focuses on the common contents as finding the minimal of average weight between node weight and edge weight, and excluding the summary. By comparing the result of inclusion-based and exclusion-based approach, the evaluation use ROUGE-N [14] compare with manual evaluation Using fifteen sets of Thai political news articles obtained from an allnews database and classified in completely related type of news relations based on the relevance of news events as a process to discover news article relation [12], two proposed methods were shown to be a powerful way to select the candidate paragraphs and comparison to human judgement. By experiment, the result shows that the inclusion-based approach outperforms the exclusion-based one with approximately 2% performance gap (80.59 to 78.21%). As a future work, we plan to improve the inclusion- and exclusion-based approach by iteration of node and edge weight after insertion or deletion nodes. Towards this work, we need to consider the performance of selecting the summary between two approaches in multiple news articles.

Acknowledgments. This work was supported by the National Research University Project of Thailand Office of Higher Education Commission, as well as the National Electronics and Computer Technology Center (NECTEC) under Project Number NT-B-22-KE-38-54-01. We would like to thank to all members at KINDML laboratory at Sirindhorn International Institute of Technology for fruitful discussion and comments, and Dr.Suriyawut Ketui on his kind helps in human evaluation.

References

1. Luhn, H.P.: The automatic creation of literature abstracts. IBM J. Res. Dev. 2(2), 159–165 (1958)
2. Baxendale, P.B.: Machine-made index for technical literature: an experiment. IBM J. Res. Dev. 2(4), 354–361 (1958)
3. Edmundson, H.P.: New methods in automatic extracting. J. ACM 16(2), 264–285 (1969)
4. Goldstein, J., Kantrowitz, M., Mittal, V., Carbonell, J.: Summarizing text documents: Sentence selection and evaluation metrics. In: Research and Development in Information Retrieval, pp. 121–128 (1999)
5. Mani, I., Bloedorn, E.: Summarizing similarities and differences among related documents. Information Retrieval 1, 35–67 (1999), 10.1023/A:1009930203452
6. Radev, D.R., Hovy, E., McKeown, K.: Introduction to the special issue on summarization (2002)
7. Radev, D.R., Jing, H., Budzikowska, M.: Centroid-based summarization of multiple documents: Sentence extraction, utility-based evaluation, and user studies (2000)
8. McKeown, K., Klavans, J., Hatzivassiloglou, V., Barzilay, R., Eskin, E.: Towards multidocument summarization by reformulation: Progress and prospects. In: AAAI/IAAI, pp. 453–460 (1999)
9. Barzilay, R., McKeown, K.R., Elhadad, M.: Information fusion in the context of multi-document summarization (1999)
10. Goldstein, J., Carbonell, J.: Summarization (1) using mmr for diversity - based reranking and (2) evaluating summaries. In: Proceedings of a Workshop on Held at Baltimore, Maryland, pp. 181–195. Association for Computational Linguistics, Morristown (1996)
11. Mani, I.: Multi-document summarization by graph search and matching. In: Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI 1997), pp. 622–628. AAAI (1997)
12. Kittiphattanabawon, N., Theeramunkong, T., Nantajeewarawat, E.: Exploration of document relation quality with consideration of term representation basis, term weighting and association measure. In: Chen, H., Chau, M., Li, S.-h., Urs, S., Srinivasa, S., Wang, G.A. (eds.) PAISI 2010. LNCS, vol. 6122, pp. 126–139. Springer, Heidelberg (2010)
13. Meknavin, S., Charoenpornsawat, P., Kijsirikul, B.: Feature-based Thai word segmentation. In: Proceedings of the Natural Language Processing Pacific Rim Symposium 1997 (1997)
14. Lin, C.-Y.: Rouge: A package for automatic evaluation of summaries. In: Proc. ACL Workshop on Text Summarization Branches Out, p. 10 (2004)

Band Selection for Hyperspectral Image Using Principal Components Analysis and Maxima-Minima Functional

Kitti Koonsanit and Chuleerat Jaruskulchai

Department of Computer Science, Faculty of Science, Kasetsart University,
Bangkok, Thailand
{g5317400074,fscichj}@ku.ac.th

Abstract. Nowadays, hyperspectral image software becomes widely used. Although hyperspectral images provide abundant information about bands, their high dimensionality also substantially increases the computational burden. An important task in hyperspectral data processing is to reduce the redundancy of the spectral and spatial information without losing any valuable details. In this paper, we present band selection technical using principal components analysis (PCA) and maxima-minima functional for hyperspectral image such as small multi-mission satellite (SMMS). Band selection method in our research not only serves as the first step of hyperspectral data processing that leads to a significant reduction of computational complexity, but also a invaluable research tool to identify optimal spectral for different satellite applications. In this paper, an integrated PCA and maxima-minima functional method is proposed for hyperspectral band selection. Based on tests in a SMMS hyperspectral image, this new method achieves good result in terms of robust clustering.

Keywords: Band Selection, Principal Components Analysis, PCA, Satellite image, Maxima-Minima Functional.

1 Introduction

Satellite application in which such fast processing is needed is the dimension reduction of hyperspectral remote sensing data. Dimension reduction can be seen as a transformation from a high order dimension to a low order dimension. Principle Component Analysis (PCA) is perhaps the most popular dimension reduction technique for remotely sensed data [1]. The growth in data volumes due to the large increase in the spectral bands and high computational demands of PCA has prompted the need to develop a fast and efficient algorithm for PCA. In this work, we present on an implementation of maxima-minima functional with PCA dimension reduction of hyperspectral data. For this paper, hyperspectral data was obtained from the Small Multi-Mission Satellite (SMMS) [2] which has a ground pixel size of 100m x 100m and a spectral resolution of 115 channels, covering the range from 450 nm to 950 nm.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 103–112, 2011.

We focus on a collection of data taken in June 12, 2010 in the northern part of Amnat Charoen province, Thailand. The data consists of 200 x 200 pixels by 115 bands of a total size of 8.86 Mbytes. Figure 1 shows the color composites image.

Fig. 1. The Color Composite Image One

In this paper, the research interest is focused on comparing the effects of integrated PCA and maxima-minima functional method (PCA-MM) of band selection on the final clustering results for hyperspectral imaging application. In the following sections, the various methods in this research, including PCA, maxima-minima functional, and the proposed integrated PCA-MM, are described, and the experimental results are presented later.

2 Methods

2.1 Dimensionality Reduction

Nowadays, hyperspectral image software becomes widely used. Although hyperspectral images provide abundant information about bands, their high dimensionality also substantially increases the computational burden. An important task in hyperspectral data processing is to reduce the redundancy of the spectral and spatial information without losing any valuable details. Therefore, these conventional methods may require a pre-processing step, namely dimension reduction. Dimension reduction can be seen as a transformation from a high order dimension to a low order which eliminates data redundancy. Principal Component Analysis (PCA) [1], [3], [9], [10] is one such data reduction technique, which is often used when analyzing remotely sensed data. The collected hyperspectral image data are in the form of three dimensional image cube, with two spatial dimensions (horizontal and vertical) and one spectral dimension (from SMMS spectrum 1 to spectrum 115 in this study). In order to reduce the dimensionality and make it convenient for the subsequent processing steps, the easiest way is to reduce the dimensions by PCA.

2.2 Background on PCA

PCA is a widely used dimension reduction technique in data analysis. It is the optimal linear scheme for reducing a set of high dimensional vectors into a set of lower dimensional vectors. There are two types of methods for performing PCA, the matrix method, and the data method. In this work, we will focus on the matrix methods. To compute PCA, we follow the general 4 steps given below [1]:

I. Find mean vector in x-space
II. Assemble covariance matrix in x-space
III. Compute eigenvalues and corresponding eigenvectors
IV. Form the components in y-space

It has been previously shown that only the first few components are likely to contain the needed information [4]. The number of components that hold the majority of the information is called the intrinsic dimensionality and each data image may have a different intrinsic dimensionality. PCA condenses all the information of an "N" band original data set into a smaller number than "N" of new bands (or principal components) in such a way that maximizes the covariance and reduces redundancy in order to achieve lower dimensionality as shown in the figure 2.

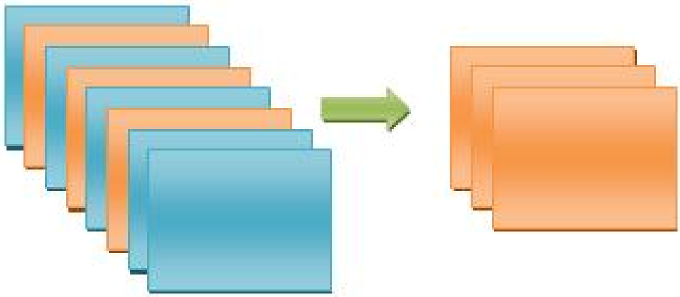


Fig. 2. PCA is a dimension reduction technique

2.3 Maxima and Minima

One more technique was integrated in our research, maxima and minima[5], known collectively as extrema, are the largest value (maximum) or smallest value (minimum), that a function takes in a point either within a given neighborhood (local extremum) or on the function domain in its entirety (global or absolute extremum). More generally, the maxima and minima of a set (as defined in set theory) are the greatest and least values in the set as shown in the figure 3. To locate extreme values is the basic objective of optimization.

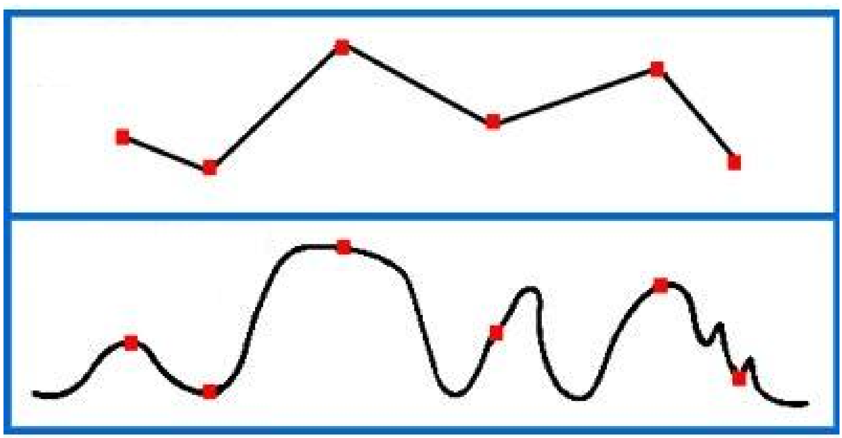


Fig. 3. Maxima and minima

Main features that classify the different objects should be extracted and preserved. The optimal bands is accordingly defined as the bands that not only maintains the major representation of the original data cube, but also maximally preserves features that separate different object classes [6]. Since the PCA method does not necessarily guarantee that the resulting transformation will preserve the classification information among different object classes, a maxima and minima method is proposed in this study to achieve a better performance to satisfy the optimal band selection criteria. We think that maxima and minima value can use preserved features that separate different object classes. Maxima and minima value can calculate and explain as shown in session 2.3.1-2.3.4.

2.3.1 Theorem on First Derivative Test

Let f (x) be a real valued differentiable function. Let be a point on an interval I such that f '(a) = 0.

(a) a is a local maxima of the function f (x) if
i) f (a) = 0
ii) f(x) changes sign from positive to negative as x increases through a.
That is, f (x) > 0 for x < a and f (x) < 0 for x > a
(b) a is a point of local minima of the function f (x) if
i) f (a) = 0
ii) f(x) changes sign from negative to positive as x increases through a.
That is, f (x) < 0 for x < a, f (x) > 0 for x > a

2.3.2 Working Rule for Finding Extremum Values Using First Derivative Test

Let f (x) be the real valued differentiable function.

Step 1: Find f '(x)

Step 2: Solve f '(x) = 0 to get the critical values for f (x). Let these values be a, b, c. These are the points of maxima or minima.

Arrange these values in ascending order.

Step 3:

Check the sign of f'(x) in the immediate neighbourhood of each critical value.

Step 4:

Let us take the critical value x= a. Find the sign of f '(x) for values of x slightly less than a and for values slightly greater than a.

(i) If the sign of f '(x) changes from positive to negative as x increases through a, then f (a) is a local maximum value.
(ii) If the sign of f '(x) changes from negative to positive as x increases through a, then f (a) is local minimum value.
(iii) If the sign of f (x) does not change as x increases through a, then f (a) is neither a local maximum value not a minimum value. In this case x = a is called a point of inflection.

2.3.3 Theorem on Second Derivative Test

Let f be a differentiable function on an interval I and let a is a member of I. Let f "(a) be continuous at a. Then

i) 'a' is a point of local maxima if f '(a) = 0 and f "(a) < 0
ii) 'a' is a point of local minima if f '(a) = 0 and f "(a) > 0
iii) The test fails if f '(a) = 0 and f "(a) = 0. In this case we have to go back to the first derivative test to find whether 'a' is a point of maxima, minima or a point of inflexion.

2.3.4 Working Rule to Determine the Local Extremum Using Second Derivative Test

Step 1

For a differentiable function f (x), find f '(x). Equate it to zero. Solve the equation f '(x) = 0 to get the Critical values of f (x).

Step 2

For a particular Critical value x = a, find f "'(a)

(i) If f "(a) < 0 then f (x) has a local maxima at x = a and f (a) is the maximum value.
(ii) If f "(a) > 0 then f (x) has a local minima at x = a and f (a) is the minimum value.
(iii) If f "(a) = 0 or infinity, the test fails and the first derivative test has to be applied to study the nature of f(a).

2.4 PCA-MM

At the band selection stage, several projection-based methods are studied, including principle components analysis (PCA), maxima and minima function (MM), and integrated PCA-MM methods. We integrated PCA- MM method which is follow as

$$\text{X Band Selected} = \text{PCA of Band} \cap \text{MM of Band} \quad (1)$$

Let we take two sets PCA of band and MM of band, (∩ = intersection) PCA of band ∩ MM of band, is the set of band selected such that the behind statement is true for the entire element x: X is a band member of band selected if one means if X is a band member of PCA and X is a band member of MM.

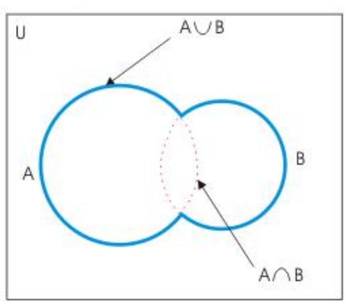


Fig. 4. The intersection set consists of elements common to two sets is a PCA-MM

This paper presents a PCA-MM method, which can effectively reduce the hyperspectral data to intrinsic dimensionality as shown in the figure 5.

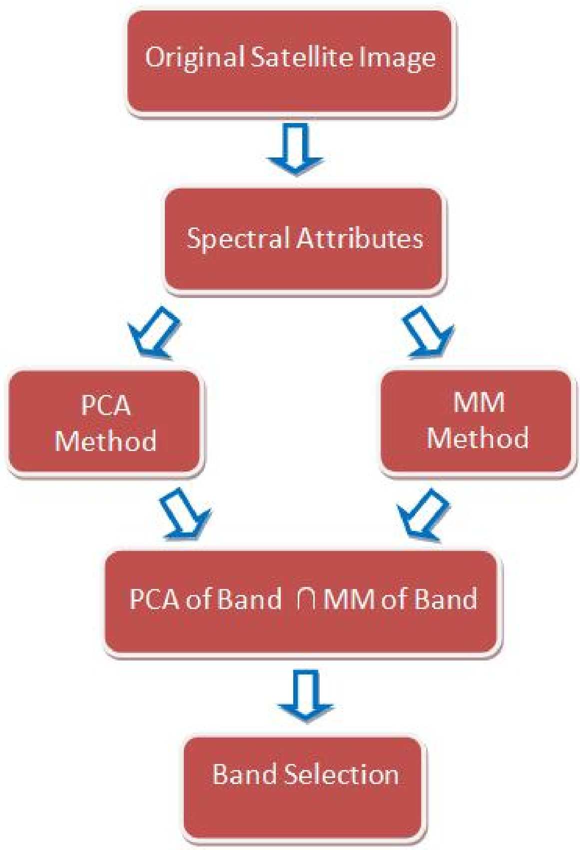


Fig. 5. Overall of our band selection technical

In the PCA-MM, we divide step into 2 parts and combine the results by intersection. In the following sections, the clustering results over the original and the resulting reduced data have been compared.

3 Result

In this section, we present and compare the experimental results obtained by applying each of the techniques, and then evaluating its effectiveness in clustering.

3.1 Experimental Setup

3.1.1 Hyperspectral Data

For this paper, hyperspectral test data was obtained from the SMMS imaging. For experiments we focus on a collection of data taken in June 12, 2010 in the northern part of Amnat charoen province, Thailand. The data consists of 200 x 200 pixels by 115 bands.

3.1.2 Unsupervised Classification Method

The experiments performed in this paper use the simple K-mean from the WEKA software package [7], [8]. The simple K-Mean is the common unsupervised classification method used with remote sensing data. The effectiveness of the K-Mean depends on reasonably accurate estimation of the k cluster for each spectral class.

3.2 Experimental Results

In this paper, the PCA-MM is performed using the two techniques described above. While performing the PCA, we have performed 3 experiments as shown in the figure 6.

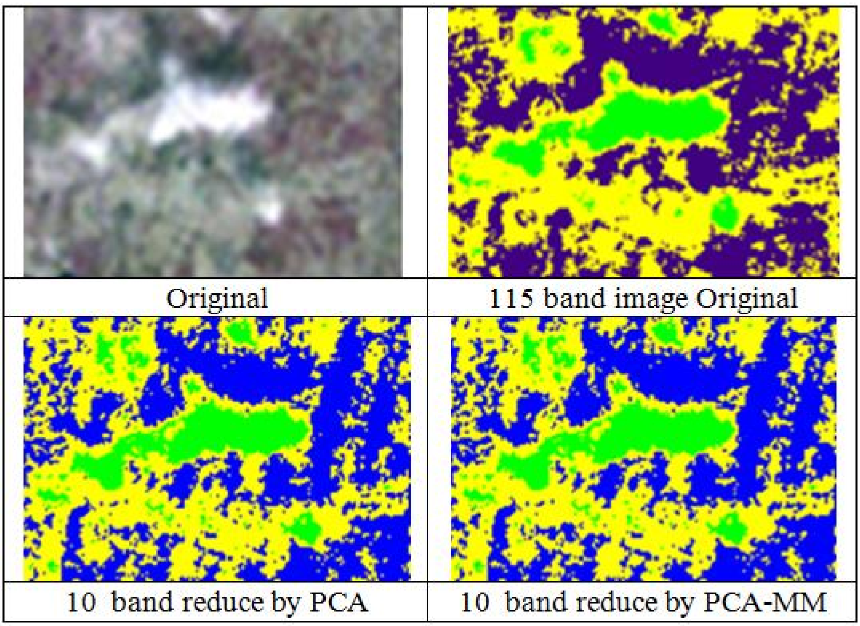


Fig. 6. Result images

Table 1. Show comparison PCA, PCA-MM, Original image by Simple K-Mean clustering

	4 Cluster 115 band Original image	4 Cluster 10 band reduce by PCA	4 Cluster 10 band reduce by PCA-MM
Cluster1	22677 (34%)	22677 (35%)	22677 (34%)
Cluster2	1974 (3%)	1398 (2%)	2093 (3%)
Cluster3	20294 (30%)	20302(31%)	20185 (30%)
Cluster4	20590 (31%)	21158(32%)	20580 (31%)
All cluster	100%	100%	100%

From the experiment, it was found that clustering using PCA combined with maxima-minima functional gives the nearest result with 115 band original image and can reduces numbers of attribute from 115 down to 10 attributes. These selected attributes were used as input for clustering algorithms. Table 1 shows the percent clustering obtained for various classes for an example experiment. It can be noticed that the differences in clustering between original image and PCA-MM techniques are very closed. Because features from maxima and minima functional can use preserved features that separate different object classes and, both the representation information and class specific information are included.

Table 2. Show comparison PCA, PCA-MM, Original image by Simple K-Mean clustering

	7 Cluster 36 band Original image	7 Cluster 7 band reduce by PCA	7 Cluster 7 band reduce by PCA-MM
Cluster1	489 (8%)	675 (10%)	608 (9%)
Cluster2	1299 (20%)	1343 (22%)	1362 (21%)
Cluster3	927 (14%)	760 (12%)	742(12%)
Cluster4	572 (9%)	598 (9%)	586 (9%)
Cluster5	1315 (20%)	1504 (23%)	1504 (23%)
Cluster6	1043 (16%)	627 (10%)	881 (14%)
Cluster7	790 (12%)	928 (14%)	752 (12%)
All cluster	100%	100%	100%

One more experiment was tested a proposed band selection on the statlog (landsat satellite) data set from UCI databases [9]. The database consists of the multi-spectral values of pixels in 3x3 neighbourhoods in a satellite image, and the classification associated with the central pixel in each neighbourhood. One frame of Landsat MSS imagery consists of four digital images of the same scene in different spectral bands. Two of these are in the visible region (corresponding approximately to green and red regions of the visible spectrum) and two are in the (near) infra-red. Each pixel is an 8-bit binary word, with 0 corresponding to black and 255 to white. The spatial resolution of a pixel is about 80m x 80m. Each image contains 2340X3380 such pixels. These data contain 6435 in-stances. Each instance consists of 36 band attributes. The proposed process was implemented on java environment, and tested on CPU 2.80 GHz Intel(R) Core 2 Duo processor with 1 GB of RAM. From table 2, it can be noticed that the differences in clustering between original image and PCA-MM techniques are very closed. Following the results of the experiment are shown in Table 2.

4 Conclusions

Hyperspectral image software becomes widely used. Although hyperspectral images provide abundant information about bands, their high dimensionality also substantially increases the computational burden. Dimensionality reduction offers one approach to hyperspectral image (HSI) analysis. In this paper, we present band selection technical using principal components analysis (PCA) and maxima-minima functional for hyperspectral image such as small multi-mission satellite (SMMS). We tested the proposed process on satellite image data such as small multi-mission satellite (hyper spectral) for unsupervised classification. We compared this classification results between original images and PCA-MM by clustering. The experimental results show that the differences in clustering between original image and PCA-MM techniques are very closed. Because features from maxima and minima functional can use preserved features that separate different object classes and, both the representation information and class specific information are included. A result of this research was developed to provide users have been selected band for hyperspectral image. The out-come of this research will be used in further steps for analysis tools in hyperspectral image processing.

Acknowledgments. The authors would like to thank TGIST. The financial support from Thailand Graduate Institute of Science and Technology (TGIST) is gratefully acknowledged. The scholar student ID is TG-22-11-53-005D and the grant number is TGIST 01-53-005.

References

1. Richards, J.A.: Remote Sensing Digital Image Analysis: An introduction. Springer, Heidelberg (1986)
2. Small Multi-Mission Satellite (SMMS) Data, `http://smms.ee.ku.ac.th/index.php`
3. Agarwal, A., El-Ghazawi, T., El-Askary, H., Le-Moigne, J.: Efficient Hierarchical-PCA Dimension Reduction for Hyperspectral Imagery. In: 2007 IEEE International Symposium on Signal Processing and Information Technology, December 15-18, pp. 353–356 (2007)
4. Kaewpijit, S., Le-Moige, J., El-Ghazawi, T.: Hyperspectral Imagery Dimension Reduction Using Pricipal Component Analysis on the HIVE. In: Science Data Processing Workshop, NASA Goddard Space Flight Center (February 2002)
5. Maxima and Minima, `http://www.tutorvista.com/content/math/calculus/application-derivatives/maxima-minima.php`
6. Cheng, X., Chen, Y.R., Tao, Y., Wang, C.Y., Kim, M.S., Lefcourt, A.M.: A novel integrated PCA and FLD method on hyperspectral image feature extraction for cucumber chilling damage inspection. ASAE Transactions 47(4), 1313–1320 (2004)
7. Bouckaert, R.R.: WEKA Manual, WAIKATO University, pp. 1–303 (January 2010)
8. Kirkby, R., Frank, E.: Weka Explorer User Guide. University of Waikato, New Zealand (2005)
9. Frank, A., Asuncion, A.: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA (2010), `http://archive.ics.uci.edu/ml/support/Statlog`
10. Jackson, J.E.: A User Guide to Principal Components. John Wiley and Sons, New York (1991)
11. Jolliffe, I.T.: Principal Component Analysis. Springer, Heidelberg (1986)

Structuring Invention Knowledge to Creatively Solve a Technological Problem

Itsuo Makino and Takaya Yuizono

School of Knowlege Science, Japan Advanced Institute of Science and Technology,
Asahidai 1-1, Nomi-City, Ishikawa-Prefecture, Japan
{makino1000,yuizono}@jaist.ac.jp

Abstract. We have proposed structuring invention knowledge to effectively solve a technological problem. Our structuring method comprises steps for determining indispensable constituent elements of an apparatus to form an essential construction thereof, for a plurality of times, from a current apparatus up to the first model thereof respectively. The invention knowledge structure will form a cylindrical invention space for making an invention. Experiment results have revealed that structuring invention knowledge is thought to be possible for patent workers. When an invention space was formed by a series of essential constructions thereof, the participants produced technological ideas of higher values in terms of five evaluation items such as practicality and feasibility. Further, when a series of essential constructions were completely conducted in the invention space, there were produced ideas of higher evaluation values in terms of practicality and productivity.

Keywords: creative problem solving, patent systems, invention, knowledge structure.

1 Introduction

In the present societies, there are various kinds of problems to be solved [1]. Technological problems are often included in social problems. There are cases in which when a technological problem is solved, the social problem will be solved.

Patent Systems have been established in many countries such as in Japan and the United States for advancing technologies. As one of the features of the Patent Systems, as seen in the United States that the Patent System provides a pioneer invention with broader scope of protection as compared to the one of an improvement invention [2],[3],[4]. As another feature of the Patent System, patent applications filed at the Patent Offices will be made public. Thus, it is also considered that Patent Systems encourage people to utilize patent publications to make other inventions.

Some problems require us to develop a highly sophisticated technology such as a control device for controlling the movement of a space satellite so that a pioneer invention is made. There are also some problems requiring us to make improvement inventions. Such improvement inventions can be made by referring to existing patent publications and adding a new and small technological development thereto.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 113–124, 2011.

Our attention is directed to patent knowledge and particularly to constructions of, for instance, an apparatus varying in time to make new improvement inventions.

Our objective is to propose structuring invention knowledge to form an invention space so that a technological problem will be efficiently and effectively solved.

2 Related Works

2.1 Creative Problem Solving Methods

Various kinds of creative problem solving methods are introduced in the Handbook of Creativity such as the Brain Storming Method [5],[6]. Some methods are directed to produce technological ideas. It has been known Theory of Inventive Problem Solving (TRIZ) invented by G. S. Altshuller for producing inventions to solve technological problems [7],[8]. Good patented inventions were analyzed to know processes how the inventions had been made. A vast number of patents were analyzed to invent TRIZ. But, the patents were used to devise the methodology of TRIZ. It classifies problems to be solved into five layers. The first layer includes problems to be solved by using knowledge individuals own. The second layer includes problems to be solved by using knowledge organizations own. The third layer includes problems to be solved by using documents or information in the one industrial field. With the fourth layer of problems, knowledge another industrial field is used to produce a new concept to solve a problem. The fifth layer relates to develop a fundamentally new system. TRIZ is effective for the second, third and fourth layers of problems. Our approach proposed in this paper is effective for the first, second and third layers of problems.

It has been known Theory of Equivalent Transformation Thinking invented by Kikuya Ichikawa about sixty years ago [9]. This theory uses analogical thinking. Essence or feature of the base is determined and then transformed into the target.

TRIZ and the theory do not, however, teach ways to discover problems to be solved.

2.2 Patent System

An approach was proposed to creatively solve a problem using knowledge in patent databases. The approach was directed particularly to a way how to find relevant patent information to solve a technological problem [10]. While, our invention knowledge structure proposed in this paper directs attention to the constructions of, for instance, an apparatus for making an invention. The construction of the apparatus can be obtained from a patent publication. Further, a series of essential construction concepts or constructions selectively formed along a time axis will structure invention knowledge to provide a three-dimensional invention space for making inventions.

The invention protection model and the concept equivalent thereto as shown in Fig. 1 were proposed respectively by Makino in 1987 and Judge Rich for the United States Court of Appeals for the Federal Circuit [3],[4].

Figure 1 shows a model diagram explaining how an invention is protected or evaluated in the United States. A small blue circle k1 represents an embodiment of an invention. An outer red circle K1 represents the broadest scope of protection obtainable by an inventor with respect to the prior art which is shown in white outside of the red circle.

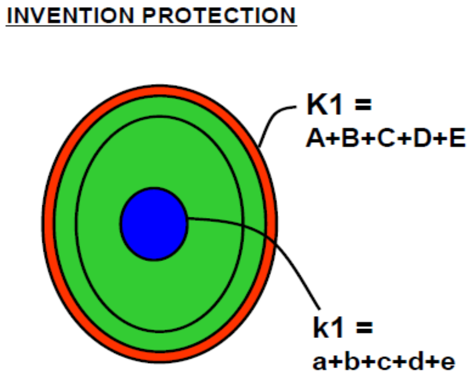


Fig. 1. Invention Protection Model

3 Structuring Invention Knowledge

We propose structuring invention knowledge to form a three-dimensional invention space in which obvious or implicit problems will be discovered and which will be useful for solving a technological problem to produce an invention.

Our attention is directed to constructions of an apparatus which vary in time in forming an invention space. The invention space is formed based on constructions of an apparatus. The cylindrical invention space for making an invention is considered compatible and integrated with the invention protection model shown in Fig. 1.

Differences between constructions of an apparatus are also important information which will enable one to make an invention.

It will be easily understood whether the resultant invention can be protected by the patent law and system, since the invention space is integrated with the invention protection model in Fig. 1. It is expected that the integration will promote making inventions. Hereinafter, our method as applied, for example, to an apparatus will be explained.

Flow diagram for the proposed method is shown in Fig. 2. Essential constructions each having indispensable constituent elements of an apparatus are selectively formed from the current apparatus up to the first model to structure invention knowledge or to form an invention space. Based on the structured invention knowledge and other information, a technological problem will be discovered, which will lead one to make improvement or new-use inventions. Attention is directed to constructions of an apparatus and differences therebetween.

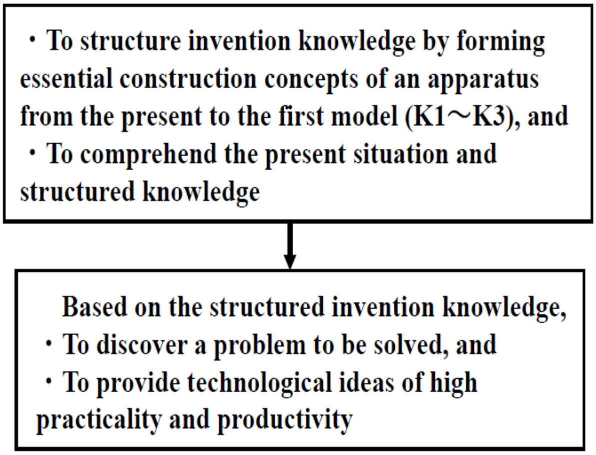


Fig. 2. Invention Process

Figure 3 shows an invention space. It also shows how the three-dimensional invention space is formed and the relationship between structuring invention knowledge and the invention space. It also shows the relationship between the structuring invention knowledge and the method to protect an invention. The invention space shown in Fig. 3 is related to the invention protection model shown in Fig. 1 in forming some constructions of an apparatus.

Referring to Fig. 3, four concentric circles are provided as K1, K2, K3 and K4 showing generalized or abstracted conceptual constructions. The circumference of a first red circle K1 is determined by five indispensable constituent elements A, B, C, D and E for an apparatus currently used. The generalized conceptual construction K1, for instance, is obtained by generalizing a construction of an apparatus comprising constituent elements a, b, c, d and e shown in Fig. 3. The circumference of a second green circle K2 is obtained by generalization and determined by four indispensable elements A, B, C and D for an apparatus manufactured at an older time instant. In the same way, the circumferences of a third purple circle K3 is obtained by generalization and determined by indispensable constituent elements A, B and C. The circumferences of a fourth yellow circle K4 is obtained by abstraction with respect to the circle K3 and determined by indispensable constituent elements B and G.

A three-dimensional invention space is made in a form of cylinder. The cylinder is determined by the circle K3 at the top, the circle k1 at the bottom and the circumference of the circle K3. At first, the present construction of a product is analyzed to determine indispensable elements thereof (k1), and then generalized to cover another example of product which functions in the same way as the present product (k1). The present construction of a product will be generalized also along a time axis up to the construction of the concept of the first product (K3) which was first put on the market. Between the present construction of a product (K1) and the construction of the product concept (K3), another generalization based on the present construction thereof is made at a time instant on the time axis (K2). In the same way, another generalization may be made along the time axis as required.

The invention space is formed by structuring invention knowledge or based on constructions of an apparatus. Within the imaginary cylinder, it is expected that problems to be solved will be discovered and thus, inventions will be made.

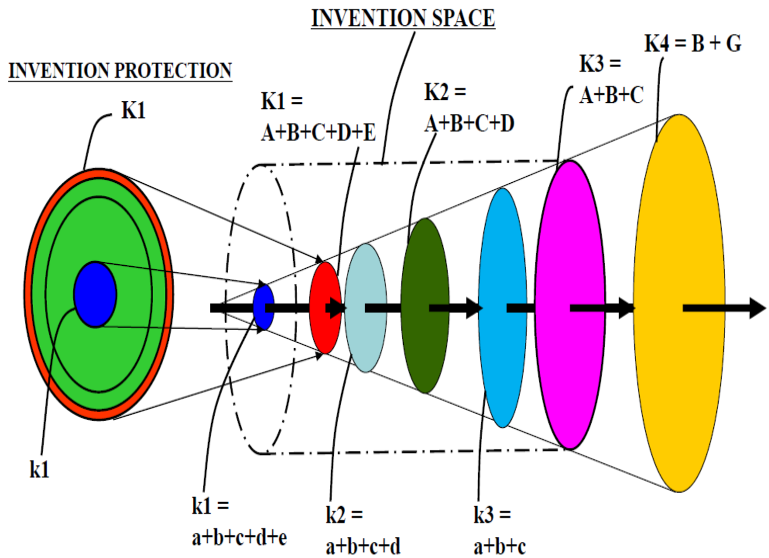


Fig. 3. Invention Space

4 Experiments

Participated are members of Intellectual Property Subcommittee in Kansai Electronic Industry Development Center (hereinafter referred to as KEC). The members have been engaged in jobs relating to patents such as assisting engineers to make inventions or filing patent applications at the Japan Patent Office, in electronics companies or a patent office from ten to thirty years. The members are not required in the companies or office to make inventions or to produce technological ideas.

4.1 Forming an Essential Construction Concept of an Apparatus

Tests were made on Karaoke apparatuses to see whether participants can form an essential construction, i.e., determine indispensable constituent elements of a Karaoke apparatus (hereinafter referred to as Karaoke). There were used a single Karaoke developed and used on the market for the first time and a so-called network-type Karaoke having units such as a server storing musical data, and a client comprising a display, computer and an operational panel which are mutually connected. We received from six participants experiment results and completed questionnaires.

4.1.1 Experiment Procedure

1. Participants are required to determine and describe indispensable constituent elements of a single Karaoke and a network-type Karaoke.
2. With the network-type Karaoke, participants are required to read Japanese Unexamined Patent Application No. 1993-165483 (hereinafter called as “the patent publication”). Then, the participants are required to describe indispensable constituent elements of the network-type Karaoke to provide the construction thereof.
3. With the single Karaoke, participants are required to describe the construction thereof by referring to the patent publication.
4. We asked all the participants,

4a. to read through the patent publication from the beginning to the end thereof,
4b. to underline portions they cannot understand,
4c. to determine and describe indispensable constituent elements of a network-type Karaoke generalized with respect to the single Karaoke,
4d. to describe the construction of a single Karaoke by referring to the patent publication,
4e. to measure times needed to do the assignments described in 4a, 4c and 4d respectively, and
4f. to answer the questionnaire.

4.1.2 Essential Construction Concept of Karaoke

The six participants described constructions of the network-type Karaoke and single Karaoke.

The construction of the network-type Karaoke is represented as ‘k1’ in Fig. 3. The essential construction concept of the single Karaoke is represented as ‘K3’ in Fig. 3.

An example of the essential construction of network-type Karaoke is shown as in the following. The example is described by one of the participants.

> A karaoke system comprising a plurality of terminal devices connected to a network, and a server connected to the network and comprising a memory for storing a plurality of kinds of songs each of which having text image data and accompaniment data and a communication control means for enabling each terminal device to be communicated separately, wherein
>
> the terminal device comprises a display unit for displaying images, a sound producing unit for outputting sounds, and an image display process means for outputting the song text image data received through the network to the sound producing unit.

4.2 Effectiveness of the Proposed Method

Participated are eight members of the KEC, who are different from the participants for the Karaoke experiment. An experiment was made on a vending machine with unwanted sound reducer to see the effectiveness of our proposed method for structuring invention knowledge to produce technological ideas. We received from the eight participants experiment results and completed questionnaires.

4.2.1 Experiment Procedure

The eight participants are divided into two groups. All of the eight participants are allowed to make searches to refer to any kinds of information and documents. The first group used the proposed method to produce ideas, while the second group produced technological ideas in their own ways without using the proposed method. Relevant assignments to produce ideas were conducted in companies the members of the groups belong to or in their homes.

To the first group

We explained our proposed method to the first group for structuring invention knowledge to produce technological ideas. We presented the first group with an unexamined patent application publication No. 2009-237719 with its title “A vending machine with feedback-type active sound reducer” and the questionnaire. We asked the members of the first group to read the publication, and to make any kinds of inventions such as improvement or small ones, to solve problems they may have or discover.

To the second group

We presented the second group with a letter explaining the experiment method and the unexamined patent application publication No. 2009-237719 and the questionnaire. We asked the members of the second group to read the publication, and to make any kinds of inventions such as improvement or small ones, to solve problems they may have or discover.

4.2.2 A Plurality of Essential Constructions of Vending Machine with Unwanted Sound Reducer

The four participants of the first group described a plurality of essential constructions of a vending machine with unwanted sound reducer.

As an example, the following shows a series of constructions formed by a participant of a vending machine with unwanted sound reducer.

Construction of current machine:

A vending machine with unwanted sound reducer comprising: a compressor for adjusting temperature, a microphone for detecting unwanted sounds, a speaker for producing control sounds, means for generating control signals to generate control sound signals, detection means for detecting unwanted sounds above a predetermined level, and a controller for controlling the means for generating control sound signals in response to an output from the detection means.

Construction concept 1:

A vending machine with unwanted sound reducer comprising: a compressor for adjusting temperature, a microphone for detecting unwanted sounds, a speaker for producing control sounds, and means for generating control signals to generate control sound signals.

Construction concept n:

An unwanted sound reducer comprising: a microphone for detecting unwanted sounds, a speaker for producing control sounds, and means for generating control signals to generate control sound signals.

Abstraction:

An unwanted sound reducer for reducing unwanted sounds by using sounds having 180 degrees out of phase with respect to the unwanted sounds.

5 Results

5.1 Essential Construction Concept of an Apparatus

Evaluations on the essential constructions of the network-type Karaoke and single Karaoke were made by three evaluators, Student A, Student B and Teacher E, who have electronics or mechanical backgrounds.

The essential construction reference of the network-type Karaoke had been drafted by the author. The constructions thereof described by the participants were compared with the concept reference respectively so that evaluations were made as Table 1 in the following.

The construction reference of the single Karaoke apparatus had been drafted by the author. The constructions thereof described by the participants were compared with the concept reference respectively so that the resultant average values were obtained as Table 1 in the following.

It is seen that the average values for the network-type Karaoke are high. While, the average values for the single Karaoke is lower than the ones.

Table 1. Evaluation on Karaoke

Evaluator	Student A	Student B	Teacher E
Average Value for network-type	81	72	78
Average Value for single-type	64	55	61

5.2 Effectiveness of Invention Knowledge Structure

Evaluation on a plurality of the constructions of the vending machine with unwanted sound reducer were made by three evaluators, Student A, Student B and Teacher E, who had understood well the function and the construction of the machine used for the experiment.

A plurality of the construction references of the vending machine with unwanted sound reducer had been drafted by the author. Each of the constructions of the vending machine described by the participants was compared with the construction reference so that an evaluation was made. An example of the evaluation is shown as Table 2 in the following. The evaluation relates to the constructions described by a participant belonging to the first group.

Table 2. Evaluation on a Series of Essential Construction Concepts

Evaluator	Current Construction	Construction concept 1	Construction concept n	Abstracted concept
Student A	100	100	88	83
Student B	93	90	50	33
Teacher E	100	100	50	67

Ten technological ideas were presented by the four participants A, B, C and D in the first group. While, eighteen ideas were presented by the four participants E, F, G and H in the second group.

Each idea is evaluated by the three evaluators with one of three, "F" having 5 points, "J" having 3 points and "K" having 1 point. For each idea presented, evaluation values given by the three evaluators are summed to be a total value in terms of six items as Originality, Practicality, Feasibility, Productivity, Inclusiveness and Insightfulness. It is to be noted that the evaluation items are taken from "Creative Cognition" by R.A. Finke et al [11].

Evaluation comparison between the first and second groups is made and its result is shown as Table 3 in the following.

There are shown average values of the ideas presented by the first group using our proposed method and by the second group producing their ideas in their own ways. A t-test is made on evaluation values.

As a result of the t-test and as seen in the Table 3, the ideas presented by the first group using our proposed method are more advantageous or preferable than the ones presented by the second group producing their ideas in their own ways in terms of five evaluation items as practicality, feasibility, productivity, inclusiveness and insightfulness.

Table 3. Idea Comparison between First and Second Groups

Evaluation item	Originality	Practicality *	Feasibility *	Productivity*	Inclusiveness*	Insightfulness*
First G. with Proposed Method	9.8	13.0	11.8	12.4	12.2	12.2
Second G. without the Method	10.7	8.4	9.0	7.9	8.7	9.1

T-test:*p<0.05

Evaluation comparison between the participants AB and CD in the first group is made and its result is shown as Table 4 in the following.

The table reveals that the ideas presented by participants A and B who conducted the proposed method completely are more advantageous or desirous than the ones provided by the other participants C and D who conducted parts of the proposed method in terms of three evaluation items.

Table 4. Idea Comparison in First Group

Partic-ipants	Origi-nality	Practi-cality	Feasi-bility	Produ-ctivity	Inclu-siveness	Insight-fulness
A, B	9.5	15.0	10.5	15.0	15.0	11.5
C, D	10.0	11.7	12.7	10.7	10.3	12.7

6 Discussion

6.1 Essential Construction Concept of an Apparatus

With the Karaoke test result, it is considered that the participated patent workers could form essential construction concepts. They obtained high evaluation values with the network Karaoke. Some of the participants had some difficulty in forming a construction of an imaginary apparatus with a single Karaoke. The average values for the network-type Karaoke are much different from the single Karaoke. It is considered that this difference comes from how the essential constructions of apparatuses were described for the respective apparatuses. With the network-type Karaoke, the participants could read the patent publication and understand the construction and operation of the network-type Karaoke and derive indispensable constituent elements of the apparatus. While, with the single Karaoke, the participants had to devise the construction of and to make a single Karaoke by themselves by referring to a network-type Karaoke. Some of the participants could cope with the assignment. But, the others could not. Some of the participants have some difficulty in forming a construction of the first model thereof. It is thought that the difficulty can be resolved, if they understand and recognize a similarity between the way to form an essential construction of an invention and the way to obtain a broader scope of protection thereon. In filing, for instance, a U.S. patent application, a patent worker is required to draft a patent claim for obtaining broader scope of protection for an invention. This job to draft a claim is similar to making an essential construction of the single Karaoke. Thus, it will be possible for patent workers to form an essential construction concept.

6.2 Effectiveness of Invention Knowledge Structure

With Table 3 for the first group using the proposed method and for the second group not using the method, it is thought that the first group structured invention knowledge to form the cylindrical invention space, and found relevant knowledge, information and technological problems in the invention space. It is thought that the invention space helped the participants of the first group to produce ideas of higher evaluation values in terms of five evaluation items such as practicality and productivity. It is to be noted that eight technological ideas out of the ten ideas presented by the first group relates to vending machines with unwanted sound reducer. We think that attention of the participants is directed to constructions of the apparatus and that these eight ideas are produced in the invention space determined by the plurality of the constructions formed.

With Table 4 for the first group having the four members A, B, C and D, it is thought that as the number of the formed constructions of the apparatus increases, differences between the constructions will be recognized and the invention space will be filled with more relevant knowledge and information so that the participants will obtain deeper understanding of the apparatus. As a result, the members A and B formed the whole invention space could produce technological ideas more preferable than the ones provided by the members C and D in terms of Practicality, Productivity and Inclusiveness.

7 Conclusion

It is thought that the proposed structuring invention knowledge to form an invention space is effective for making improvement inventions. We think that our invention space method will assist ones to make inventions.

It is considered that the patent workers participated can form essential constructions of an apparatus.

When the proposed invention space is used, there will be produced technological ideas of higher evaluation values in terms of five out of six evaluation items such as practicality, feasibility or productivity.

Further, when the whole imaginary invention space is formed, there will be produced ideas of higher evaluation values in terms of practicality, productivity and inclusiveness.

For future work, we are going to make the same tests on students to have more relevant data. With the experiments, participants were patent professionals.

We have structured invention knowledge for making improvement inventions. We are going to make researches to find how well the invention space will work in making pioneer inventions as well. Further researches will be made to devise business model inventions or patents leading to innovations.

References

1. Komiyama, H.: Japan, Advanced in Facing many World-new Problems (Kadaisenshinkoku Nihon). Chuo Koronshinsya, Tokyo (2007) (in Japanese)
2. U.S. Constitution, Article 1, Section 8, cl.8., `http://www.usconstitution.net/const.html` #A1Sec8

3. Rich, G.S.: Wilson Sporting Goods Co. v. David Geoffrey & Associates. 14USPQ.2d 1942, Fed. Cir. (1990)
4. Makino, I.: Lessons from Court Decisions Relating to Patent Infringement (Chitekizaisan-kenno Koboni Manabu). In: Intellectual Properties (Chitekizaisan), pp. 109–129. Kansai Electronic Industry Development Center (Kansai Denshikogyo Shinkosenta), Osaka Japan (2000) (in Japanese)
5. Takahashi, M.: Handbook of Creativity (Shinhen Souzouryoku-jiten), New Edition. Juse Press, Ltd. (Nikkagirensyuppansya), Tokyo (2002) (in Japanese)
6. Osborn, A.F.: Applied Imagination. Charles Scribner's Sons, New York (1953)
7. Altshuller, G.S., Shapiro, R.V.: About A Technology of Creativity. Questions of Psychology 6, 37–49 (1956)
8. Sanno University CPM/TRIZ Research Group (Sangyonouritsudaigaku CPM/TRIZ kenkyukai). TRIZ Theory and its Development (TRIZnoriron to sonotenkai). Sanno University Publishing Department (Sangyonoritsudaigaku syupanbu) (2003) (in Japanese)
9. Ichikawa, K.: Creative Engineering (Sozokogaku), Rateisu, Tokyo (1977) (in Japanese)
10. Makino, I., Yuizono, T., Nakamori, Y.: An Approach to Creatively Solving a Problem using Knowledge in Patent Databases. In: The 4th International Conference on Knowledge, Information and Creativity Support Systems, pp. 109–116 (2009)
11. Finke, R.A., Ward, T.B., Smith, S.M.: Creative Cognition Theory, Research, and Applications. A Bradford Book. The MIT Press, Cambridge, Massachusetts (1992)

Development of a Search System for Heterogeneous Image Database

Sanparith Marukatat

IMG lab
NECTEC
112 Thailand Science Park, Pathumthani, Thailand
sanparith.marukatat@nectec.or.th

Abstract. Nowadays, multimedia data, especially image, are of increasing number. Content-based image retrieval system is becoming an important tool to assist user in managing his/her collection of images. This work presents a development of an image search system for a particular image database containing various types of images. We present a study of visual descriptors for this database and a simple strategy to speed-up the retrieval process. We also present a relevance feedback technique based on semi-supervised learning technique. Experimental result of the proposed system seems promising.

Keywords: Image retrieval, visual descriptor, relevance feedback.

1 Introduction

Nowadays, multimedia data, especially image, are of increasing number. Multimedia management systems, both storage and retrieval systems, are thus important for user. Basic image search system relies on image description or text that appears near image to index and then to retrieve using text query. This approach gives reasonable result for mixed media data like web page. However, for an image database a content-based image retrieval (shorten as *CBIR* hereafter) is needed.

CBIR system extracts some visual descriptors from image and used them to index and to compare images in the retrieval process. Existing CBIR systems (e.g. [1,4,5]) differ principally in the choice of visual descriptors used. It should be noted that the choice of features also depends on the type of images present in the database. For this work, we are interesting in developing a CBIR system for a heterogeneous image database. Indeed, we are given a set of images of various types such as drawing with pencil, ink, or water color, photography images of people, building, archaeological site, mural painting, etc. Figure 1 shows some of the images from the database. An objective of this work is to study the visual features for this heterogeneous database.

Besides, visual descriptors no matter how sophisticated they are still cannot cover all notion of similarity for all users. Therefore, this work also investigates the relevance feedback mechanism for image retrieval. Indeed, the relevance feedback is considered in a semi-supervised learning framework. Graph-based technique is used to filter images that could correspond to user requirement.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 125–133, 2011.

Fig. 1. Example of images in the database

In the following, Section 2 describes the study of different visual descriptors used in this work. Section 3 presents a strategy used to speed up the retrieval process. Section 4 presents the relevance feedback mechanism used and Section 5 concludes this work.

2 Visual Descriptors

The given database is composed of 21667 images which can be roughly classified in two main groups namely color images and grayscale images. This classification can be done using the average difference between the maximum and the minimum values of the three channels (RGB) for each pixel. Using this simple rule, we can identify 4120 color images and 17457 grayscale images. From visual inspection, one can see that grayscale images are more textural than color images. Therefore, different visual descriptors are used for these two groups.

2.1 Color Descriptor

Colorspace. The most fundamental color descriptor is the color histogram. Given 2 color histograms x and y, the similarity between them is measured by the *histogram intersection*:

$$HI(x, y) = \sum_{i=1}^{m} \min\{x_i, y_i\} \tag{1}$$

with m the total number of colors or *bins*. This similarity measure is used to retrieve similar image with respect to the query image.

Two classical colorspaces were considered in this work namely Red-Green-Blue or RGB colorspace and Hue-Saturation-Value or HSV colorspace. Figure 2 show search results from RGB colorspace compared to those obtained from HSV colorspace. For each result, the top left image is the most similar image to the query. Note that in all cases, the most similar image is exactly the query image. Figure 2 show results obtained from different number of bins, i.e. $3\times3\times3$, $4\times4\times4$, $5\times5\times5$, and $8\times8\times8$ bins. From these results, one may see that for low number of bins, the results from HSV colorspace are visually better that that of RGB colorspace. As the number of bins increases, the results from both colorspace become more and more similar with, however, a slight advantage for HSV colorspace. As a consequence, the HSV colorspace will be used in the following.

Fig. 2. Comparison between RGB colorspace and HSV colorspace

Number of Bins. In previous experiment the same number of bins is used for all channels H, S, and V. However, different channels in HSV colorspace has different meaning. Thus, they should have potentially different importance. Figure 3 shows example of search results obtained with different number of bins for H, S, and V channel. The total number of bins is kept constant, and we varied the relative important of each channel by assigning more or less bins to them. From these results, it is still unclear what is the optimal configuration of this histogram. Therefore, in the following, the color histogram in HSV colorspace with equal partition for each channel will be used.

Spatial Information. Color histogram alone can capture only the relative frequency of each color in the image, but cannot describe how these colors are distributed in the image. This kind of information requires additional knowledge about spatial distribution of colors in the image. In [3], the authors proposed a notion of *coherence*. A pixel is considered as coherent, if it is spatially connected of a large group of pixels having the

Fig. 3. Search results using HSV colorspace with different number of bins nh $\times$ ns $\times$ nv where nh is the number of bins for the H channel, ns for S channel, and nv for V channel

same color. For each color, the number of coherent pixels and the number of incoherent pixels are kept as *color coherence vector* or *CCV*. The authors proposed using norm-1 distance to compare two CCV. In term of memory requirement, this feature requires the double of normal color histogram.

The idea of coherence is interesting, but their calculation requires computing connected components for each color. Also noted that, both coherent and incoherent pixels are used in final description. One may deduce then that the incoherent pixels do contain information about the similarity between image as well. Besides, the CCV does specify the number of incoherent pixels but do not describe *how incoherence* are they. From this observation we propose a new spatial information based on histogram of difference between color of each pixel and the average color in local window centered at that pixel. For instance, only the hue value of is used for computing this difference. Integral image technique [6] is adopted to rapidly compute the average hue in local window with complexity independent from the size of local window.

Figure 4 show example of images retrieved by color histogram of 512 bins (a), and by color histogram and histogram of spatial hue difference using different window's size; 4x4 (b), 10x10 (c), and 20x20 (d). The spatial hue difference is quantized into 18

(a) (b) (c) (d)

Fig. 4. Example of images retrieved by color histogram (a), and by color histogram and histogram of spatial difference with different window's size; 4x4 (b), 10x10 (c), and 20x20 (d)

bins. From these results, one may see that the retrieved images look globally similar but the one retrieved with the spatial hue difference seems locally close to the query image as well. The results obtained from the window's size 4x4 seems slightly better than the other. Hence, in the following the color histogram of 512 bins with histogram of spatial hue difference quantized into 18 bins with local window of 4x4 will be used resulting in a set of 530 features as visual descriptor for color images.

2.2 Grayscale Descriptor

The most simple visual descriptor for grayscale images is the histogram of grayscale levels, similar to the color histogram for color images. This descriptor describes the color aspect of image but not the texture of the considered image. To describe texture, two types of visual descriptor are often used in literature namely the Gabor descriptor and the local binary pattern or LBP [2].

Gabor descriptor is obtained by passing the grayscale image into a bank of Gabor filter of different orientations and different scales. The sum of absolute values from the each filter response is used as feature. In this work, 4 orientations and 4 scales are considered leading to a feature vector of 16 dimensions. For a large area containing several textures, a single Gabor feature vector may not be suitable. Therefore, the image is partition into 3x3 grids, and a Gabor feature vector is computed for each area in addition to the global Gabor feature vector. In total a feature vector of 160 dimensions is used to describe a grayscale image. Euclidean distance is used to compare two images.

LBP descriptor is a histogram of LBP codes obtaining by comparing the grayscale value of each pixel to that of its 8 neighbors. This comparison results in 8 binary values which can be converted into an integer between 0 and 255. Followed the suggestion in [2], the 36 rotation invariant LBP codes are considered in this work. As for Gabor descriptor, a global LBP histogram and 9 LBP histograms computed from 3x3 area of the image are used. In total a feature vector of 360 dimensions is used to describe a grayscale image.

Figure 5 show example of images retrieved by grayscale histogram (a), Gabor descriptor (b), and LBP descriptor (c). In every cases, the top-left image retrieved is matched with the query image. From these results, one can see that the LBP feature yields more satisfactory result in most cases. As a consequences, in the following we shall use the LBP descriptor to describe grayscale images.

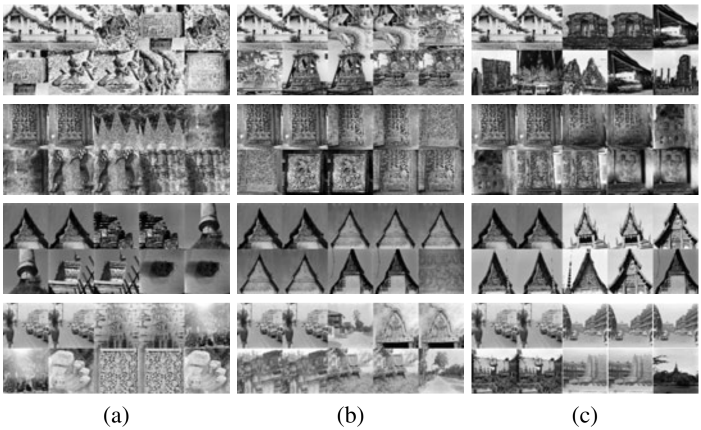


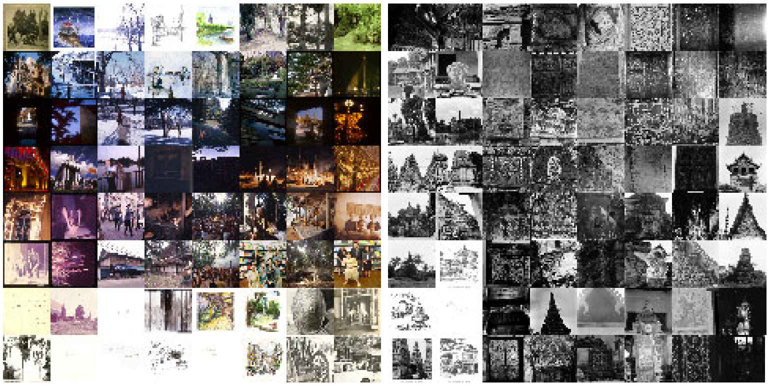
Fig. 5. Example of images retrieved by grayscale histogram (a), Gabor descriptor (b), and LBP descriptor (c)

3 Speed up the Search

The speed of the retrieval process depends on the number and the type of operations that are used to compare image descriptor. With the selected descriptors, 530 operations on float are needed in order to compare two color images while 360 operations on float are needed for grayscale images. As the size of database could be increased in future, it become interesting to prepare the retrieval system to handle large dataset. A simple technique to speed up the search is to partitioning the images into cluster, then to consider only images belonging to the same or the nearby cluster as the query image. In this work, self-organizing map (SOM) is used to clustering images. Figure 6 shows the 8x8 nodes SOM obtained from color images (a) and from grayscale images (b).

Given the above color SOM (respectively grayscale SOM), each color image (respectively grayscale image) can be encoded into a 64-bit string by first setting all bits to 0. Then compute the node whose prototype is closest to the considered image, and set this node and its neighbors to 1.

In the retrieval process, images whose binary string is close to the binary string of the query image, in the sense of Hamming distance, are first selected. These selected images are then sorted using the visual descriptor as usual. Using this simple strategy, the number of candidates can be reduced to 100 images or less for color images and to about 300 images for grayscale images without losing the accuracy.



(a) (b)

Fig. 6. SOM obtained from color images (a) and from grayscale images (b)

4 Relevance Feedback

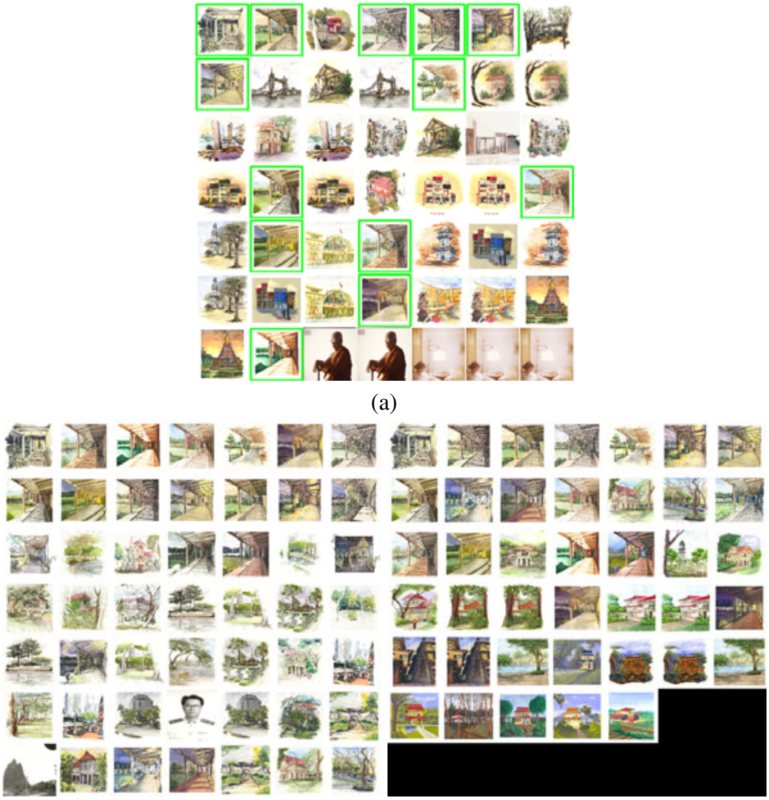
Even if the selected visual descriptors seem to perform well in our preliminary test, different users may be satisfied by the search result differently. To increase the satisfaction of the user, a relevance feedback mechanism is considered. Indeed, user is allowed to give a feedback on each retrieved image, whether he/she likes it or not. Then the system processes this feedback in order to retrieve images which could better satisfy the user.

The simplest method to do the relevance feedback consists in modifying the query's feature vector to get close to feature vector of images that user likes and to move away from the images that user dislikes. The same adaptation is used to construct SOM can also be applied here. Then the retrieval can be done using this modified query's feature vector. Figure 7 (a) shows images that have been retrieved by the system. Images that are marked in green are labeled like by user while all other images are disliked. Figure 7 (b) shows images obtained by modifying the original query's feature vectors using the first round's relevance feedback. One can observe that the images that user likes get better rank, and the newly retrieved images do appear similar to what user likes. However, there are still some retrieved images that appear to be erroneous.

To better analyze the relevance feedback information, one can consider it as a semi-supervised learning problem. In fact, we have a large number of unlabeled images and a small set of images labeled by user either like or dislike. Semi-supervised learning techniques can be used to propagate these labels to all images in the database automatically. A graph-based semi-supervised learning technique [8] is used in this work. Indeed, if we consider the probability $P(like|i)$ for each image $i = 1, ..., n$ in the database. This probability can be updated as follow

$$P(like|i)^{(t+1)} = \sum_{j \in Neighbor(i)} P(like|j)^{(t)} P(j|i) \quad (2)$$

where $P(like|i)^{(t)}$ is its estimation after t iterations, and $P(j|i) \propto SIM(i, j)$.



(a)

(b) (c)

Fig. 7. Images that have been retrieved in first round (a). The images that user likes are marked in green, all other images are disliked. Figure (b) shows images obtained by modifying the original query's feature vectors using the first round's relevance feedback. Figure (c) shows images obtained by using graph-based semi-supervised learning followed by the modified query's feature vector.

After the convergence, the images are sorted in decreasing order of $P(like|i)$ and the value at position $n \times \frac{n_+}{n_+ + n_-}$ is used as a threshold value, n is the total number of images, n_+ is number of images that user likes and n_- is the number of disliked images. All images having probability above this threshold are retained and resorted in decreasing order of similarity to the modified query's feature vector. Figure 7 (c) shows images retrieved by this relevance feedback mechanism. One can see that these images do correspond to user desired better than the images retrieved using modified query's feature vector alone.

5 Summary and Conclusion

We have presented the development of an image search system. In this development, we focus on three aspects in particular namely the selection of visual features, the speed of the retrieval process, and the relevance feedback mechanism. For visual features, we have proposed a new method to integrate spatial information into color feature. Experimentally, this additional information is proved to be useful to capture spatial information of color in the image. The comparison between the proposed method and existing method such as CCV is currently under investigation. We have also proposed a simple method to encode images into binary string to speed up the retrieval process. Indeed, our first objective of constructing SOM is for visualizing the database's images. Its use for binary encoding is a by-product of the SOM construction. There are other techniques that can be used for better converting the feature vectors into binary strings such as spectral hashing [7]. This will be further investigated in our future work. Finally, we have developed a relevance feedback mechanism based on a semi-supervised learning technique. This allows improving the accuracy of the retrieval system.

References

1. Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P.: Query by image and video content: The qbic system. IEEE Computer 28(9), 23–32 (1995)
2. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary pattern. IEEE Transactions on Pattern Analysis and Machine Intelligence (2002)
3. Pass, G., Zabih, R., Miller, J.: Comparing images using color coherence vectors. In: Proceedings of the ACM Multimedia (1996)
4. Smith, J.R., Chang, S.F.: Query by color regions using visualseek content-based visual query system. In: Proceedings of the ACM Multimedia (1996)
5. Vasconcelos, N.: From pixels to semantic spaces: Advances in content-based image retrieval. IEEE Computer, 20–26 (2008)
6. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
7. Weiss, Y., Torralba, A., Fergus, R.: Spectral hashing. In: Advance in Neural Information Processing Systems, NIPS (2008)
8. Zhu, X., Ghahramani, Z., Lafferty, J.: Semi-supervised learning using gaussian fields and harmonic functions. In: Proceedings of the International Conference on Machine Learning, ICML (2003)

Distribution of Evolved Knowledge Based on Adapting Context in a Region

Taizo Miyachi, Yusuke Ito, Atsushi Ogiue, and Yuhei Yanagawa

School of Information Science and Technology, Tokai University
4-1-1 Kitakaname, Hiratsuka, Kanagawa 259-1292, Japan
miyachi@keyaki.cc.u-tokai.ac.jp

Abstract. Timely problem solving services for people all over to every corner of society is one of main purposes in ubiquitous computing society. People should timely share awareness and useful knowledge with context in order to take suitable actions in emergency time etc. We call such relevant knowledge "Evolved knowledge." We propose a methodology in order to represent evolved knowledge with context and to distribution it to the people by success network. We also discuss what kinds of evolved knowledge with contexts is useful in a region and how evolved regional knowledge should be timely transmitted and expanded to the people corresponding to the change of environments in a long time period. Effective network architecture for switching knowledge between "foreground" and "background" in a context based on autonomous context node is also discussed.

Keywords: success network, action list in a context, evolved regional knowledge, knowledge distribution, and autonomous context node.

1 Introduction

Timely solution services for people all over to every corner of society based on the observations by sensor networks and human is one of main purposes in ubiquitous computing society. Knowledge by both experts and a person with well experiences should be timely transmitted to the people corresponding to user's context and dynamic change in the environment. Expanding useful knowledge over different regions is also important for the people in order to take effective actions in emergency time.

We propose a methodology of distribution and expansion of such evolved knowledge and to both residents and visitors corresponding to the dynamic change of environments in a certain long time period. We also discuss a network architecture that is suitable for such adapting knowledge distribution based on an autonomous context node and reflection node.

2 Useful Regional Knowledge and Context

Timely knowledge distribution in a region is very useful. We suppose that there are two major types of contents; (a) **happy content** and (b) **contents for avoiding risk**.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 134–143, 2011.

(a) Happy content depends on user's preference that dynamically changes while taking actions. (b) is very important since people must survive in a serious case. People all over to every corner of the society should know the indispensable knowledge to avoid the risk in a serious case like tsunami. Knowing the solution of a problem is important but it is not enough to solve the whole case since the case often has the other problems. In such a case knowing the context is more useful than knowing a part of problems with solutions. A person can take a good action chain corresponding to the target context.

■**User's positions for contents** can be classified into (1) to (4).

(1) **Easy information access**. Enjoy accesses to content for a pleasure.
(2) **Deep emotion**. Be impressed by a work of art or a structure etc.
(3) **Action improvement**. Take better actions based on timely awareness in a context.
(4) **Breakthrough in crisis**. Acquire knowledge for safe life: how to take refuge, how to find an available hospital for a baby, etc.

Users all over the world usually acquire the same information like ranking information by Internet search engines in the position (1). Context becomes more important for (2), (3) and (4) than (1). The main reasons of this are (a) to (f).

(a) A variety of meaning. Experts and a person with rich experiences describe a variety of meaning in a short description. Users can easily understand neither all of them nor their relationships in a short time period when they encounter the problem.
(b) Deep meaning. Human can understand deep meanings step by step based on his/her background knowledge and experiences. Experts explain lots of possibility with rich background knowledge and experiences. They also prospect typical results.
(c) Change of value based on a context. Values of knowledge change and even important knowledge switches to the other one because of the change in the context.
(d) Switch of the key functions occurs based on the change of the context.
(e) Quick actions. The best quick actions for an emergent issue should be easily understood and taken by everyone in every corner of the society.
(f)Continual actions without carelessness in a certain long time. Adequate actions should be continually taken in a certain long time period. A sensor network system keeps observing lots of targets and the environment for 24 hours. Human sometimes misunderstands that the environment becomes safe.

2.1 Context and Human Actions

Anyone in every corner of society should be able to timely acquire suitable awareness and knowledge for his/her context [8] by easy terminals such as Internet TV, digital signage and electronic book like iPad. Context can be clear by lots of sensors in a certain time period. "Managing long duration" is indispensable not only for the observation but also for distribution of useful knowledge in the ubiquitous computing society since human can neither watch lots of targets for a long time nor timely distribute essential knowledge to lots of people in a certain time period.

(1) Definition of Adapting Context

Adapting Context (AC) consists of nine attributes: Name, Region-map, Duration, Foreground, Background, Pre-conditions, Post-condition, Action-list, and Multi-view. All attributes are described in text, voice, image, animation and movie. AC describes thought, instinct, state of mind and emotion both in a virtual reality space and in the real space. Name is also a trigger of an image of AC (Refer to Fig.1).

Adapting-Context ::= Name | Region-Map |Duration | Foreground | Background | Pre-conditions | Post-condition | Action-list | Multi-view, Region-Map ::= | Place | Position | Relationship | Time | Region-risk | Region-environments | Potential | Depth
Pre-conditions ::= Direct-conditions | Background-conditions
Action-list ::= (AID | Action | Place | Position | Relationship | Time | Risk | Environments | Potential | Depth)*,
Environments ::= Spatial | Social, Depth ::= Future-extension | Source | Worst-case,
Multi-view ::= (Bias | Expand | Adapting-Context)*

An adapting context has a target region and its map in order to share essential knowledge. Duration means not only the exact time but also irregular occurrences in certain long time period. Examples of the duration are occurrence after five minutes, repetitions, and by 3rd September 2010. AC describes not only "Foreground" but also "Background." Foreground describes about main targets. Background describes about the second targets or a background. Pre-conditions describe a change of environments in "Direct-conditions" with "Background-conditions" that is a kind of indirect cause of the context. The clear description of relationship between Direct-conditions and Background-conditions is important to understand detailed context and to decide the order of actions. Background-conditions also describe preparation for actions. Action-list describes a list of actions with Place, Position, Relationship, Time, Risk, Environment, Potential and Depth. Time describes not only the adequate time but also both interval time between actions and a certain time period. Easy actions are chosen in case of weak relationship. Potential describes a possibility or an easy job for human's ability [9]. Multi-view shows a set of targets in order to expand adapting knowledge in emergency time and to extend pleasure for a target person and to avoid risks. Bias is used to adjust user's prejudice and to choose an appropriate knowledge. A visitor has a prejudice. Suitable AC should be chosen corresponding to the conditions which is described in Pre-conditions and Multi-view.

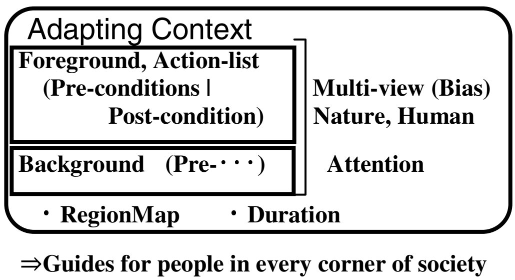


Fig. 1. An image of structure of AC

Example 1. An AC in case of typhoon. A farmer wishes to go to his/her field in order to check either the safety net of crops like apple or the water in a paddy field avoiding the danger in a typhoon. The check of the crops is described in Foreground.

Avoiding the risk in the typhoon is described in Background when he/she goes to the field.

■An example of the description of a AC

ContextEx1 ::= Fieldwatch-in-typhoon, Crop-map-typhoon, 02-Sep-2010to03-Sep-2010, ("Check crops and the fields": 1, 2, 3, 4),
("Keep safety in a typhoon by yourself":11, 12, 13, 14)
[{("If the typhoon would neither be so strong nor have heavy rain")}, {("If you can not ensure your safety), (If you feel some danger)}],
[(1, Walk to the field, Ground, Agriculture area, Owner, Typhoon, 40%, Hill, Strong typhoon 60%, familiar), (2, Check the field, Field, Near the field, Owner, Typhoon, 40%, Hill, Strong typhoon 60%, familiar), (3, Check crops, Field, in the field, Owner, Typhoon, 50%, Hill, Strong typhoon 70%, familiar), (4, Come back to the house, Agriculture area, Owner, Typhoon, 60%, Hill, Strong typhoon 80%, familiar), (5, Record the damaged objects and the reason, ・・・),
(11, Never walk close to water, Ground, Agriculture area, Owner, Typhoon, 80%, Hill, Strong typhoon 90%, familiar), (12, Never walk in the water, Ground, Agriculture area, Owner, Typhoon, 80%, Hill, Strong typhoon 90%, familiar), (13, Keep a distance from the water area, ・・・), (14, Find a safe route, ・・・), (15, Come back in the center of a safe road in the safe route, ・・・), (16, Call 119 (911), ・・・) , (17, you must not go outside and stay in a house, ・・・) , (18, You must go to a safe building, ・・・)],
[expand(Agriculture-area, Paddy, Riverside, Mountain, Fisherman-area → Fruit, Beach side, Town, Ria (saw-tooth) coastline area), {D:/video/video & picture in the field}].

(2) Context in high context culture

People usually share lots of embedded meaning and context in short expressions in **"**high context culture**"** [1, 2]**.** Such sharing allows a person to avoid some troubles and to effectively complete a job. Context is especially useful in a highest context culture. They say that "People that have eaten rice in the same iron pot can easily share a context with little information" in Japan. People can highly share AC in a region in high context culture and can mutually suppose the intention of the other person with neither efforts nor skill in communication in the high context culture.

Example 2. Collaboration based on a context in an emergency time

A farmer looks up the course and the change of strength of typhoon when a typhoon suddenly comes for the farmer's fields of apple. The other farmer who is sharing the context in an emergency time calls in order to check the stock of safety nets for preventing apples falling from trees without saying anything together.

2.2 Success Network and Adapting Contexts

Context becomes more important in "low context culture" than in "high context culture" [1]. "Success network" represents major success stories with AC in a region that have a list of actions in the AC for the success (Refer to Fig. 2). It guides users to the success. A success network consists of two layers (i) success context layer and (ii) success creation layer.

(i) Success context layer (SC layer): In "success context layer" success network represents a list of actions in an AC for a great success. Success context layer allows a user to know a local knowledge and specific actions that are chosen for both a target region and a specific situation in order to solve the local problems. Success network can describe useful thought, instinct, emotion, and collaboration in a region. It also has meta-data related to user's preferences and user's bias for a target since users including visitor in a region usually have a different culture and different bias.

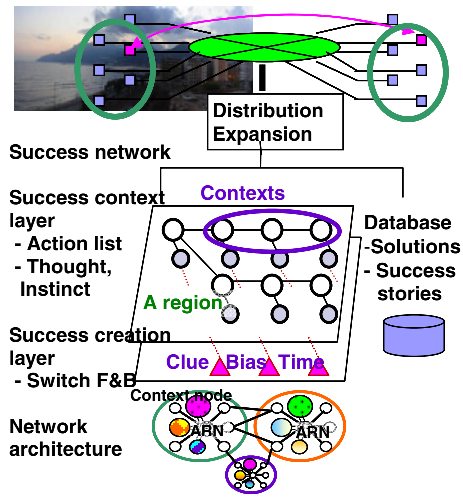


Fig. 2. An image of success network with contexts and context node with active

(ii) Success creation layer (SR layer): Success creation layer in Success network has meta-data related to time information, clue, bias and key information in order to timely improve success chains in SC layer. User's bias for a target issue changes both recognition of an encountered context and decision of solution. SR layer also manages switching from foreground to background corresponding to an emergent change in environments like a strong typhoon.

3 Evolved Knowledge in a Region

Internet users can easily acquire simple information like ranking information via the search engines. However they can not often acquire indispensable information in order to avoid risks and to solve problems because of the flood of information in the Internet. Two kinds of knowledge services **(A) aware-ness service in a context** and **(B) timely solution service in a context in regional environments** should be timely distributed to people all over to every corner of the society based on their background knowledge. Major relevant knowledge for the services can be classified into five categories (i) General knowledge, (ii) Deep knowledge, (iii) Predictable knowledge, (iv) Scheduled knowledge and (v) Adapting knowledge.

3.1 Distribution in Knowledge Service

(1) Distribution of predictable knowledge in a region in certain time period
The regions can be mainly classified into (a) geometric and cultural region and (b) region of a field. Predictable knowledge is indispensable for human since human must sleep and work busy. Success network timely distributes such knowledge in a target region.

Example 3. Flood by sudden heavy rain (region (a))
People who enjoy camping in a sand bank in a river sometimes fall in sleep there although they know about flood by sudden rain at night. The success network can timely inform them the flood when it detects sudden rain in near mountain in place of human. The people can survive the information by a mobile terminal.

(2) Distribution of scheduled deep knowledge in a long time period
People all over to every corner of the society should know the indispensable knowledge to avoid the risk in a serious case like tsunami. Natural disaster is a complex system and has lots of problems. In such a case knowing the context is more useful than knowing a part of problems with solutions. A person can take a good action chain corresponding to the target context. Useful knowledge should not only timely make a user be aware of problems but also show how to solve them corresponding to the progress of the user in a long time period. Regional knowledge should also repeatedly be supplied to all users in a target region at an adequate time.

Example 4. TSUNAMI. A tsunami, also called seismic wave, can travel great distances very quickly without losing much of its force at speeds of up to 500 miles per hour. In this way, a tsunami can cause great damage. As well, a tsunami's height in open water may only be a few feet, but its height can increase greatly in shallower water near land. Success network transmits not only quick refuge from tsunami on the hilltop in the first tsunami but also the deep knowledge scheduled in a long time period fifty minutes to all users that live in the beachside in a saw tooth coastline. The deep knowledge is “Both second and third tsunami are bigger than the first tsunami even if the first tsunami would be small. People should stay in the hill area.”

(3) Adapting knowledge: Switch from Foreground to Background in a context
A person has an original aim in “Foreground” in an AC. However he/she can not help switching from the aim in Foreground to an aim in Background corresponding to the change of environment in an emergency time and should takes quick actions.

Example 5. Switch in case of strong typhoon
Watching the crops in the field is in Foreground in example 1. However it should suddenly move into “background” when a typhoon becomes too strong. “Keeping the safety in the strong typhoon” should become a new foreground for the person and “watching crops” move in the background for him/her. Foreground and Background switch each other because of the bad weather (refer to Fig 3).The knowledge system

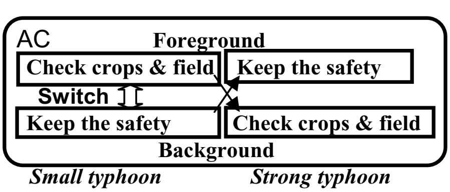


Fig. 3. Switch foreground and background

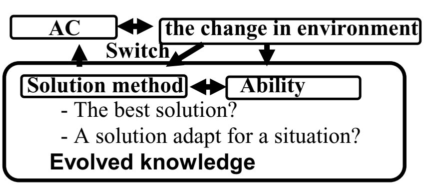


Fig. 4. Switching solution corresponding to the change in environments

with sensors should let him/her be aware of the switch corresponding to the emergent change of conditions of the environment in a certain time period (refer to Fig. 4)

An example of switched context of ContextEx1 is described in SwitchEx1.

SwitchEx1::= Fieldwatch-in-typhoon, Crop-map-typhoon, 02-Sep-to03-Sep-2010, ("Keep safety in a typhoon by yourself":11, 12, 13, 14),
("Check crops and the fields": 1, 2, 3, 4),
[{("If you can not ensure your safety), (If you feel some danger)}, {("If the typhoon would neither be so strong nor have heavy rain")}], Action-list, · · ·

The success network distributes suitable knowledge to people all over to every corner of society in target areas corresponding to the change of environments. The change can be automatically detected by lots of sensors in sensor networks in the ubiquitous computing environment.

3.2 Evolving Knowledge and Expanding Evolved Knowledge

People can acquire a basic law in explicit knowledge and utilize it in case that a causal relationship is clear. We call such knowledge "General knowledge" (Refer to Fig. 5). Human sometimes unconsciously chooses a behavior and shares such unconsciousness with the other persons. Carl Gustav Jung found that human naturally has collective unconsciousness that might be a base of symbiotic relationship between human [4, 3]. Human has been evolved from the ape for a very long time and acquired the ability of sharing a variety of knowledge based on genetic evolution and using letters. A person with well experiences has ability to timely solve problems even in complex systems like natural phenomenon. Genetic evolution for a long time and accumulation of experience made the ability [7]. However human can easily solve real time problems by sharing evolved context AC that are timely transmitted to people all over to every corners of the society in the internet. Not only explicit knowledge but also some implicit knowledge including intuition could be timely shared for a regional change. We call the knowledge in such contexts "Evolved Knowledge". Human can evolve a shared consciousness by this new methodology faster than genetic evolution.

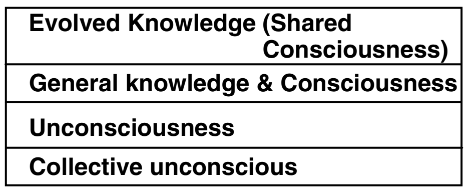


Fig. 5. An image of Evolved Knowledge

(1)How to create and use evolved knowledge

Experts and any persons with well experiences can describe knowledge for a target region. The users access the knowledge with the voted points by mobile phone and digital signage terminal etc. The experts review and choose highly ranked evolved knowledge that users take actions in a certain long time period in their environments. Important knowledge is also updated by the experts. The users can access both best solutions and special case solutions in the evolved knowledge. The knowledge distribution system periodically throws away useless knowledge based on the voted points etc.

(2)How to expand adapting knowledge – trigger and basic mechanism

Serious situation happens in lots of region in a dynamic change by natural phenomenon etc. The kind of problem changes by trouble factors. Heavy rain and strong wind causes many kinds of damages. These problems have relationships each other based on the characteristics of typhoon etc.

The change in the environment should be detected by not only sensor networks but also decision of actions for the change by a person with well experiences or an expert. The detection is done by not only the evaluation in multiple check points but also considering the experiences and intuitions by experts.

Success network expands adapting knowledge by five mechanisms (a) to (e).

(a) Context condition. When a person that knows the condition in context ensures that it becomes true he/she transmits the adapting knowledge to the others.
(b) Sensor information. When a sensor detect that the conditions become true the system automatically expand the adapting knowledge.
(c) Resident decision. When a resident ensures that the conditions become true he/she sends the context to the system. The system expands the knowledge.
(d) Expert decision. When an expert supposes that the condition becomes true he/she sends the context to the system in order to expand the adapting knowledge.
(e) Rescue team decision. When a rescue team finds that the conditions become true during his work it sends the context with typical actions to the system.

Multi-view in a context defines such relationships so as to expand adapting knowledge.

4 Network Affordance with Active Reflection Node and Context

Easy multimedia knowledge should be timely distributed to people all over to every corner in a region even in emergency time. Lots of accesses to the large volume of such knowledge concentrate in a certain time period in emergency time. Such concentration of massive accesses causes a bottle neck of the Internet access. We propose new network architecture in order to effectively distribute indispensable regional knowledge and context to the people even in emergency time.

An “active reflection node (ARN)” [5, 6] sends a message to itself via a “context node” that corresponds to a target context (Refer to Fig. 6, Fig. 2). “Context node” is a kind of slow node to slowly transmit knowledge in the context. It has a region-map, views for a target field and many kinds of sensor data. A context node autonomously plays roles (1) to (5).

(1) Estimating the risks based on information from many kinds of sensors,
(2) Gathering supplemental information and solutions for the newest environments
(3) Merging a target context and the supplemental information into a new context as a part of solution
(4) Transmitting a merged solution with a schedule to the ARN
(5) Forgetting useless knowledge in few access

ARN can quickly access the knowledge in text nodes. The ARN transmits the merged solution to all users in the region as a "Network Affordance" according to the defined schedule. This architecture reduces the communication outside of the region and allows users to quickly acquire evolved knowledge even in emergency time. It also saves the energy consumed by the outside servers.

A context node can easily be added corresponding to a context in a new focus.

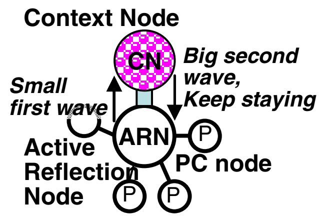


Fig. 6. An image of roles by context node and active reflection node

Example 4-2. TSUNAMI People forget the risk of tsunami and come home in case Example 4 since they feel easy in case that the first tsunami was small. A bigger wave by the roll back of tsunami also happens in case of a strong earthquake like the Indian earthquake. Success network should distribute a message "Both second tsunami and third tsunami are bigger than the first one (refer to Fig. 6). Everybody should stay in the hill area for 50 minutes" within 5 minutes after the first tsunami in the target regions.

The context node can merge actions with messages for a problem and ARN should timely transmit the merged solution to people all over to every corner of a region step by step in certain long time period according to the change of environments. People can easily access the relevant knowledge by choosing an emergency button in a mobile terminal, an Internet TV and a digital signage.

The ARN and context node architecture manage solutions in a long time period corresponding to the local characteristic in a region. The architecture should have four kinds of long time period transmission. (1) Emergency oriented transmission, (2) Sign oriented transmission, (3) Sensor oriented transmission, and (4) Due date oriented transmission.

5 Conclusions

We propose "success network" based on regional contexts for solution services that has action-lists for people all over to every corner of society in order to timely acquire evolved knowledge and to take adequate actions for emergency etc. in certain long time period. We also discuss evolved knowledge that can cope with the change in environment detected by either sensor networks or a few kind of human with deep knowledge. Such evolved knowledge plays an important role of network affordance.

"Context node and active reflection node" architecture in order to effectively transmit evolved regional knowledge to people all over to every corner in regional society is also discussed. The architecture needs some experiments in order to evaluate the efficiency in the Internet even in emergency.

References

[1] Hall, E.T.: Beyond Culture, Anchor Books, A Division of Random House, Inc. (1977)

[2] Gupta, A., et al.: Beyond ”culture”: Space, Identity, and the Politics of Difference. Cultural Anthropology 7(1), 6–23 (1992)

[3] Nakatsu, R.: Future of Communications Empowered by Technologies. Ohm-sha Publishing Co. (2010)

[4] Jung, C.: Archetypes and the Collective Unconscious (1959)

[5] Villazón, A.: A Reflective Active Network Node. In: Yasuda, H. (ed.) IWAN 2000. LNCS, vol. 1942, p. 87. Springer, Heidelberg (2000)

[6] Hughes, D., et al.: A Framework for Developing Reflective and Dynamic P2P Networks (RaDP2P). In: Proc. of the Fourth International Conference on Peer-to-Peer Computing, P2P 2004 (2004)

[7] Boyd, R., et al.: Not by Genes Alone: How Culture Transformed Human Evolution. Chicago University Press (2005)

[8] Miyachi, T., Balvig, J.J., Kisada, W., Hayakawa, K., Suzuki, T.: A Quiet Navigation for Safe Crosswalk by Ultrasonic Beams. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part III. LNCS (LNAI), vol. 4694, pp. 1049–1057. Springer, Heidelberg (2007)

[9] Miyachi, T., Balvig, J.J., Kuroda, I., Suzuki, T.: Structuring Spatial Knowledge and Fail-Safe Expression for Directive Voice Navigation. In: Lovrek, I., Howlett, R.J., Jain, L.C. (eds.) KES 2008, Part III. LNCS (LNAI), vol. 5179, pp. 664–672. Springer, Heidelberg (2008)

A Method to Analyze Learners' Interest Changes with Concept Maps

Masae Nakazawa and Mitsuru Ikeda

School of Knowledge Science, JAIST,
1-1 Asahidai Nomi, Ishikawa, Japan
{nakazawa,ikeda}@jaist.ac.jp

Abstract. Learners' interests may be changed through learning activities in their courses. We focus on changes in their interests; we think that this information is useful to monitor learner interests, and modify the lecturers' instructional strategies. Our aim is to establish ontological approaches to reveal these changes. In this paper, we propose a way to collect concept maps and analyze learner interests, and show results of the experiments.

Keywords: Ontology, concept mapping, analysis method, learner interests.

1 Introduction

One of important experiences in higher education is a communication between students and their supervisors concerning students' research ideas. Throughout the cognitive apprenticeship starting with those communications, students learn their supervisors' acuity of vision (e.g. good judgment about what is important, his/her ability to find the problem domain which will become important in the future), attitude as a researcher, and exacting evaluation criteria. Through the communication learners refine their own interests.

On the other hand, learners' interests may be changed through learning activities. The changes may be caused by either positive or negative feedback from their past learning activities. One of the most important aspects to evaluate learning activities is to clarify how the learning experiences affect the learners' interests or how well-motivated and well-oriented their learning is.

We focused particularly on lecture attendance; lectures were one of the learning activities. In addition, we closely observed the data about how learner interests changed after the lectures, which lectures affected learners, and how the changes and the lecture were related. There are two reasons for focusing on these questions. The first reason is that the data are helpful for lecturers to adjust their instructional strategies. For instance, using this information shows how lectures they presented affect learners. The second reason is that the data are useful for learners to control and monitor their learning strategies. Learners can obtain this information, to distinguish which lecture is better for own interest refinements. However, although methods for describing interests were known, we

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 144–154, 2011.

did not have a method for objectively analyzing interest changes. Therefore, we aimed to establish a technique to clarify learner interest changes.

The method we developed is based on concept maps and ontological engineering studies [1]. Concepts maps, as defined by Novak, are graphical tools for organizing and representing knowledge[2]. In this study, we introduced concept maps as a medium to express learner interests and lecturer intentions. Methodologies for the analysis of concept maps have been proposed in various ways using the bi-directional approaches of quantity and quality (e.g., [3]; [4]). We wanted a new approach, with a qualitative method of analysis, to render the details of each anecdotic data item in a concrete form. We thus constructed an ontology as the foundation for analysis. In this paper, we will explain the analysis method we have been developing, showing the development process of our method.

At the current stage of our research, we have collected most data in the form of hand written paper sheets, and applied the analysis method to the data semi-manually. Our ultimate goal is to build an IT environment for instructors to analyze learners' interests.

2 Outline of Our Approach

In this section, we briefly mention how we approach learner interest changes. In the method, we use the three kinds of concept maps shown in Table 1. The difference between (1) and (2) represents the change in learner interests. The goal of this research is to enable analysis of the potential relation between (3) and the change (See Figure 1). In this paper, we use the terms, "learner interest", "concept map" and "ontology", in the following senses.

Learner interest: a learner's interest that is specific or not. If learner interest is specified, it is concretized as a form of purpose. e.g. research purpose "I want to make theory."

Concept map: a diagram expressing learner interests and lecturer intentions

Ontology: the system of concepts, which becomes a criterion for comparing concept maps

The method we established combines a method to collect concept maps (in Section 4.3) and a method to interpret the concept maps (in Section 5.1).

At the first, we developed a method to collect concept maps as analyzable data, and we made a method to interpret the three points, (1) central-content

Table 1. The Three Kinds of Concept Maps

Concept Map	Explanation
(1) Learner interests before lectures	Each learner is requested to express her interest as a concept map before lecture
(2) Learner interests after lectures	When lecture completed, each learner is requested to express her interest as a concept map
(3) Lecture contents	Each Lecturer is requested to express instructional goal of her lecture as a concept map

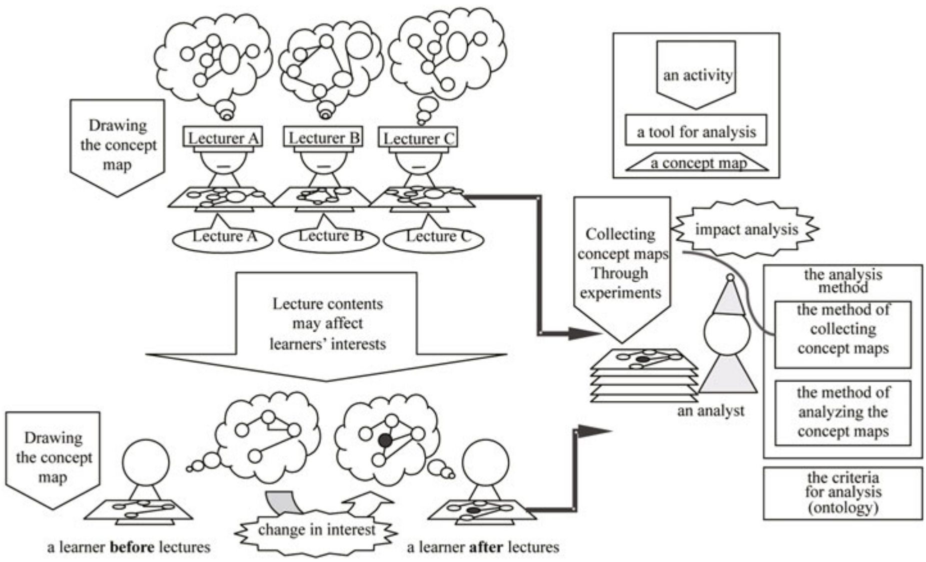


Fig. 1. An Overview of the Analysis Method

in learner interest, (2) the level of concreteness of their interest (hereinafter, interest-concreteness), and (3) reconciliation between a learner's interest and a lecturer's intention, using the information obtained from the concept map data and the ontology we developed. We conducted experiments (Section 4.2), and we referred to the experimental data to establish the method. We describe how the ontology and concept maps perform in the method, in the next section.

3 Concept Mapping and Ontology

3.1 Concept Map in This Study

Concept maps are intended to represent meaningful relationships between concepts in the form of propositions. Employing concept maps can reinforce learner understanding and learning, because creating a concept map allows one to visualize key concepts and summarize their relations to other concepts [5]. The concept maps in our studies are expressive diagrams of lecturer intentions and learner interests. It is possible to comprehend learner interests before/after a learning activity, and lecturer intentions for lectures they presented.

3.2 The Ontology in This Study

Ontologies are defined as an explicit specification of a conceptualization [6]. Constructing an ontology enables consistent relationships between concepts based on consideration. An ontology is a system of concepts to analyze concept maps in our study. An analysis based on ontology has the following two aspects:

1. The first aspect is to enable the clarification of the central-content of concept maps (links of the most important keywords in a concept map) obtained from learners/lecturers.
2. The second aspect is to enable comparing the different levels of abstraction of keywords between learners and lecturers in concept maps, e.g., a link between communication and support, and a link between informal communication and support.

Furthermore, these aspects were useful in making the design of concept maps much simpler. If we can regulate the abstraction of keywords using ontology, we do not have to think about how to regulate keywords on concept maps, and if we can consider the meaning of links on concept maps using ontology, we do not need both learners and lecturers to specify the meaning of each link.

The ontology we made is divided into two parts by two major features, shown below:

1. **Ontology about purposes (the Purpose-Concept Ontology)**
   This works as the criterion for analysis of the central-contents of learner interests and lecturer intentions. This is a system of concepts about their purposes, such as wanting to know about something, or wanting to make a theory of something.
2. **Ontology about the domain (the Domain-Concept Ontology)**
   This works as the criterion for analysis to coordinate the degree of abstractness of keywords, and to consider the meaning of links on concept maps learners and lecturers have made. This is a system about concepts on theories, methodologies, and objects of research in the academic domains they belong to, e.g., SECI-model, which is one knowledge-creating-method.

4 Study Design and Results of Experiments

4.1 Study Subjects

Subjects were selected from master course students at the School of Knowledge Science, JAIST (Japan Advanced Institute of Science and Technology). Knowledge science is a cross-disciplinary research field between information science and social science. We expected that these learners' interests would change much more than those of learners from other fields, because the faculty often face difficulties in following up on the students' research interests changes.

4.2 Outline of Experiments

We carried out experiments as follows:

1. **Preliminary-experiment**
   The purpose is to find a better way to gather concept maps about learner interests and lecturer intentions.

2. **Main experiments**
   The purpose is to gather data to establish a method of concept map interpretation for analyzing the change of learner interests.
   (a) **Experiment with learners**
   The purpose is to collect data of learner interests before learning activities (i.e., before attending lectures in Quarter 1-2).
   (b) **Experiment with learners**
   The purpose is to collect data of learner interests after learning activities (i.e., after attending lectures in Quarter 1-2).
   (c) **Experiment with lecturers**
   The purpose is to investigate teaching intentions for the lectures given.

The details of the experiments are shown in Table 2.

Table 2. Details of the Experiments

Type of experiments			**Study subject**	**# of people**	**Duration of experiment**
Pre			M2	7	25 to 80 min/person
	Learners	Before	M1 before learning activities	31	
Main	Learners	After	M1 before learning activities	31	30 to 60 min/person
	Lecturers	—	Lecturers of 1-2 lectures	12	Avg. 35 min

∗ M stands for master course, e.g. M1(M2) means 1st(2nd) year students in Master Course program.

4.3 Procedures of Main Experiments

We set procedures of the main experiments, referring to the results of preliminary-experiment as follows:

1. We asked subjects (learners and lecturers) to write 10 keywords about knowledge science. Especially, in the experiments with lecturers, the emphasis was to focus on what they expressed to learners through their lectures.
2. We asked the subjects to evaluate the importance of the 10 keywords written in Step 1, on a scale of 1-10. (10 being most important.)
3. We asked subjects to write about what they wanted to accomplish at the School of Knowledge Science in the form of short sentences.
4. We then gave a work sheet for making a concept map to each subject. We had prepared four keywords, (theory, tool (methodology), object, and phenomenon) on that sheet. We asked subjects to write the 10 keywords they wrote in Step 1, on the sheet, and to link the keywords that they thought were related (including the four new keywords). We recommended to the subjects that the number of links be five or less for each word.
5. Finally, we asked subjects to add 10 new keywords, then set and link the keywords.

We collected the data on importance in Step 2 to discuss the procedures for pre-editing (see 5.1.1). At the same time, we gathered the data on short sentences in Step3 to establish a procedure of analyzing learner interest changes (see 5.1.2).

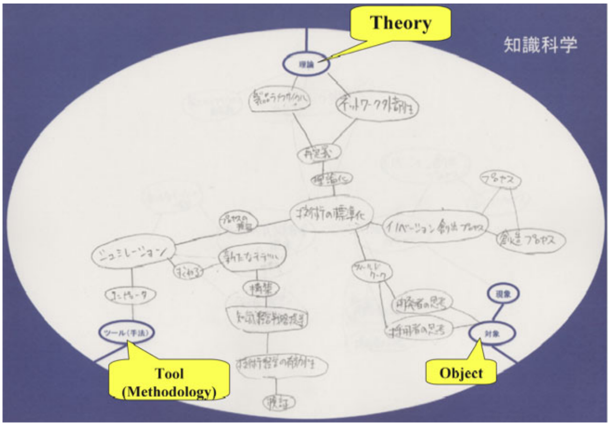


Fig. 2. An Example of Concept Map (Raw Data)

4.4 Results of Experiments

Figures 3 and 4 shows examples of the data collected by experiments using the procedures described in the previous section. A study subject drew this map after taking coursework. He effectively used the four nodes we prepared. The nodes of ergonomics and the decision-making processes directory linked to theory are deeply related to the lecture he attended.

Thus that experimental procedure gives us valuable information to consider the relation between learner interest and lecture contents. However, we need data to interpret data to get the information from concept maps.

How can we interpret this data? How can we reduce interpretation-variation? We make a method for data interpretation to answer these questions.

5 Our Analysis Method

In this section, we introduce the analysis method we established, and an instance of the application of the method. We review the information obtained using the method.

5.1 Procedure of Analysis

We set the analysis procedures as follows:

1. To gather the concept maps
2. To pre-edit the concept maps
3. To analyze the changes of learner interests based on ontology
4. To analyze influences of lectures on learner interests, based on ontology

As Step 1 has already been mentioned, in this section we describe Steps 2, 3, and 4.

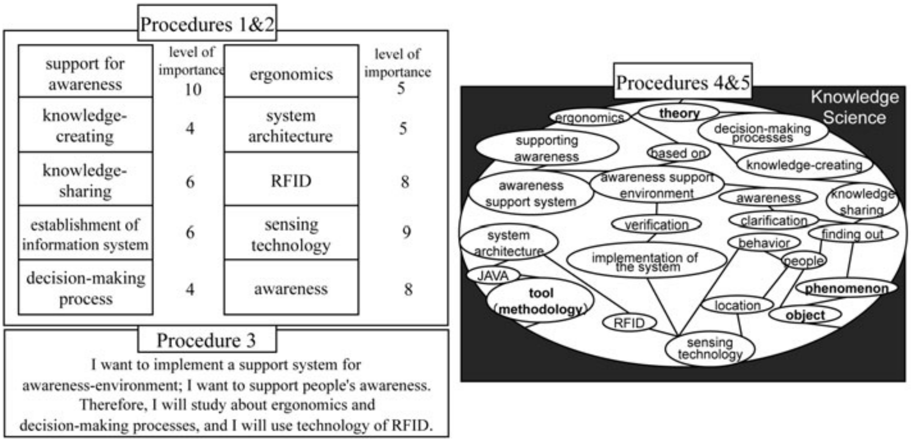


Fig. 3. An Example of Experimental Results

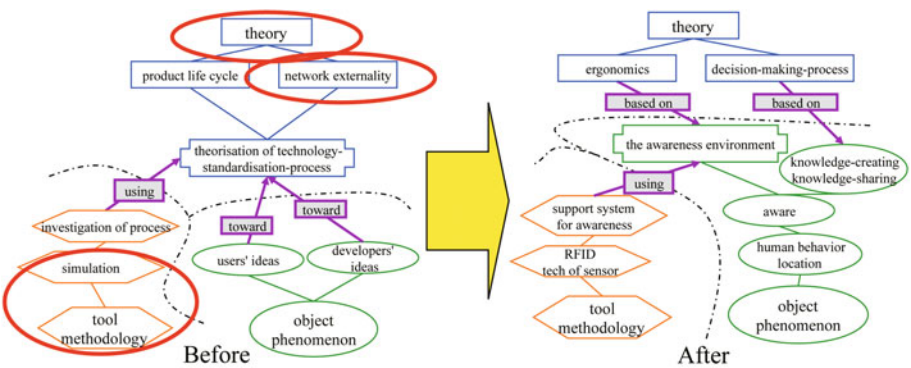


Fig. 4. S27's Interests before and after lectures

Pre-editing. Pre-editing highlights the traits of concept maps applied to the rules that we set. For example, we rearrange the nodes on concept maps to simplify links, and categorize them based on ontology into three domains, i.e., theory, methodology and object. Furthermore, we assigned specific meanings to borderlines among the categories, since such borderlines are likely to evolve into important links, which each have unique meanings requiring specific relations among the categories. Hereinafter, those borderlines assign specific meanings are called Highlighted Links. This step, based on ontology, enables us to apply the next step to concept maps. We show Figure 5 as data after the step. In Figure 5, each keyword clearly belongs to one of three domains, such as "simulation is a method" or "network externality belongs to theory."

At Step 3 of analysis, we compare the learners' concept maps before and after lectures obtained from the pre-editing shown in Section 5.1.1, as shown below:

(A)	(B)	(C)
purpose: to make a theory (research)	purpose: to make an environment (research)	purpose: to make an information system (research)
method: review of the literature	method: review of the literature	method: review of the literature
method: searching for an object	method: study about the env.	method: study about the info-system
method: preparing to make a theory	method: preparing tools to make env.	method: preparing skills and tools
method: modeling	method: constructing the env.	method: constructing the info-system
method: validating of the model	method: validating the env.	method: validating the info-system
method: summarizing the results	method: summarizing the results	method: summarizing the results

Fig. 5. The Purpose and Methods Relation Table

1. We view the output of pre-edit, and consider the changes of learner interests based on the Purpose-Concept Ontology and the Domain-Concept Ontology.
2. We look at the concreteness of learner interests by using the "Purpose and Methods Relation Table." The Purpose and Methods Relation Table summarizes relations, such as, if we want to accomplish A, we have to obtain knowledge about B-technology. Figure 6 shows a part of this table. We can seriously consider details of changes by using this table; it shows how much methods change depending on purpose.

Through this step, we clarify the features of learner interests, the level of their interest-concreteness, and the relation central-content in their interest and each keyword. We compare the features of the data before and after attending lectures.

Getting information through this step, enables learners to understand the degree of their interest-concreteness, and their lack of knowledge.

We describe a specific embodiment of this step, in Section 5.2.

Analyzing Changes as Influenced by Lectures. In Step 4 of this procedure, we focus on the organization of Highlighted Links of the learner maps. The links are highlighted by the pre-editing, expressing central-content in learner interests.

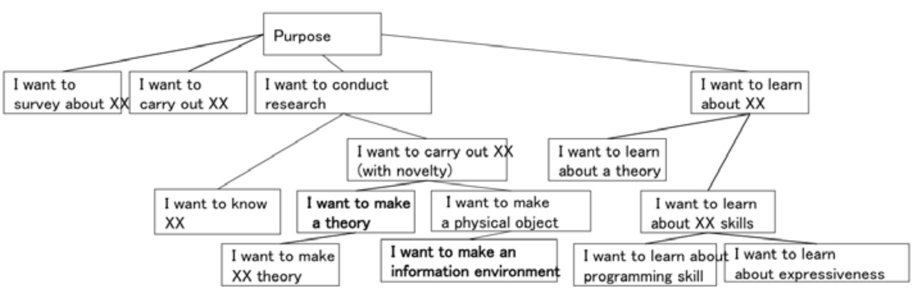


Fig. 6. A Part of the Purpose-Concept Ontology

Additionally, we view the organization of Highlighted Links of the lecturers' maps. We look for indirect or direct relationships between learner and lecturer organization of Highlighted Links (for example, the indirect matching through is-a relation). Output of this step is helpful for lecturers, because this information shows how a learner incorporates lecture contents into his/her interests.

We describe a specific embodiment of this analytic procedure, in Section 5.2.

5.2 The Case of S27: Interest Changed from Economics to Computer Science

Analysis of the Change in Learner's Interest. Figure 5 is the data obtained from the pre-edit of (before and after lectures) concept maps from a learner (Subject 27, hereinafter S27). We analyzed changes using the Purpose-Concept Ontology (Figure 7). Although the actual ontology has more detailed concepts and attributions, we show some essential parts of the ontology for analysis here to keep the viewing understandable.

The central-content of S27's interest before taking the coursework is "I want to make a theory," as shown in Fsigure 7. The purpose "I want to make a theory" was included in "I want to conduct research", especially "I want to carry out XX." (Figure 7, in this figure, lines express "is-a" relations).

Learner research-purposes are made up of constructive-concepts, including theory to be used as the basis, methodology to be applied, region of interest for research, and originality. For example, Figure 8 shows the instance of the concept "I want to make a theory." That instance materialized to suit the purpose of S27. In this way, we can clarify learner's purposes by concretizing purpose-concepts using domain-concepts.

Moreover, we can reveal the techniques and knowledge learners should learn, by matching methods to purpose. For example, we can see that the degree of change is greater when a learner switches from the purpose of "I want to make a theory" (Figure 6(A)) to "I want to make an environment" (Figure 6(B)), rather than when the learner changes from the purpose "I want to make an information

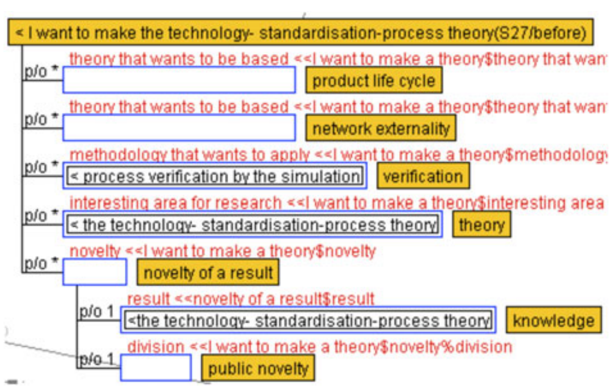


Fig. 7. The Instance of "I want to make a theory"

system" (Figure 6(C)) to "I want to make an environment", by using the Purpose and Methods Relation Table. This is because the former changes are dramatic and involve almost all constructive-concepts except review of the literature and summarizing the results, while the latter change is much less dramatic.

Using this information, learners can understand not only their own change in purpose, but also new things that must be done for that purpose.

Analysis of Influence of Lectures. We refer chiefly to the Domain-Concept Ontology, in this step. Using the ontology that summarizes the keywords in an academic domain, we are able to comprehend by matching at different abstraction levels, as well as simple matching of keywords. In the case of S27, he wanted to apply theories of ergonomics or the theory of decision-making processes to knowledge-creating or awareness, after the lectures. (See Figure 5.) The ontology shows that the decision-making processes is a cognitive theory. ((A) in Figure 9.) At the same time, knowledge-creating and awareness are kinds of individual thinking processes. There are varying models of the definition of individual thinking processes. In the Wallas' model, there are four stages in the thinking process, with the 3rd stage as illumination. (See (B) in Figure 9.) The idea of illumination in Wallas' formula is a synonym for awareness obtained from S27. The decision of S27 to apply decision-making processes theory to knowledge-creating and awareness may have been influenced by the lecturer's intention,

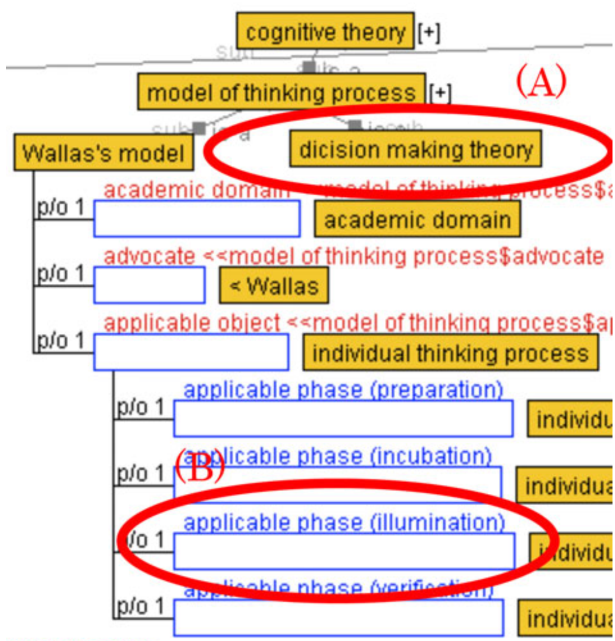


Fig. 8. A Part of the Domain-Concept Ontology

which was to apply the theories of cognitive science to human behavior, such as thinking or knowledge-creating.

This information is useful for lecturers. For example, when there are a lot of learners who want to make an information environment, and who are influenced by the lecturers' intention, which is to apply knowledge-creating to cognitive theories, e.g., S27 in Figure 5, the lecturer may think "I will tell not only about the analysis of the knowledge-creating process based on cognitive theory, but also about examples of implementations for computer environment of the analysis."

6 Conclusion

In this paper, we introduced an analysis method using concept maps based on an ontological engineering approach.

Through establishing this technique, we got some findings, summarized below,

1. We can get data about learner interests and lecturer intentions, using concept maps.
2. We can compare and interpret the central-content of learner interests and lecturer intentions in concept maps, using the Purpose-Concept Ontology.
3. We can analyze the details of learner interests more than the content in concept maps, using the ontology we developed, and the "Purpose and Methods Relation Table."
4. We can coordinate the degree of abstractness of keywords, thus we can see indirect matching of the links between learners and lecturers.

If we provide feedback on this information to lecturers and learners, they can refine their lecture contents or interests. The future task is to improve this procedure through more analysis of learners' concept maps. We are aiming at the construction of a computer environment that is able to replicate the analysis results obtained by using this analysis method.

References

1. Mizoguchi, R.: Tutorial on ontological engineering - Part 1: Introduction to Ontological Engineering. New Generation Computing 21(4), 365–384 (2003)
2. Novak, J.D., Gowin, D.B.: Learning how to learn. Cambridge University Press, New York (2006)
3. Novak, J.D., Canas, A.J.: The theory underlying concept maps and how to construct them. Technical Report IHMC CmapTools 2006-01, Florida Institute for Human and Machine Cognition (2006)
4. Mavers, D., Somekh, B., Restorick, J.: Interpreting the externalized images of pupils' conceptions of ICT: Methods for the analysis of concept maps. Computers and Education 38(1-3), 187–207 (2002)
5. Somekh, B., Mavers, D.: Mapping learning potential: Students' conceptions of ICT in their world. Assessment in Education 10(3), 409–420 (1980)
6. Gruber, T.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisitions 5(2), 199–220 (1993)

A Naïve-Bayesian Methodology to Classify Echo Cardiographic Images through SQL

S. Nandagopalan[1], B.S. Adiga[2], T.S.B. Sudarshan[3],
C. Dhanalakshmi[4], and C.N. Manjunath[5]

[1] Department of Computer Science & Engineering, Amrita Vishwa Vidyapeetham, Amrita School of Engineering, Bangalore, India
[2] Tata Consultancy Services, Parallel Computing Division, Bangalore, India
[3] Dept. of Computer Science & Engineering, Amrita Vishwa Vidyapeetham, Amrita School of Engineering, Bangalore, India
[4] Department of Echocardiography, Sri Jayadeva Institute of Cardiovascular Sciences and Research, Bangalore, India
[5] Sri Jayadeva Institute of Cardiovascular Sciences and Research, Bangalore, India
{snandagopalan,adiga.bs}@gmail.com
tsb_sudarshan@blr.amrita.edu,dhanalakshmi_chandrasekaran@yahoo.co.in

Abstract. Generally patient data in healthcare environments exist in relational databases. Classification of echocardiographic images is an important data mining task that helps hospitals without transferring the data in any form. This paper proposes a novel method to accomplish this task using naïve Bayesian model via SQL. The proposed method has two phases. The first phase builds a knowledge base using many normal and abnormal subjects and the second phase uses this knowledge to categorize an unseen subject into appropriate class. The image features such as cardiac chamber dimensions (specifically Left Ventricle - LV), Ejection Fraction, Mitral valve (MV) orifice area, etc., are computed by first segmenting the image by employing advanced image processing techniques. For instance, to segment echo images we employ an efficient SQL based Fast K-Means algorithm combined with a greedy active contour algorithm for accurate boundary detection. Additional features such as textural, statistical, and histogram are computed and added to the classifier model by analyzing color Doppler echo images to strengthen the classifier accuracy. Our SQL based naïve Bayesian classifier model is built with 7 schema, simple yet efficient SQL queries and thus providing an accurate classification of patients as normal or abnormal. The model is trained with 112 patient data and we believe that the clinical decision is simplified and can happen on-the-fly. Experimental results presented in this paper show an increased accuracy of 87.48% against the other state-of-the art segmentation and classification methods reported.

Keywords: Echocardiographic image, Naïve-Bayseian Classification, K-Means, Segmentation, SQL, Active Contour.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 155–165, 2011.

1 Introduction

Most data in real enterprises exists in multiple relational databases each with well designed schema. Classification being one of the most popular predictive data mining tasks, we need practical algorithms that can directly be applied on existing medical databases [9]. Echocardiography is one of the popular techniques in human heart diagnosis for the study of heart abnormalities. Classification lies at the core of many knowledge discovery and data mining (KDD) applications whose success depends critically on the quality of the classifier [10]. Bayesian classifiers stand out for their robustness, interpretability, and accuracy [1]. It is a useful tool to support the assessment of individual risk of relapse or progression in patients diagnosed with brain tumor undergoing RT postoperatively [13].

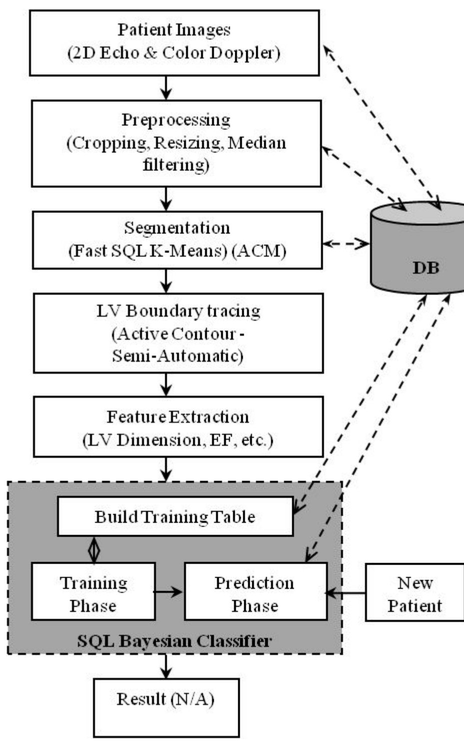


Fig. 1. Block diagram of the proposed system

Our work mainly focuses on classifying 2D and Doppler echo image modalities into normal or abnormal based on several domain specific attributes [3] [6] computed specifically for this purpose. There are two phases involved in our work. The first phase is to analyze the echo images by applying Fast K-Means SQL algorithm followed by active contour ("*snake*") for segmenting the various objects (Left Ventricle, Right Ventricle, etc) [2] [4]. This quantification in turn helps us to extract relevant features such as LV diameter, LV height, LV area, LV volume, etc. A similar set of features are computed for other cardiac chambers: Right Ventricle, Left Atrium, and Right Atrium. These features constitute the so called feature vector for classification. The said quantifications are applicable for 2D echo images and on the other hand color Doppler images require a different type of analysis and feature extraction.

Color Doppler imaging is mainly used as a qualitative tool for diagnostic assessment of vessel stenosis or occlusion [3]. Sonographers and physicians currently rely on their expertise to evaluate and interpret the images. We believe that the introduction of image understanding techniques can simplify and standardize the analysis of the images. To achieve the same, a new approach is employed. This is called as pixel classification method which is simple, fast, and accurate [6]. After identifying the color segment in the Doppler image, the qualitative features that are clinically useful for classification such as energy, entropy, contrast, skewness, kurtosis, etc., are computed. Figure 1 shows the outline of the proposed classifier model.

In the second phase we compute the posterior and priori probabilities to build the training database on the naïve Bayesian theorem. This is accomplished with the consolidated feature vector (computed in the first phase), a set of DBMS tables, and a series of SQL statements. With this training table, it is easy to classify any unseen patient data with utmost one join.

The rest of the paper is organized as follows: Section 2 presents the earlier works done in this area. Section 3 describes various methods to extract qualitative and quantitative features of 2D and color Doppler images. Section 4 presents our proposed Bayesian model explaining schema design, algorithms, and SQL queries. The experimental set up and results (classifier accuracy) are covered in Section 5 and conclusions with improvements are discussed in Section 6.

2 Previous Work

K-Means algorithm was used for biomedical image segmentation using adaptive techniques and morphological operations by Jierong Cheng, et al [2]. Three algorithms were proposed by Carlos Ordonez using DBMS SQL and C++ and demonstrated how K-Means can be of practical importance for clustering large data sets [4]. However, his work was more generalized and less efficient than our method. In our work, the basic concept of SQL based implementation is adopted. Moreover it is customized for echo images and its segmentation with careful selection of k value.

For LV boundary tracing, automatic and semi-automatic methods were presented by Jierong Cheng [2]. These authors use Watershed algorithm for segmentation. One of the main drawbacks of this method is over segmentation. For echo image segmentation, we need to identify precise segments, say LV, RV, LA, RA, etc., and not the other areas, because each of these objects will later be used for quantification purpose. Any error caused will have direct impact on the deviation from the normal values suggested by ASE and [8].

Though many authors have contributed on analysis of 2D echo images, only few authors had thrown light over color Doppler image segmentation. Ashraf A. Saad [3] presented a novel technique for segmenting blood vessels in ultrasound color Doppler images based on shape decomposition. Their method uses color Doppler images as binary images, but in our case we try to accurately extract the color segment. Two authors Carlos and Anne Denton contributed on Bayesian classifiers using SQL and P-tress [7] respectively. However, approaches and the goal of these methods were totally different. For example, in their method the data was first clustered and then the classification was carried out.

3 Feature Extraction of 2D and Color Doppler Echo Images

In the first phase the 2D echo image is median filtered to remove the speckle noise and then segmented using a fast K-Means algorithm of [4]. The traditional K-means algorithm is implemented by setting $k = 3$ (since we are interested in knowing myocardial, endocardial, and epicardial boundaries) and with set of SQL statements that computes Euclidean distance more efficiently. The outcome of this step is an image with just three colors: black, white, and gray. This transformed image has uniform intensities in their respective color regions, thus proving an efficient way to trace the boundaries of LV, RV, etc.

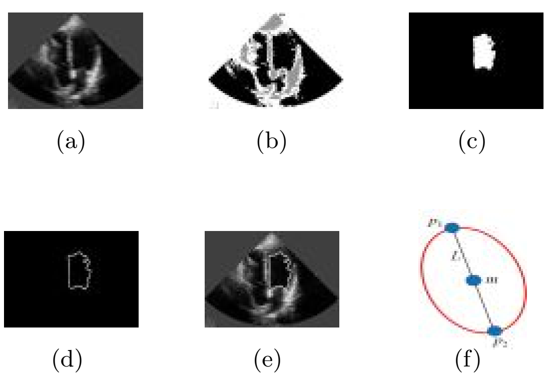


Fig. 2. Image Quantification: (a) original image - four-chamber view (b) segmented image (c) LV boundary (d) Sobel edge detected (e) final image (f) quantification

Next step is to mark the ROI on a particular object, say for example LV, which is marked as the initial contour. This initial contour quickly converges due to uniform pixel value and stops as soon as it reaches the boundary [5] based on active contour model.

$$E_{Snake} = \int_C \alpha(s)E_{cont}(s) + \beta(s)E_{curve} + \delta(s)\,ds \tag{1}$$

This automatic tracing is achieved finding energy minimal as per equation 1. The parameters α, β, and γ are used to balance the relative influence of the three terms. We set $\alpha = 1.5$, $\beta = 1.0$, and $\gamma = 1.5$ to avoid bad tracing. A series of image processing and geometric methods are applied to quantify and obtain the features of cardiac chambers (see Figure 2.)

Next stage involves finding the edges of this object using Sobel edge detector operator (Figure 2(d)). This image is thresholded to remove the noise in the neighboring regions of the edge ($T = 184$). The final image shown in Figure 2(e) is the result of adding 2(a) and 2(d).

The dimensions of LV such as area, diameter, height, volume, EF, etc., are computed by approximating the object as an ellipse [8]. For example, to find the volume of the LV, we use the formula,

$$V = \frac{(0.85 \times A^2)}{L} \tag{2}$$

where A is the area and L is the major axis. Here L is nothing but the distance (d) between $p_1(x_1, y_1)$ and $p_2(x_2, y_2)$, where $p_1 = (\frac{(min\_x+max\_x)}{2}, min\_y)$ and $p_2 = (\frac{(min\_x+max\_x)}{2}, max\_y)$. Using Euclidean space, the distance $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. Since the LV object is traced automatically, the number of pixels inside this object gives the area. Alternatively, we can compute the area by first finding the width, W of the ellipse using $m = \frac{(p_1+p_2)}{2}$ and extrapolating on either directions. Then the formula shown in equation 3 gives the area of LV. However, the first method gives a better result.

$$A = 0.7854 \times L \times W \tag{3}$$

To compute the ejection fraction, we must know the volume of LV at end of diastole and systole. Then ejection fraction is calculated with the formula as shown in equation 4.

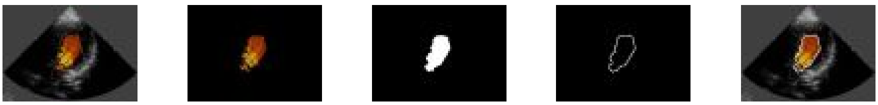
$$EF = \frac{(EDV - ESV)}{EDV} \times 100\% \tag{4}$$

Here, EDV is the volume at end of diastole and ESV is the volume at end of systole. For the other cardiac chambers the same formulae can be applied, because these cavities also resemble ellipse shape.

To extract qualitative features of Color Doppler images we propose two algorithms that could segment the color image efficiently. The first one is called as Color Doppler K-Means (CDKM) algorithm (extension of [6]) which identifies only the color pixels by setting a predetermined threshold value. Next we populate the NACData table from NAData table for further processing. When L = 0, it is a color pixel, else it is grayscale as shown below:

```
Update NAData set(L) = (
   Case  When (R = G and R = B) or (Abs(R-G) <= 16)
         or (Abs(G-B) <= 16) Then 4 Else 0 End);
Insert into NACData (
  Select i,x,y,R,G,B From NAData Where L = 0);
```

The second algorithm is called as Pixel Classification (PC) and is much faster than the previous method as no distance computation is carried out. It scans through the entire image and checks each pixel whether it belongs to color or grayscale based on the threshold logic as explained already. Only pixels that are identified as color are stored in the output image and the rest of the pixels being marked as black as shown in Figure 3. Figure 3(a) shows the original image and Figure 3(b) is after extracting color pixels (note that the rest of the pixels are black).



(a) (b) (c) (d) (e)

Fig. 3. Color Doppler image processing (a) original image (b) segmented image (c) binary image (d) Sobel edge (e) final image

Next to trace the boundary of this color segment, we automatically set the initial contour points at four corners of the image. After approximately 150 iterations, the contour settles on the boundary of the color segment. This happens very fast because of the uniform gradient from the initial contour position up to the boundary of the segment. This object, now, is analyzed for computing quantitative measurements and various textural, histogram, and statistical features [6]. In case if the color segment is present in more than one location of the image, the largest color segment is considered for analysis.

4 SQL Based Naïve Bayesian Classifier

Baye's Theorem provides a mathematical method that could be used to calculate, given occurrences in prior trials, the likelihood of a target occurrence in future trials. According to Bayesian logic, the only way to quantify a situation with an uncertain outcome is through determining its probability [13]. A simple Bayesian classifier, naïve Bayesian classifier, has comparable performance with decision tree and selected neural network classifiers [11].The main focus of our work is to deploy Bayesian classifier in database tables and SQL queries. This involves two phases: training phase and prediction phase of any new patent.

Let D be a training set of tuples with n samples. Each tuple has three categories (c) 'L''/''C''/''H', and m classes $C_1, C_2, ..., C_m$ as their associated class labels. Assuming l attributes for each tuple and $X = (x_1, x_2, ..., x_l)$, the aim of classification is to derive the maximum posteriori, i.e., the maximal $P(C_i \mid X)$. This can be derived from Bayes' theorem as shown in equation 5.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{5}$$

Assuming that the attributes are conditionally independent (that is, P(X) is constant), we have

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times ... \times P(x_l|C_i) \qquad (6)$$

For our problem, only the following attributes are considered: LVLD (LV Length Diastole), LVWD (LV Width Diastole), LVVD (LV Volume Diastole), LVVS (LV Volume Systole), EF (Ejection Fraction), LVAD (LV Area Diastole), LVAS (LV Area Systole), and BC (class label.) The normal values for these parameters are taken from [8] and discretized as C/L/H meaning 'Correct', 'Low', and 'High' respectively which signifies the severity of the disease.

4.1 Training and Prediction Phase Schema

In order to implement this phase and compute priori and posterior probabilities we use 7 tables. The detailed schema is as follows: EDT (PID, LVLD, LVWD, LVVD, LVVS, EF, LVAD, LVAS, BC), ACLASS (PID, CAT), ECLASS (ATTB, CAT), EPC (CLASS, C, PCI), EPX (ATTB, CAT, CLASS, XC), EPXI (ATTB, CAT, CLASS, XC, PR), EFP (CLASS, PR). Table 1 gives a detailed tabular structure along with the dimensions for the proposed model. Here n, l, m, c represent the number of sample images in EDT, number of attributes, number of classes, and number of categories respectively. To follow a generalized design, two tables AClass (number of categories for each attribute) and EClass (each attribute and its associated categories) are populated using relevant SQL queries from EDT.

Table 1. Database Table Structure

Sl. No.	Table name and Attributes	Size
1	EDT	$n\times$ l
2	ACLASS (id, cat)	c
3	ECLASS (attb, cat)	c× l
4	EPC (class, c, pci)	m
5	EPX (attb, cat, class, xc)	$l \times$ c× m
6	EPXi (attb, cat, class, xc, pr)	$l\times$ c× m
7	EFP (class, pr)	m
8	Target	$r\times$ l

Let us describe the purpose of each of these tables. Table EDT contains the feature data of all patient images considered for training with attributes PID, LVLD, LVWD, LVVD, LVVS, EF, LVAD, LVAS, and BC. The table EPC stores class prior probability, whereas EPX stores the count of tuples grouped by each class, category, and class. To store posterior probability and subsequently as training table, we use a separate table EPXi. Thus, this table contains the probabilities for all combinations of categories making it as eager learning model.

With the training table (i.e. knowledge base), the class labels of sample data in the Target table (attributes are same as EDT with BC initialized to 'U') can easily be computed. For this we use EFP table and from this table, we retrieve the class corresponding to the maximum probability and update BC as Normal or Abnormal (i.e., 'N' or 'A') in the Target table. This means all the new patient data in Target table will be classified. Further, this table can be used for measuring the accuracy of the proposed model.

4.2 SQL-Bayesian Algorithms

An important step in our research work is how to write the SQL queries. Considering the DBMS tables shown in the previous section and the Baye's theorem SQL queries are written to compute posterior and class probabilities.

```
Algorithm SQLBayesianClassifier(EDT)
[Training Phase]
Step 1: [Extract attributes from EDT for each c]
    EClass = AClass * EDT
Step 2: [Find class probability P(Ci), where i = 2]
    EPC = EDTgroup(bc)
Step 3: [Initialize EPX probability with 0]
    EPX = EClass * EPC
Step 4: [Update EPX with count of tuples]
    for i = 1 to l do
      UPDATE EPX = EDTi * EPX
Step 5: [Compute P(X | Ci) of all combinations]
    EPXi = EPC * EPX
End.
```

```
Algorithm SQLBayesianPrediction(EPXi, Target)
[Prediction Phase]
Step 1: [Compute probabilities w.r.t X]
    EFP = EPXi * Target
Step 2: [Compute P(X | Ci)
    Update EFP = EFP * EPC
Step 3: [Normalize the probabilities]
    Update EFP = EFP * EFP
End.
```

The algorithms for for training and prediction phases are shown above. Steps 1 to 5 of the first algorithm (*SQLBayesianClassifier*) computes the probabilities for each class of every attribute and stores it in EPXi table. This trained data from EPXi is joined with an unknown sample X (single tuple) from Target table and the result of this gives two probabilities; one for normal and another for abnormal. The larger of these two probabilities is stored in EFP and further normalized to get the final class label. This is shown in the algorithm *SQL-BayesianPrediction*.

4.3 SQL Queries

We show only imperative queries in this section. For each attribute (total number of attributes is 7) there are 3 categories and two class labels. Therefore, the table EPX is updated and contains a total of 42 rows (7*3*2) along with the count values. This data is obtained from EDT table. Next, to compute the posterior probability, we join EPX with EPC. These two queries are shown below:

```
Update EPX e
Set (xc) = (Select NVL(p.xc + x.xc, 0)
From (Select  LVLD, bc, count(*) as xc
   From  EDT Group by lvld,bc) p, EPX x
   Where p.lvld(+) = x.cat and p.bc(+) = x.class and
         x.attb = 'LVLD' and e.cat = x.cat and
         e.class = x.class )Where e.attb ='LVLD';
Insert into EPXi (Select attb,e.cat,e.class,xc,xc/c pr
   From EPC p, EPX e Where e.class = p.class);
```

Now EPXi acts as the trained knowledgebase. To determine the class label of a new sample say X, we first join EPXi and Target and then multiply the probabilities of each attribute under each category. Thus there are two tuple in EFP with class-conditional probabilities corresponding to $P(X \mid N)$ and $P(X \mid A)$. Here, N refers to Normal and A refers to Abnormal. The following query shows this:

```
Insert into EFP ( Select class, exp(sum(ln(pr))) as
                    prob From EPXi e, Target t1
              Where e.attb = t1.attb and e.cat = t1.cat
              Group by class );

Update EFP E
Set (class, pr) =
( Select EFP.class, (EPC.pci * EFP.pr) as pr
  From EFP, EPC Where EPC.class = EFP.class and
       E.class = EFP.class );
```

In order to find the posterior probabilities, $P(N \mid X)$ and $P(A \mid X)$, we join EFP and EPC and multiply the probabilities corresponding to each class as shown above.

5 Experimental Setup and Results

To evaluate the accuracy of our classifier the patient images are acquired using a Philips Envisor C HD ultrasound machine with S4-2 adult probe. The training data include normal and abnormal patient images under four-chamber, short-axis, and color Doppler modalities. A total of 48 normal and 64 abnormal patient images are tested on segmentation, LV tracing, quantification, and classification algorithms. The software tools that we used for the research experiment are Visual Studio 2008, ODAC (Oracle Data Access Components (ODAC) on C#.NET framework and Oracle 10g DBMS running on Dell T3400 Desktop with Intel Core2 Duo Processor @ 2.33GHz and 2GB RAM.

5.1 Estimating Bayesian SQL Accuracy

The extracted data from the echo images are discretized as nominal values using entropy based method. Then these nominal values are stored in EDT table for further processing. We compute the accuracy of the classifier using equations 7, 8, and 9.

$$Sensitivity = \frac{\#\ of\ true\ positives}{\#\ of\ positive\ tuples\ (p)} \tag{7}$$

$$Specificity = \frac{\#\ of\ true\ negatives}{\#\ of\ negative\ tuples\ (n)} \tag{8}$$

$$Accuracy = Sensitivity \times \frac{p}{p+n} + Specificity \times \frac{n}{n+p} \tag{9}$$

Table 2. Results in terms of confusion matrix of the proposed Bayseian SQL classifier

Patient Image	C1 (Normal)	C2 (Abnormal)	Total
C1 (Normal)	43	5	48
C2 (Abnormal)	9	55	64
Sensitivity	89.58 %		
Specificity	85.93 %		
Accuracy	87.48 %		

Holdout method is adopted for accuracy measure. We use a confusion matrix and compute the sensitivity and specificity to arrive at the final accuracy [12]. Table 2 shows the results of the performance indicated in terms of true positives, true negatives, false positives, false negatives, and accuracy. There are 5 normal and 9 abnormal false hits out of 112 cases mainly due to poor image quality and incorrect image acquisition. The accuracy shown is much higher than the results reported in [12] [13].

6 Conclusions

This paper presented a novel approach for building a classifier model using naïve Bayesian model through SQL particularly for echocardiographic images. All the algorithms are based on Oracle SQL as the patient data is normally kept in a database. 2D echo and color Doppler are the two modalities considered for classification. The 2D echo feature extraction is semi-automatic whereas color Doppler is fully automatic. The classifier model is intelligently integrated with other modules to obtain a complete system for cardiovascular environments. The algorithms are tested on live images of patients and the accuracy is computed with standard formulae. The results are encouraging and can further be improved by better image qualities.

Acknowledgement. The authors would like to thank the doctors and technicians of Sri Jayadeva Institute of Cardiovascular Sciences and Research Hospital, Bangalore for their support in carrying out this research especially in obtaining echo images of live patients.

References

1. Ordonez, C.: Bayesian Classifiers Programmed in SQL. IEEE Transactions on Knowledge and Data Engineering 22(1), 139–144 (2010)
2. Cheng, J., Foo, S.W., Krishnan, S.M.: Watershed-Presegmented Snake for Boundary Detection and Tracking of Left Ventricle in Echocardiographic Images. IEEE Transactions on Information technology in Biomedicine 10(2), 412–416 (2006)
3. Saad, A.A., Shapiro, L.G.: Shape decomposition approach for ultrasound color Doppler image segmentation. In: 18th Intl. Conf. on Pattern Recgnition, ICPR 2006 (2006)
4. Nandagopalan, S., Adiga, B.S., Dhanalakshmi, C., Deepak, N.: A Fast K-Means Algorithm for the Segmentation of Echocardiographic Images Using DBMS-SQL. In: Proc. International Conf. IEEE, ICCAE 2010, Singapore, pp. 162–166 (February 2010)
5. Nandagopalan, S., Adiga, B.S., Dhanalakshmi, C., Deepak, N.: Automatic Segmentation and Ventricular Border Detection of 2D Echocardiographic Images Combining K-Means Clustering and Active Contour Model. In: Proc. International Conf. IEEE, ICCNT 2010, Bangkok, pp. 447–451 (April 2010)
6. Nandagopalan, S., Adiga, B.S., Dhanalakshmi, C., Deepak, N.: Color Doppler Echocardiographic Image Analysis via Shape and Texture Features. In: Proc. International Conf. IEEE, ICBBT 2010, China, pp. 139–143 (April 2010)
7. Denton, A., Perrizo, W.: A Kernel-Based Semi-Naive Bayesian Classifier Using P-Trees. In: Fourth SIAM International Conf. on Data Mining, pp. 427–431 (2004)
8. Feigenbaum, H.: Echocardiography. Williams and Wilkis (1993)
9. Manjunath, G., et al.: A Heterogeneous Naive-Bayesian Classifier for Relational Databases. In: KDD, Paris (2009)
10. Fuxman, A., et al.: Improving Classification Accuracy Using Automatically Extracted Training Data. In: International Conf. KDD 2009, Paris (2009)
11. Langley, P.: An Analysis of Bayesian Classifiers. In: Proc. of AAAI 1992 (1992)
12. Zhang, Y., Cheng, H.D., et al.: Classification of breast masses using color Doppler flow imaging. In: 11th Join Conf. on Information (2008)
13. Kazmierskaa, J., Malickib, J.: Application of the Naive Bayesian Classifier to optimize treatment decisions. Radiotherapy and Oncology 86, 211–216 (2008), www.thegreenjournal.com Elsevier Ireland Ltd. (2007)

Acquisition of Practical Knowledge in Medical Services Based on Externalizing Service Task-Grasp of Medical Staff

Taisuke Ogawa[1], Tomoyoshi Yamazaki[1], Mitsuru Ikeda[1], Muneou Suzuki[2], Kenji Araki[2], and Koiti Hasida[3]

[1] School of Knowledge Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai Nomi, Ishikawa, Japan
{t-ogawa,ikeda}@jaist.ac.jp
[2] Medical Information Technology, University of Miyazaki Hospital, Japan
[3] Social Intelligence Technology Research Laboratory, National Institute of Advanced Industrial Science and Technology, Japan

Abstract. It is ideal to provide medical services as patient-oriented. The medical staff share the goal to help patients recover. Toward that goal, each staff members uses practical knowledge to achieve patient-oriented medical services. But each medical staff members have his/her own priorities and sense of values, that are derived from their expertise. And the results (decisions or actions) based on practical knowledge sometimes conflict. The aim of this research is to develop an intelligent system to support externalizing practical wisdom, and sharing this wisdom among medical experts. In this paper, the authors propose a method to model each medical staff member's sense of values as his/her way of task-grasp (or task-perception) in medical service workflow, and to obtain the practical knowledge using ontological models. The method was tested by developing a knowledge-sharing system based on the method and running it at Miyazaki University Hospital.

Keywords: Knowledge Acquisition, Service Science, Ontology.

1 Introduction

Service science is "based on the idea that services are the fundamental activities of human society, we can understand and improve our society through research services and methods to improve services" [1]. In medical service, how to develop services from the viewpoint of patients is quite important. Medical staff have practical knowledge (P.K.) to deal with the patient taking their needs into account in providing services. But it is difficult to support sharing P.K.

The difficulties about sharing P.K. are that P.K. depends on contexts of services, that include patients' feelings, environment (their family and work), sense of values, etc. P.K and its contexts are vague. The aim of this research is to develop a system to support sharing P.K. in medical services.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 166–177, 2011.

In knowledge engineering, to achieve intelligent information systems, the methods for knowledge acquisition were researched (e.g. Protégé [2]). To axquire expert knowledge, it is important to distinguish the knowledge about the problem-solving task itself from all knowledge about problems (This includes domain knowledge). The task knowledge gives a framework to the knowledge interview systems. For instance, The KADS methodology [3] clarifies the problem-solving knowledge based on Newell's knowledge-level principle. Chandrasecaran's Generic Task [4], and Mizoguchi's Task Ontology [5] give essential and abstract task knowledge to knowledge interview systems. The knowledge about tasks gives a base to obtain knowledge about how to solve the problem. These interview systems mainly aim to use computers to solve problems (diagnosis, scheduling, etc.).

In contrast, the author try to improve P.K. usage and sharing by human experts, and give another viewpoint on problem-solving tasks. The viewpoint is the value the expert find in the tasks in the service they provide. The ways of service provision are dependent not only on customer and situation, but also on the experts' sense of values. The authors suppose if we want to understand P.K., we have to understand the sense of values as the background knowledge of P.K. And the P.K. acquisition and sharing system should be designed based on this idea.

In this paper, a method to model medical experts' task-grasp (or task-perception) in medical workflow is explained. The authors think the task-perception model can help to express of sense of values. And through the development of the P.K. interview and shareing system, the feasibility of the idea, that P.K sharing support based on the modeled sense of values of experts, is shown. In the next section, the problems and needs are described. In Section 3, the method and its implementation as a system is explained. In Section 4, the trial use of the system is shown.

2 P.K. Sharing in Clinical Pathway

P.K. is the medical experts' wisdom embedded in medical service activity. Medical staff think about patient feelings and situation, and take them into account and decide how to provide services. P.K. is service-context-dependent. The difficulties in sharing P.K. come from the difficulty in externalizing and organizing the service context. Moreover each medical staff has his/her own sense of values. This influences the priority of services and service provision ways. The sense of value is vague. P.K. is handed down from experts to novices in practical service provision.

In recent years, the clinical pathway (C.P.) is widely used to ensure the baseline quality of medical services. A C.P. defines the optimal sequence and timing of interventions by physicians, nurses, and other staff for a particular diagnosis or procedure [6]. C.P.s are the documents on which the procedures and goals of activities (e.g. examination and treatment) in a hospital are written. On the other hand, there are some critical feelings in hospitals. Medical experts are anxious that "The novice would only perform the medical activities defined on the pathways" or "The training occasions are lost."

But the authors suppose the C.P. give the chance to obtain and to share P.K. in a new way. To externalize the task-grasp (or task-perception) of medical staff members as service task structure models, and to obtain P.K. by interviews generated from the models in path making. And the obtained P.K. is shared among medical staff using C.P. on electronic medical records.

C.P. are only the knowledge about sequences of medical activities. The details of activities are designed by medical staff who execute them on demand. At that time, the medical staff take into account the mental or emotional aspects of patients. P.K. is the knowledge about creative execution of the medical activities defined in C.P for patient-oriented medical services.

C.P. design is done by medical staff teams. The team members, doctors, nurses, pharmacists, etc have their specialities, and C.P. design includes the opportunities to discuss about P.K. and to boost mutual understanding. But there is an impediment to boosting mutual understanding. It is the differences in their sense of values. The differences derive from the different priorities in medical activities, and P.K of the details of activities. For instance, the requirements for physiotherapy by doctors and nurses are different. A doctor may value fast recovery, and a nurse may value comfort. To support the P.K. discussion in C.P. design, the author advance the method to externalize the task-perception of medical staffs as service task structure models, and to obtain P.K. by interviews generated from the models.

3 P.K. Sharing Support System

The system is developed for externalizing medical staff's task-perception in medical services as models based on ontology, and using the models for P.K. acquisition interview. In this section, the authors show the rough sketch of the system at first, and explain the function to model the task-perception and the function to interview medical staff members regarding P.K. using the model.

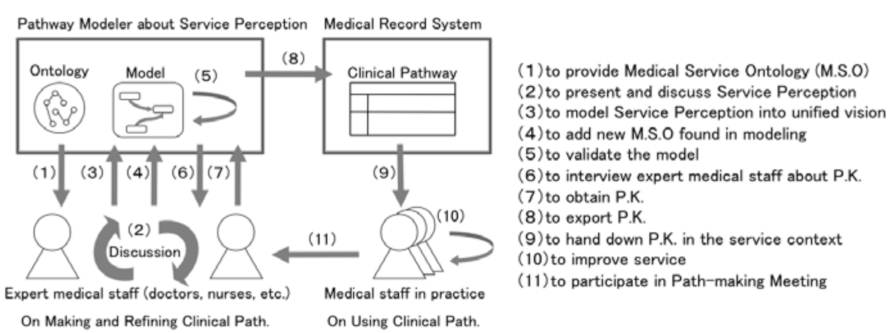


Fig. 1. P.K. Sharing Support System

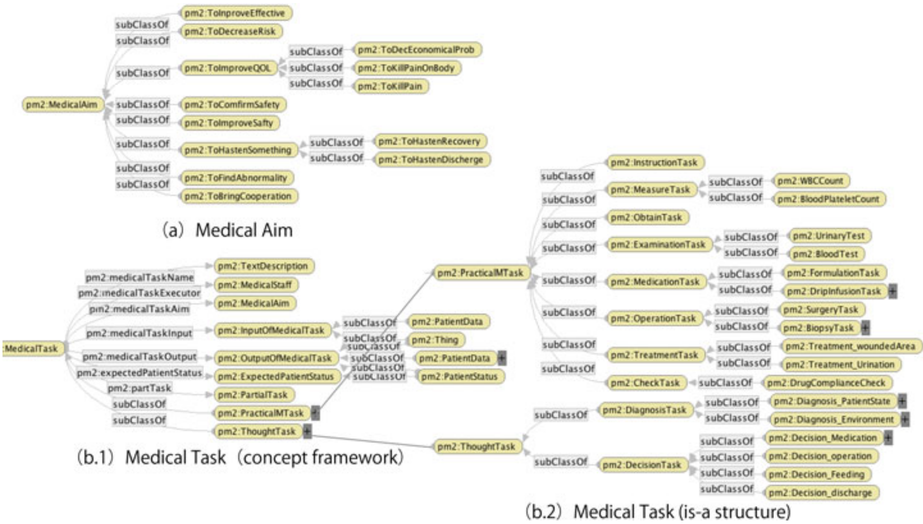


Fig. 2. Medical Service Ontology

3.1 The Future and Aim of Medical Service Ontology

Fig.1 shows how the system supports the activities of sharing P.K. C.P. are designed by expert medical staff, who have different specialities (doctors, nurses, etc.). In general, C.P. defines the typical medical task sequences only, P.K. is not described in C.P. The main functions of this system are to obtain P.K. in C.P. design and to distribute P.K. in medical practice. The phases of support are below.

1. To externalize the task-perception of medical staff in medical service as task models through pathway making.
2. To obtain P.K using the task models.
3. To export P.K. to electronic medical records.
4. To provide medical services referring to and learning and using P.K.
5. To join the pathway-making meeting, and refine the task models and P.K.

To support phase (1) and (2), the authors develop the application software "Pathway Modeler". It is Java-based, and the ontologies are shared on the ontology repository "Semantic Server" [7]. About this support system, the ontological modeling of the task-perception is explained in Section 3.2, and P.K. acquisition is in Section 3.3.

3.2 The Modeling of Medical Service as Task-Grasp of Medical Staff

In recent years, medical knowledge sharing systems are developed based on ontologies, with the spread of the Internet. For instance, in EON [8] project, the patient-oriented medical decision support system, based on ontologized clinical guidelines, is developed. In SAGE project [9] the methodology of modeling clinical guidelines is developed. In these project the medical workflow modeling is researched. C.P. is one expression of the medical workflow. In [10], the concepts included in C.P. are analyzed. An aim of these research studies is to make or support decisions using computers.

On the other hand, the authors consider it is important to support medical staff to design and arrange medical services in a patient-oriented way. The aim of C.P. modeling is to obtain P.K. But the P.K. depends on service contexts. Context-dependent mean not only what people do in certain situations, but what kind of value, priorities or aims they find in services. The purpose of C.P. modeling is to externalize service-perception as each medical staff member's sense of values. The medical service ontology is developed for this purpose.

The difficulties in development of this ontology are the following two points.

- There are no organized vocabulary to explain service task-perception.
- Their service task-perception is externalized when they talk about concrete service scenarios.

We want to recognize and organize vocabulary before the discussion about medical service, but the vocabulary appears through the discussions. Hence, we develop the ontology in parallel with discussions. The ontology development activities are divided into two phases, "Initial Phase" and "Continuing Phase".

In Initial phase, the knowledge engineers define the basic structure of ontology, that guides how to organize the vocabulary. And in Continuing Phase the medical staff explain their service task-perception with using the initial ontology. If there are no ontologies needed for models, they add vocabulary as ontology. As the ontology is developed, the medical staff are not familiar with ontology, so we should take these two points into account.

- tolerance of vagueness and mistakes in ontology definition
- Immediate usefulness fo the ontology

In general, there are some conflicts and confusion in developing dictionaries for sharing knowledge. Because some words have different meanings in different specialized domains. The initial ontology works as the guideline regarding how to organize the new vocabulary and how to use them, for reducing the confusion in C.P. design.

Medical Service Ontology

The authors analyzed the C.P. in the University of Miyazaki Hospital, and obtained the initial ontology. We use Semantic Editor [7] as an ontology development environment.

The medical service task-perception is modeled as below.

- What kind of aim does the medical staff find in each medical task?
- How does the medical staff structure the medical tasks?

Fig.2(a) shows the medical aim ontology. The top level includes "to progress toward a cure", "to improve efficiency of treatment", "to improve QOL (Quality of Life) of patients", etc. More specific concepts are defined as leaf nodes. A generic concept and its specific concept relationships are defined as IS-A relationships.

Fig.2(b.1) shows the concept definition of a medical task. Medical task has an "executor" who executes the task, "input" which is processed, and "output" which include things and data from patients. A medical task has "sub-tasks". And each task and sub-task have more than one "medical aim". Fig.(b.2) shows part of the medical task ontology. It defines a medical task as a class structure. The top level includes "practical task", which brings concrete results, and "thought task" which brings abstract results like construal and decision.

Modeled Service Understanding

Fig.3 shows the model of the C.P. about the task "Let patient walk to toilet". This task is configured after some operations. The aim of the task is "to assess patient's recovery". Before this task, the patient is urinating with the catheter. This task consists of 3 sub-tasks: "assess whether patient can walk", "remove catheter", "let patient walk". The tasks are executed by a doctor, and nurses. The tasks' aims are "to secure safety", "to de-crease risks of infectious diseases", and "to bring the recovery forward". This model explains that medical staff who are in different occupational categories recognize different aims for one task "Let patient walk to toilet", and that medical staff understand one task from different view-points. In the C.P. used in practice, only the task "Let patient walk to toilet" is described, so the medical staff task-perception is not expressed clearly enough. At the construction of initial ontologies, it is difficult to gather these relationships between outcomes (or tasks) and aims comprehensively. In other words, it is difficult to clarify what task-perceptions are held by medical staff members toward tasks. Hence we have to find these relationships and bring them into ontologies through the discussions on C.P. construction.

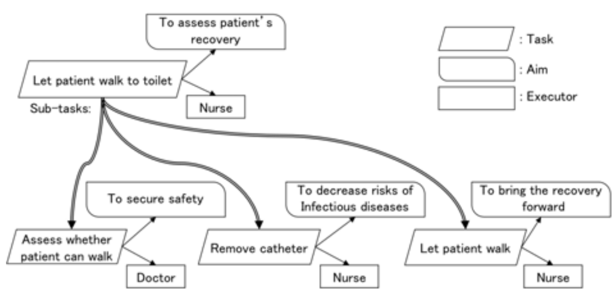


Fig. 3. Model of a task "Let patient walk to toilet"

Table 1. Rule and Template for Interview

Example of Interview	Overview	Rule	Template
Ex.1	Search a task that deals with status of patient' s pain, and get P.K. about how to ask amount of pain.	<task.sub-task>.num≧1,<task.sub-task.output>⊆pain_status	The sub-tasks {<task.sub-task n>, 1..n} are assessment task, and may make patient nervous. In assessment, you give some mental care and some explanation. Please answer in detail.
Ex.2	Search tasks that have part-of relation and each task executer is different, and then get P.K. about how to collaborate.	<task.included-task>.num≧1,< task.included-task.executor >≠<task.executor >	Toward <task.included-task>, <task.executor> needs some collaboration with <task.included-task.executor>, what are the points to remember?
Ex.3	Search a task that participates in more then 2 tasks, each of them has different medical aim. And get P.K to execute the task to trade off or to unify the aims.	<task.included-task>.num≧2, enumerate(<task.included-task.medicalAim>).num≧2	This task participates in multi tasks {"<task.included-task>.n", 1..n} and they have different aims {"enumerate(<task.included-task.medicalAim>).n", 1..n}, To execute this task, what is the key points to trade off or to unify the aims?

3.3 P.K. Interview Based on Models

The P.K. interview function is realized as below. The function infers what kind of P.K exist in each medical task, using rules, and generates interview message using text template (Table 1) based on the modeled service task-perception. The P.K. types are externalized and organized in the group discussions of medical staffs. And the authors (knowledge engineers) develop the rules in considering the P.K. types and the attribute and structures of models. Thus the types and the rules are developed ad hoc at present.

The system and its users (medical staff who make C.P.) interact through the interface (Fig.4).

- The user models their service task-perception on the panel (a) referring to the ontology, which is displayed on the panel (b).
- On the panel (a), the task relations are expressed. And each task's details (attributes) are described in panel (c).
- Next the system checks the model. and detects what kind of P.K. may exist in each task, and generates interview messages Table 2.
- The interview messages are displayed on the panel (d).
- The users answer the interview questions. And the obtained P.K (Table 3) is kept in the system.
- The P.K are exported to the electronic medical record system, and displayed on it.

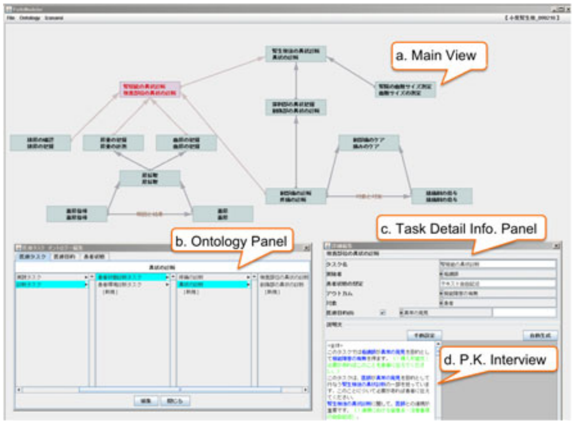


Fig. 4. Interface of "Pathway Modeler"

4 Trial Use of the System

The aim of the trial use is to confirm whether the system will function intended. In the trial, 5 pathways are modeled and P.K is obtained from them.

The criterion of the pathway selection is that the pathway is used over a long time and has almost no needs to modify. Because the aim of modeling is not to refine the contents of the pathways, but to externalize the medical staff members' task-perception about the pathways. The steps of the pathway modeler usage is below.

1. The tasks that must include different task-perceptions are gathered and modeled by some interviews by the doctors in medical information department in the hospital.
2. The problems and dubious points in the task explanation and interview messages generated from the models are obtained by checking them by medical staff in each department in the hospital.
3. The practical knowledge is obtained by answering to the interview messages by the medical staff in each department.
4. The results of the step 2 and 3 are considered by the doctors in medical information department and the authors.

4.1 Result of Modification of Task Explanation Message

As previously mentioned, the role of the task explanation message generated by the system is to show situations about medical tasks to be imagined by the pathway designers.By modifying the task explanation massages, the modeling mistakes and the different task-perceptions between the staff members who have modeled the tasks and other staff members. In this section below, the authors examine the necessity of the task-perception modeling to obtain the P.K.

Table 2. Sample of an Interview Message

Task "watch for signs of abnormality of the kidney" in "kidney biopsy" Path.	Q. No.
For Medical Staff:	
(1) In This task, <Nurse> gets the <abnormality of the kidney> <to find abnormalities in the kidney>.	
(2) This task is a part of the task <diagnosis abnormality of kidney function> by <Doctor>. The aim is <to assure safety>. This task needs the collaboration with <Doctor>.	
(3) This task has sub-tasks "<Check urination>", "<Get amount of urine>", and "<Check bloody urine>".	
(4) The task "Check bloody urine" needs assessment in terms of the aim "to find abnormalities". What are the key points in the assessment?	(Q.1)
(5) In the sub-task "Check bloody urine", you deal with patient's anxiety factor.	
Please be careful about it. What are the key points about the care? [Write the explanation for patient in column (A)]	(Q.2)
Explanation for Patients	
(A) "Check bloody urine" task relates to patient's anxiety factor. How to explain this task to reduce the anxiety?	(Q.3)

Table 3. Sample of Obtained P.K.

Questions	Answers
Q.1	In the case the patient is on bed rest after the exam, the first time urination is reported to the doctor, when the nurse confirms without fail, and there is a gross hematuria.
	In the case the urethra balloon is used, the nurse confirms the tone of urine in measuring the vital sign. And if there is a gross hematuria, contact the doctor.
Q.2	After the patient can move to the rest room. The patient himself or his family confirms whether there is gross hematuria.
Q.3	After the kidney biopsy, the gross hematuria might be caused. But usually improving it by the rest .
	When the gross hematuria continues, it is reported to the physician in charge, the ultrasonography. The blood test preliminarily be prepared and done immediately.

There were one modeling mistake. There is a task "Measurement of the blood tumor size" after the operation of the kidney biopsy. At first, this was defined as a partial task of "Treatment of the wound area" that is done by a nurse. However, about the explanation about this with "Treatment of the wound area", a comment was given. It was "This blood tumor was caused in the kidney. Doctors used the ultrasound echo and measured the size. Then the task is not included." Then, the measurement task is defined as a partial task of "Abnormal diagnosis of the kidney" task is done by doctors. The modeling mistakes were examined toward models by the authors and the doctors in the medical information department, and were reexamined at the P.K. trial use phase (section 5.3). However, there was this mistake only. Hence, it means that the system usage "To consider system user's ontology literacy, the model is examined through the correction of the task explanation messages. " works as intended.

Next, as a declaration of service understanding different from the model, a comment for task instruction to match the medical tasks and aims modification to their usual work level understanding was obtained.For instance, about "Grasp of the volume of urine" task of TUR-BT (transurethral resection of the bladder tumor) pathway, there was instruction to correct its inclusive task "Grasp of abnormality of the bladder function" to "Grasp of urinating abnormality", and about "Drinking water guidance" task, there was instruction to correct its aim "To decrease of the infectious disease" to "To prevent the catheter blockage". These instructions show the medical staff members' task understanding about the medical knowledge that "An abnormality of the bladder function is detected form the amount of urination", and "(Because the risk of the infectious disease rises when the catheter is blockaded.) If the blockage of the catheter is prevented, the risk of the infectious disease can be kept low. " as service practices.

The correction instruction was verified together with the persons who made the model and who corrected the task explanation messages. As a result, the both explanation messages corrected and not corrected are needed ideally. Though "Grasp of abnormality of the bladder function" task is a task that doctors finally judge, the nurse bears its partial task "Grasp of the volume of urine". At this task, for the nurse, there are medical aims "Grasp of abnormality of the bladder function" and "Grasp of urinating abnormality" in the task understanding of "Grasp of the volume of urine" task. The nurse who executes "Drinking water guidance" task is executing "Abnormal confirmation of the catheter" task. In such a situation, it is better for nurses to express the aim of "Drinking water guidance" task as "To blockage prevention of the catheter" than "Risk decrease of the infectious disease". It can be said that it will express the relation of the medical tasks as natural understanding in their medical practice. However, It is said that the previous aim is better to show the aim from the medical viewpoint. As mentioned above, there are service understanding that catches medical logic and service understanding that catches the point to make the work smooth. To find out P.K. by questioning from the medical staff, this example suggests that at least these two kinds of service understanding should be expressed.

4.2 Result of P.K. Acquisition

The number of interview messages is 97. The message are given for the 24 tasks in the 5 pathways. 88 messages get answers. The other 9 messages do not get answers. But 7 messages among them get the reasons for no answers. (2 messages had no answer without reason.) The reasons are "The P.K. can not exist, because the models have some mistakes and were modified in the interview" (for 3 messages), and "Avoid overlapping of answers" (for 4 messages). Most interview messages get answers. This result means the interview function based on the service perception model has worked as intended.

In the acquisition of practice wisdom, the case that show the necessity of externalizing task understanding was obtained. For instance, there was "Grasp of the pain level" task after the operation in the caesarean section pathway, and it was modeled assuming that comprised to each of "Grasp of the abnormality in function of internal organs" task and "Abnormal grasp in the operative site" task. However, there was an instruction of excluding it from "Grasp of the abnormality in function of internal organs" task. There was a comment that "The grasp of the pain level was a pain in foundation, and it was unrelated to internal organs" as a reason. When inquiring in detail, the answer "The result of the grasp of the pain level is concerning the pain of foundation. It is supposed to be described to the clinical record. In addition, whether the analgesic is used is decided as a result . The message was corrected to express such a task flow. The grasp of the abnormality in function of internal organs is not included in this flow." was obtained. On the other hand, the answer with "We observe the pain thinking about two possibilities in "Grasp of the pain level" was obtained.

In this example, an important point is the following. In "Grasp of the pain level" task, the medical staff has P.K. of doubting two possibilities. However, when P.K. is asked by questioning based on the task explanation that the task is a part of "Grasp of the abnormality in function of internal organs" task, she only comments that "the explanation is unsuitable for her own service understanding". On the other hand, int the questioning based on the task explanation regarding the task as a part of "Abnormal grasp in the operative site" task, they reply the P.K. This suggests that P.K. is closely related to the service understanding of the medical staff. In the acquisition of P.K., it is thought that the insistence on the this research that it is necessary to base the practitioner's service understanding is affirmed.

4.3 Result of Usage of Acquired P.K.

The aim of trial use of the obtained P.K. is to examine the P.K. The trial condition is below.

- Place: Departments of gynecology, urology, radiology, surgery and pediatrics.
- Period: 10th Nov. 2008. to 15th Jan. 2009. (The P.K. is used after trial continuously.)
- Interview about usage: 21st, Jan. 2009.

In the usage interview, there are no notes about the problem in explanation message of service task-perception and P.K. obtained. And the following comments are given by interviewees, "We can explain each medical task showing causes.", "The P.K. makes it easier to give rational guide and explanation to patients." and "It makes clear the reason for each task to novice staff and work rotation staff". The P.K. are mostly evaluated affirmatively. On the other hand, some improvements in system functions are hoped for "Some task explanation messages have redundancy" and "Some messages are unnatural" about the template message, and the P.K. presentation should be optimized to the users based on their expertise and work experiences.

In this trial use, whether the method functioned was only confirmed as the author had intended it. The method will be evaluated through the long-term operation of the system.

5 Conclusion

In this paper, the authors show the method and the systems to obtain the practical knowledge to provide patient-oriented medical services, based on modeling the service task-grasp (or task-perception) of medical staff. The service task-perception model externalizes the task understanding of each medical specialist. The medical staff have different sense of values based on their specialities. And the differences of sense of values request different types of P.K. The model give the basis to P.K. acquisition. Through the development and trial use of the system, the functions worked intended. This means the idea, to obtain and share P.K. based on modeled senses of worth of experts, is feasible.

In the future work, through using the system, the author will investigate how the ontology is organized, and how the different task-perception are externalized. The system now gives the framework to externalize the service task-perception only. We will implement the functions to check the model towards the medical aims, and to guide the modeling. To realize these functions, the authors will investigate the method to externalize not only service task-perception as a result of the thinking process, but also the thinking process itself, for the modeling guidance function. And the model searching rules to generate the interview messages are developed ad hoc. It should be organized with consideration of types of P.K.

References

1. Yoshikawa, H.: Introduction to service science (in japanese). Journal of the Japanese Society for Artificial Intelligence 23(6), 111–121 (2008)
2. Gennari, J.H., et al.: The evolution of protégé: An environment for knowledge-based systems development. International Journal of Human-Computer Studies 58(1), 89–123 (2003)
3. Schreiber, G., et al.: Knowledge Engineering and Management: The CommonKADS Methodology. MIT Press (2000)
4. Chandrasekaran, B.: Generic tasks in knowledge-based reasoning: High-level building blocks for expert system design. IEEE Expert 1(3), 23–30 (1986)
5. Mizoguchi, R.: Tutorial on ontological engineering - part 1: Introduction to ontological engineering. New Generation Computing 21(4), 365–384 (2003)
6. Coffey, R.R.: An introduction to critical paths. Quality Management in Health Care 14(1), 46–55 (2005)
7. Hasida, K.: Semantic Authoring and Semantic Computing. In: Sakurai, A., Hasida, K., Nitta, K. (eds.) JSAI 2003. LNCS (LNAI), vol. 3609, pp. 137–149. Springer, Heidelberg (2007)
8. Musen, M.A., et al.: Chapter 20: Clinical Decision-Support Systems in "Biomecical Informatics". Springer, New York (2006)
9. Tu, S.W., et al.: The sage guideline model: Achievements and overview. Journal of the American Medical Informatics Association 14(5), 589–598 (2007)
10. Hurley, K., Abidi, S.S.R.: Ontology engineering to model clinical pathways: Towards the computerization and execution of clinical pathways. In: Proc. in the Twentieth IEEE International Symposium on Computer-Based Medical Systems, pp. 536–541 (2007)

Real Time Hand Tracking as a User Input Device

Kittasil Silanon and Nikom Suvonvorn

Department of Computer Engineering
Faculty of Engineering, Prince of Songkla University,
P.O. Box 2 Kohong, Hat Yai, Songkla 90112 Thailand
jalego3@hotmail.com, kom@coe.psu.ac.th

Abstract. In this paper, we propose a system that facilitates the two dimension user-input as general mouse device for controlling applications. The method is based on hand movement analysis by applying image processing technique. The Haar-like with a cascade of boost classifiers is applied for hand detection, and then tracked with skin color using CamShift. Extracted hand features are computed and used for recognizing commands via a finite state machine. We evaluated the performance of system under real time constraint in real environment with demonstration application.

Keywords: Hand detection, Hand Tracking, Command Recognition.

1 Introduction

Trends in human machine interfacing have invented very rapid in consumer devices such as multi-touch of mini-computer like iPad, mobile phone iPhone, motion sensing of game devices, and etc. Relatively, many applications to support these platforms are actively developed. In this paper, we introduce a hand motion capture system using single camera that enables to track 2D hand position for commanding the desirable applications in real-time. The system offers very easy way of short range human-machine interfacing, which is very effective in terms of application constraints and prices, and generally very appropriate for the PC-based applications without any additional integrated material devices. There are many research try to solve the same problem. Wei Du [1] introduced a system with one camera that can recognize four gestures and tracking hand by extracting the feature points on hand contour. Nebojsa Jojic [2] developed a real-time system for detecting pointing gestures and estimating the direction of pointing using stereo cameras for controlling the cursor on a wall screen. Juan [3] proposed a vision-based system that can interpret a user's hand gesture in real time to manipulate objects and optimized for the medical data visualization environment.

We propose a very simple system that used only a consumer grade webcam as additional material, generally installed on top of the computer screen looking down towards the user's hands, for interfacing with applications. Our main contribution is then focused on the image processing techniques that try to detect hand pose and extract hand features, such as position, axis, and angle, which are used later in the command recognition process.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 178–189, 2011.

2 Proposed System

In this section, we introduce the system for command recognition based on hand motion by analyzing the image sequences, obtained from the consumer grade web-cam attached over the computer screen. Images are captured and analyzed in order to extract the essential parameters of the specific pose of hand. The parameters are reformulated into hand features which is necessary for command recognition.

There are two main parts of our system: hand detection and tracking, and command recognition. The overall of system is illustrated in the figure 2. First part, a hand detector is implemented using a cascade of boost classifiers, which allows obtaining very robust object detector. Two types of hand pose with vertical position are considered such as open and close hand, which is necessary for executing a command as start and stop symbol respectively. However, hand pose can have many figures caused by translation and rotation in 3D, which needs to be found. The tracking method, CamShift and Kalman filter, is then applied in order to extract hand through image sequence. In the second part, hand parameters are determined such as position, axis, area, and angle, which are defined as hand feature. A Finite Stage Machine is established based on these features for recognizing commands.

Figure 1 (left) shows the sequence of steps for hand detection and tracking; figure 1 (right) shows the recognition process. These two parts will be detailed in the section 3 and 4 respectively. In section 5, the experimentation result is discussed for evaluating the performance of system. For testing in real environment, we have implemented the system in windows 7, which our module can control any applications as a mouse interface. In the last section, we conclude our method.

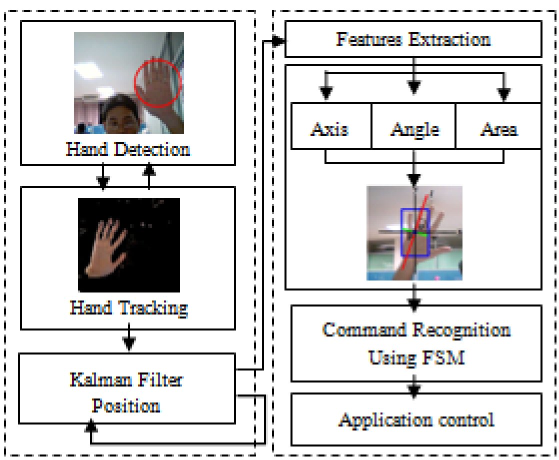


Fig. 1. System overview: (Left) Detection and tracking. (Right) Command recognition process.

3 Hand Detection and Tracking

3.1 Hand Detection

A Boost Cascade of Classifier [4][5] for object detection is originally developed by Viola and Jones. This method uses Haar-like features and a cascade of boosted tree classifier as a supervised statistical model of object recognition. Haar-like feature consists of two or three connected "black" and "white" rectangles. Figure 2 shows a basic set of Haar-like features.

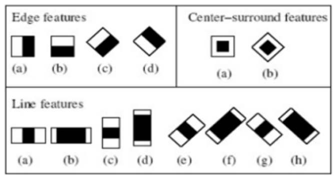
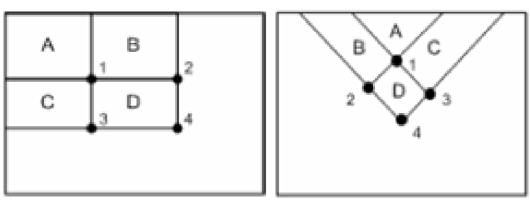


Fig. 2. A set of Haar-like feature

The feature value is the difference between the sums of pixel values within the black and white rectangles. By using integral image, these features can be calculated extremely fast. The integral image at point (x, y) is called Summed Area Table (SAT).

$$SAT(X,Y) = \sum_{x<X,\, y<Y} I(X,Y) \tag{1}$$

Then, the sum of pixels in upright rectangle can be determined from four integral image value of it vertices: Sum (D) = 1+4 – (2+3). For rotated rectangles, it is defined as the sum of pixels of a rotated rectangle with the bottom most corner at (x, y) and extending upward till the boundaries of the image. Similar to the upright rectangles, sum of pixel in 45^o rotated rectangles can be calculated from four integral image value: Sum (D) = 1+4 – (2+3).



(a) (b)

Fig. 3. Integral image (a) upright rectangle (b) 45^0 rotated rectangle

The AdaBoost algorithm is introduced to improve the classification performance that designed to select rectangle feature which best separates the positive and negative example. In the first iteration, the algorithm trains a weak classifier *h(x)* using one Haar-like feature that achieves the best recognition performance for the training sample. The classifier consists of a feature *f(x)*, a threshold θ_t and parity p_t indicating the direction of inequality sign.

$$h_t(x) = \begin{cases} 1 & \text{if } p_t f_t(x) < p_t \theta_t \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

In the second iteration, the training samples that were misclassified by the first weak classifier receive higher weights. The iteration goes on and the final result is a cascade of linear combinations of the selected weak classifiers $H(x)$, which achieves the required accuracy.

$$H(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \tag{3}$$

In the detection process by using the trained cascade, the sub-windows must be testing each stage of the cascade. A negative outcome at any point leads to the immediate rejection of the sub-window.

(a) (b)

Fig. 4. Hand detection (a) Input image (b) Hand detection result

3.2 Hand Tracking

CamShift [6] is a non-parametric technique using for color object tracking deriving from the Mean Shift algorithm. The main difference between CamShift and Mean Shift algorithm is that CamShift updates continuously its probability distributions; in generally the target object in image sequences changes significantly its shape size or color, while Mean Shift is based on static distributions. That why CamShift is suitable for tracking the rigid object. In the hand tracking, the process can be described as the following.

Step 1: the color probability distribution of detected hand image is determined from its histogram via hue component of HSV color space, related to skin color.
Step 2: this target distribution of detected hand is systematically tracked on the searching window of next frame in image using mean shift algorithm. The mean shift vector, which is aimed for finding an optimized path that climbs the gradient of a probability distribution to the mode (peak) of nearest dominant peak, is necessary to be computed.
Step 3: the back-projection technique, which associates the pixel values in the image (tracking hand) with the value of the corresponding distribution, is applied.
Step 4: the center of mass and size of tracking hand (projected image) is computed and defined as hand features. This step will be detailed in the next section (hand features).

Step 5: on the next iterative, the current position of hand in image is used for defining the searching window on the next frame. The process is repeated at step 2 continuously.

Note that the step 1 will be re-executed systematically if the detected hand by Haar-like features with boost cascade of classifier found in the searching window. We found that Haar-like method provides very accurate results when hand is paralleled to the vertical axis, compared with CamShift, but missed mostly in other directions, so that CamShift is applied in order to solve the problem.

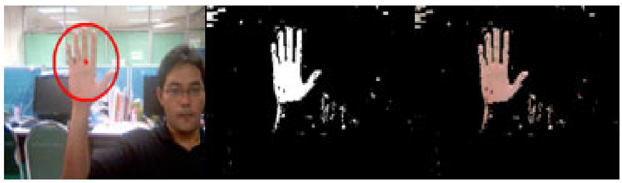


Fig. 5. Hand tracking with CamShift

3.3 Hand Movement Estimation

The Kalman filter [8] is a recursive linear filtering method. It addresses the general problem of trying to estimate the state of discrete time process that is described by the linear stochastic differential equation by the following.

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \tag{4}$$

With a measurement $z \in \Re^m$ that is

$$z_k = Hx_k + v_k \tag{5}$$

The random variables w_k and v_k represent Gaussian noise of the process and measurement respectively. The algorithm of Kalman filter estimates a process by using feedback control technique: estimating the process state by an appropriate model and doing feedback by noisy measurements. As such, the equations of Kalman filter are formed into two groups: prediction and correction equations. In the post tracking of hand, the algorithm can be described as the following figure 6.

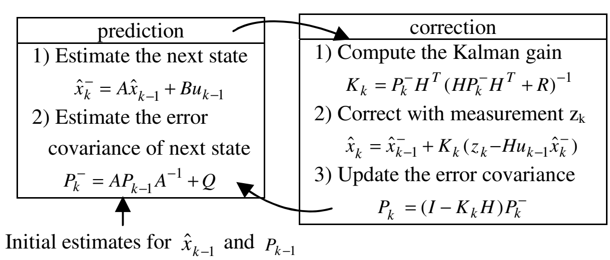


Fig. 6. The operation of the Kalman filter

Experimentally, we found that hand detector by Haar-like features with boost cascade of classifier cannot provide total results and CamShift may wrongly track especially when there are other parts of body such as face or background, having color in skin color range, move close to tracking hand. Therefore, the Kalman filter is applied in order to estimate new hand position in the next frame based on previous frame, obtained by CamShift, to enhance hand tracking.

(a) (b)

Fig. 7. Kalman filter (a) Position from CamShift for prediction in the next frame (b) Estimate new hand position (green dot) between CamShift (blue dot) and prediction (red dot)

4 Command Recognition

4.1 Hand Features

During the step 4 of hand tracking process by CamShift algorithm, in each frame, the hand characteristics are computed. The following parameters are considered: center of mass (p), axis angle (θ), area (z) and distance (d). The center of mass of the hand in a frame can be easily computed from the 0^{th} and 1^{st} order moments of pixels in hand's region.

Step 1: Find the 0^{th} moment: $M_{00} = \sum_x \sum_y I(x, y)$ (6)

Find the 1^{st} moment:

$$M_{10} = \sum_x \sum_y xI(x, y), \; M_{01} = \sum_x \sum_y yI(x, y) \tag{7}$$

Find 2^{nd} moment:

$$M_{20} = \sum_x \sum_y x^2 I(x, y), \; M_{02} = \sum_x \sum_y y^2 I(x, y) \tag{8}$$

Step 2: the center of mass X_c, Y_c is calculated as

$$X_c = \frac{M_{10}}{M_{00}} \text{ and } Y_c = \frac{M_{01}}{M_{00}} \tag{9}$$

In the above equations, I(x,y) is the pixel value at the position (x,y) of the image, x and y are range over the hand's region. At the same time, calculate the 2^{nd} moment; obtain length-axis, short-axis and angle of the hand's region.

Step 3: the angle of hand is

$$\theta = \frac{1}{2}\arctan\left[\frac{2\left(\frac{M11}{M00} - XcYc\right)}{\left(\frac{M20}{M00} - Xc^2\right) - \left(\frac{M02}{M00} - Yc^2\right)}\right] \tag{10}$$

Step 4: the long-axis l, short-axis w of tracking hand can be presented as :

$$l = \sqrt{\frac{(a+c)+\sqrt{b^2+(a-c)^2}}{2}},\quad w = \sqrt{\frac{(a+c)-\sqrt{b^2+(a-c)^2}}{2}} \tag{11}$$

$$a = \frac{M20}{M00} - Xc^2,\quad b = 2\left(\frac{M11}{M00} - XcYc\right)\quad ,c = \frac{M02}{M00} - Yc^2 \tag{12}$$

When used in tracking, the above equations give us angle, length-axis and width-axis. Then, we use axis for determined area (z) is *l x w* and distance (d) is $\sqrt{l^2+w^2}$.

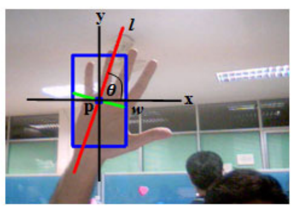


Fig. 8. Features of hand's region

By considering the parameters with respect to time t, we formulate the hand feature by the following:

$$f(t) = \left[p_t(x_c, y_c) \quad \theta_t \quad z_t \quad d_t\right] \tag{13}$$

4.2 Command Recognition

A simple recognition system of commands based on grid is established in order to interface the hand movement as user input of interactive application. The following commands are considered: (1) cursor displacement in four directions, and (2) clicks: left, middle, right. The main idea for designing the command recognition system is that to use the extracted parameters as features *f*.

For displacement commands, only $\partial p_t / \partial t$ is used for constructing command gird. A dynamic grid is used for command recognition. When hand has detected, which the 3x3 grid is created enclosing its center $p(x_c, y_c)$. Then, move hand from center to the neighborhood cells will activate the displacement command with respect to the specific direction. For the click commands, we consider the value of $\partial\theta_t / \partial t$ and $\partial z_t / \partial t$ defined by the following: click left if only if $\partial\theta_t / \partial t$ is positive, click right if only if $\partial\theta_t / \partial t$ is negative, and middle click if only if $\partial z_t / \partial t$ is less than $z/2$. Figure 9 summarizes the FSM states and transitions.

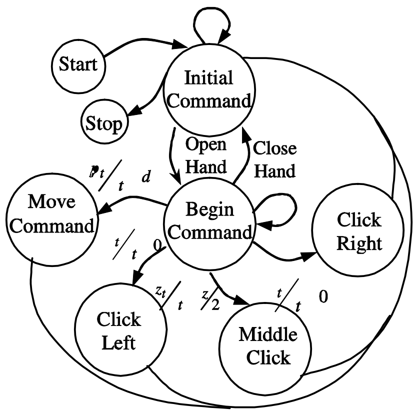


Fig. 9. Finite State Machine for command recognition

Figure 10 shows the example result of move command and click commands, for click left, middle click, and click right respectively.

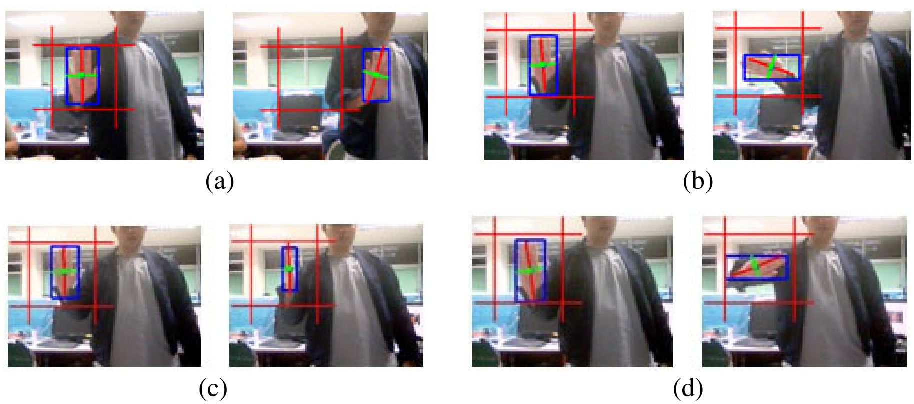


Fig. 10. System command: (a) Move command (b) Left click command (c) Middle click command (d) Right click command

5 System Performance and Discussion

Our system is implemented using OpenCV library. Testing system is run on Intel processor Core 2 Duo, 2 Ghz, 2 GB memory. The video sequence is analyzed with image resolutions 320x240 pixels at 30 fps.

5.1 Hand Detection

In our system, a command begins with open hand gesture and stop using close hand gesture. Each frame, these two gestures must be detected. In the training processing, we have collected ~5,000 positive samples for each hand style from students in our university in various conditions such as: indoor with neon light and outdoor with natural light. Systematically, around 10,000 negative samples are selected from landscape, building, and human faces or body images. Figure 11 show some of these samples.

(a) (b)

Fig. 11. Training Samples: (a) Positive samples (b) Negative samples

Each gesture type was trained two rounds. The first round is for discriminating quickly the target gesture from the outliers with small samples. By observing the false positive and false negative, we determine the additional positives and negative samples in order to overcome the better recognition rate while reducing noise.

(a) (b)

Fig. 12. Hand detection results: (a) 1st round (b) 2nd round

Table 1 show the detection rate results from our experimentation of the two hand gestures with respective configurations. We notice that a good detection rate is raised at least more than 91% with 16 training stage at minimum.

Table 1. Hand detection rate

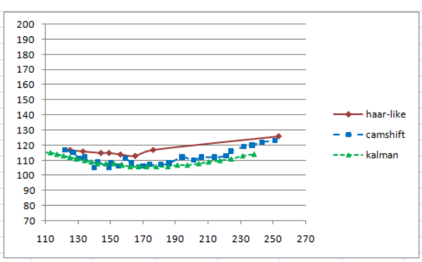
Hand Gesture	Window Size	Training and Performance			
		Positive	*Negative*	*Stage*	*Rec. Rate*
Open	32x32	5,058	10,448	19	91.17%
Close	32x32	5,678	10,448	16	99.06%

5.2 Hand Tracking

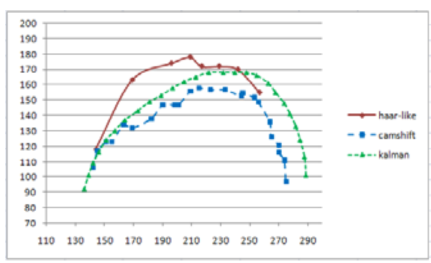
We evaluate the precision of trajectory obtained from hand tracking process. The test is done by executing the hand movement following three types of ideal path: straight line, curve and v-shave respective. Then, we compute the errors, defined by the difference between ideal path and tracking position. Table 2 shows results of ten times test with average. We found that the error in average of curve is bigger than v-shape and line respectively.

Table 2. Hand tracking with errors

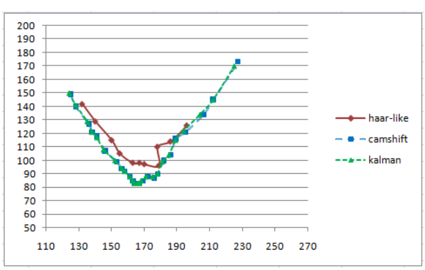
No	Testing errors (pixels)		
	Line	***Curve***	***V-shape***
1	1.54	3.79	8.05
2	1.70	3.61	3.31
3	1.39	6.87	4.22
4	1.59	6.68	9.13
5	1.97	7.55	4.98
6	0.97	8.93	3.51
7	1.40	8.09	4.31
8	2.02	4.23	4.44
9	1.76	6.05	7.05
10	1.84	6.97	6.38
μ	1.61	6.27	5.53
σ	0.29	1.74	1.89



(a)



(b)



(c)

Fig. 13. The precision of trajectory: (a) Line (b) Curve (c) V-shape

5.3 Command Recognition

The performance of command recognition is tested depends on two factors: distance of hand from camera, and hand speed for executing a command. Testing is done under indoor environment, no change of light condition, with the following test: three fix distances with three levels of speed, ten times of testing in each case.

Table 3. Experimentation results

Distance (cm)	Hand speed (s)		
	0.5s	1s	2s
30	10/10	10/10	10/10
60	10/10	10/10	10/10
90	7/1	8/10	9/10

Table 3 shows our testing results. We note at low speed commands can be executed more precise than high speed one. Intuitively, during recognition process more frames are analyzed under low speed that increases eventually the effectiveness of the method. We found the commands at near distance provide high recognition rate than far distance. Certainly, when the image region of hand in image is too small, then hand cannot be detected. We demonstrate our system as a user input that interface to Window Media Center application on Windows 7. You can see video showing the real-time interactive at **http://www.youtube.com/watch?v=loVxPEP5fME**.

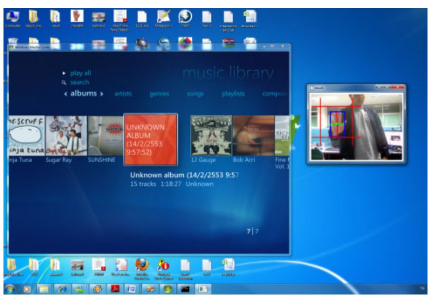


Fig. 14. Our module is implemented in Windows 7 for controlling Window Media Center

6 Conclusion

We have introduced the image processing methods for hand detection tracking and recognition aimed for human applications interfacing like a general mouse input. The hand trajectory and other features are precisely extracted, using for command recognition. Eleven basic commands are recognized at very high rate with respect to different speeds and distances. The implementation of our system on Windows 7 shows an example test under real-time constraint with real application.

References

1. Du, W., Li, H.: Vision based gesture recognition system with single camera. In: 5th International Conference on ICSP, vol. 2, pp. 1351–1357 (2000)
2. Jojic, N., Brumitt, B., Meyers, B., Harris, S., Huang, T.: Detection and estimation of pointing gestures in dense disparity maps. In: Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 468–475
3. Wachs, J., Stern, H., Edan, Y., Gillam, M., Feied, C., Smith, M., Handler, J.: A Real-Time Hand Gesture System Based on Evolutionary Search. Vision 22(3) (October 2006)
4. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE CVPR, vol. 1, pp. I-511– I-518 (2001)
5. Lienhart, R., Maydt, J.: An extended set of Haar-like features for rapid object detection. In: International Conference on Image Processing, vol. 1, pp. I-900–I-903 (2002)
6. Bradski, G.R.: Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal Q2 (1998)
7. Welch, G., Bishop, G.: An Introduction to the Kalman Filter. In: Annual Conference on Computer Graphics & Interactive Techniques, August 12-17, ACM Press, Addison-Wesley, Los Angeles, CA (2001); SIGGRAPH 2001 course pack edition

Freedom of Choice and Creativity in Multicriteria Decision Making

Andrzej M.J. Skulimowski

Department of Decision Science, Chair of Automatic Control, AGH University of Science and Technology, al. Mickiewicza 30, 30-050 Kraków, Poland
and
International Centre for Decision Sciences and Forecasting
Progress & Business Foundation, Kraków, Poland

Abstract. This paper presents new approaches to formulating and solving complex real-life decision-making problems, making use of the creativity concept. We assume that the decision-making process is embedded in the system of views and mutual relations between the decision-makers and their surrounding environment, so that creativity, as defined formally in Sec. 2, could play a primary role in the decision-making process. We will investigate multicriteria decision problems, where the decision-maker is unable to fully follow decision-making rules resulting from a standard mathematical formulation of multicriteria optimization problem. This is either due to external conditions (such as the need to make a quick decision, loss of data, or lack of data processing capabilities) or when the decision-maker can manifest creativity related to the hidden internal states of the decision-making process. We will provide a formal definition of freedom of choice (FOC), specifying three levels of FOC for multicriteria decision-making (MCDM) problems. Then we will point out that creativity in decision-making can be explained within the framework of autonomous and free decisions, and that decision-making freedom is a necessary prerequisite for creativity. The methods presented here can be applied to analyzing human decision-making processes and conditions allowing the expression of creativity as well as to designing pathways leading to creative decision-making in artificial autonomous decision systems (AADS). The applications of the latter include visual information retrieval, financial decision-making with feature identification, intelligent recommenders, to name just a few.

Keywords: Multicriteria decision-making, freedom of choice, creativity, artificial autonomous decision systems, reference sets.

1 Preliminaries

The aim of this paper is to provide the fundamentals for intelligent, autonomous and flexible embedded applications that can govern independent systems or engage in interactions with humans and artificial intelligent systems. We will directly address the cognitive aspects of such systems, focusing on new approaches to human-computer interaction. Although neither purely normative, nor purely descriptive decision

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 190–203, 2011.

theory is used, both are merged in intelligent intention-understanding (and –reconstructing) systems based on modeling the cognitive phenomena of human decision-making. The results presented here should constitute the first steps towards designing a prototype generator for such systems as well as specialized systems derived from the generator. These systems would be tailored to specific applications, including visual preference processing and *decision pilots,* a subclass of recommenders that provide rankings and implement constraints but not the final choice.

Throughout this paper we will refer to decision problems that can be modeled as a multicriteria optimization problem

$$(F:U \rightarrow E) \rightarrow min(P) \quad (1)$$

with a set of selection rules *S,* by definition

$$S \subset \{s:2^{F(U)} \times K \rightarrow 2^U: \forall V \subset U, \omega \in K\ s(F(V),\omega) \subset P(U,F))\}, \quad (2)$$

where U is the set of admissible decisions, the function $F=(F_1,\ldots,F_N)$ is a vector criterion, E is an ordered space with a partial order P, which is constructed from individual preferences concerning the values of F_i, $i=1,\ldots,N$ and some additional conditions imposed on the values of F. Each selection rule S maps the known evaluation with F of a subset V of U and an external knowledge state $\omega \in K$ into the subset of potential best-compromise decisions in V with respect to ω, which – to conform with (1) - are elements of the set of *Pareto optimal* (or *non-dominated)* decisions $P(U,F)$.

Most often, $E=\mathbb{R}^N$, P is defined by a closed and convex cone, selection rules are applied sequentially to single elements of $F(U)$, and K is a set of relations contained in the expert knowledge of the decision-maker or consultant. For instance, K may contain information concerning the constraints to be imposed on trade-off coefficients between criteria F_i. In order to enable the theory presented in this paper to be applied to artificial decision systems, we will examine the possibility of retrieving criteria from the environment using feature identification and retrieval methods, cf. e.g. [5], [7],[8]. We will also assume that the external knowledge and decision selection rules use trade-off and direct information simultaneously, the latter given in form of reference sets [10],[12], anticipated consequences [9],[14] and the history of previous choices.

When studying the theoretical background of intelligent intention-understanding systems, regarded as a class of cognitive decision support systems, we encountered a problem – the limited ability of artificial systems to make creative decisions. Creativity, in its common-sense meaning, was possible only up to a specific extent and constrained by superordinated (human) system objectives. This problem led in turn (cf. [13]) to the definition of the principles of *freedom of choice (FOC),* which we will now formulate in a more formal form as

Definition 1. *Freedom of choice of the 1st order* in problem (1)-(2) is the ability to choose an optimal solution from a given set of admissible alternatives with respect to a set of selected optimality criteria that are specified explicitly. *Freedom of choice of the 2nd order* allows the decision-maker to relax the constraints [11], or – more generally - to embed the initial set of admissible decisions into a superset. Finally, *freedom of choice of the 3rd order* is the power to select one's criteria of choice in the feature space of real-life objects, which are the subjects of the decision maker's final choice.

The criteria considered when defining FOC may be accompanied by external preferences and decision-support procedures, which can also be a subject to choice in some circumstances. Based on the above notions, in the next section we will define creativity as the ability to act within a knowledge-based variable decision model with information inflow. Note that the essence of the notion of *autonomy* is to *have the ability to make decisions.*

The above types of freedom of choice will be attributed to the *primal decision-makers*, which is a basic notion. Their features, actions, relations and interactions between them can be modeled by an ontology. The basic class of this ontology can be extended by adding *secondary decision-makers*, who will be given the ability to learn from their past decision experience and by *supervisors*, who can assign a level of FOC and creativity to other decision-makers. In this paper we will consider a situation where the ability to transfer freedom of choice is restricted to artificial autonomous systems, such as embedded intelligent systems for autonomous vehicles, research robots etc.

The behavior of decision makers when solving multicriteria decision-making (MCDM) problems can be properly explained by fundamental research on neural and psychological decision-making mechanisms, including creativity and additional cognitive phenomena. In Sec. 3 we will show that the use of direct preference information such as reference sets together with feature detection procedures can allow for an automatic formulation of criteria of choice, which – in turn – gives artificial autonomous systems second-order freedom of choice. The greatest freedom can be explored in common multicriteria decision problems, which can be interpreted as multi-step games with nature, the latter implementing an unknown, non-feedback strategy. The solution methods of extended MCDM problems, including cognitive phenomena models, depend strongly on the application area. Therefore different mathematical models will be applied depending on the different mental processes that are behind the solutions of decision problems in different areas. This may happen even if the multicriteria optimization problem formulations are of the same kind, yet the additional preference structures differ considerably. The investigation of these companion circumstances can allow us to assign appropriate MCDM models and methods to different classes of complex decision problems, which is one of the key issues in MCDM.

Furthermore, in Sec.4 we will discuss the problems related to the design of intelligent and creative decision support engines capable of understanding the implicit preferences in real-life human decision problems by reconstructing and disclosing the latent intentions of the decision makers. We will show that to achieve this, computer-based artificial decision systems must possess the ability to search and select criteria of choice, i.e. to have third-order freedom of choice. Such capability has never before been given to artificial systems in an unrestricted form, and numerous authors, from professional foresight to sci-fi writers, have warned about intelligent artificial beings able to make and implement a decision. In Sec. 4 we will also strive to discuss the question as to whether decision support systems (DSS) should possess freedom of choice and creativity and to what extent such systems should use cognitive principles to understand the intentions of human beings. The above-mentioned high-level artificial systems will be endowed with freedom of choice and cognitive skills such as the ability to search, understand, classify and select information, in particular audiovisual streams.

As a field of application, we will point out the multicriteria decision-making problems related to gathering, filtering and selecting images on the web using pattern discovery, recognition and classification, as well as cognitive decision theory methods [8]. The main real-life applications will be systems able to represent, analyze and interpret human behavior. Another specific application field of artificial systems able to follow cognitive decision-making mechanisms is embedded systems. In the latter, human decision makers lacking the ability to make a decision can be assisted in part or even replaced by an artificial system capable of making and implementing decisions. In addition, the above framework will allow us to create new DSS architectures and model the interactive decision processes.

2 Main Problems and Ideas Related to Creative Decision Making

Although creative decision making should be regarded as one of the highest-level cognitive functions of the human mind, its relation to other branches of cognitive science is sometimes unclear. There are four main reasons for this:

- The fragmented partial nature of research on eliciting and modeling human preferences, where psychometric research and the prevailing theoretical studies on multicriteria decision analysis are not fully consistent, and the decision support systems are often designed without paying enough attention to real-life human decision making mechanisms.
- The tendency to restrict the modeling of human decisions to cases where the decision maker is either able to formulate criteria of choice or able to explicitly define the set of admissible alternatives. What follows is a mathematical programming or gaming problem, and subsequent efforts are focused on solving it, without taking into account cognitive phenomena like quickly changing preferences, an extension/contraction of the decision scope resulting from different cognitive processes, replacement of consequence modeling by a single score called a utility or value function.
- The lack of adequate decision models for complex system behavior when the information underlying the decision making has a graphical form. In such situations, autonomous systems are usually restricted to performing pre-defined tasks, according to their clearly defined hierarchy of goals (like an unmanned vehicle's task to reach certain a destination with a load at minimum cost or time, avoiding obstacles and other vehicles)
- The vivid debate concerning the existence of the human decision maker's utility functions, which has been the subject of numerous studies over past decades and has re-directed the interest of researchers to questions other than creativity issues.

This is why insufficient emphasis has been put on researching creative mechanisms of decision-making processes so far (a noteworthy exception is [4]).

As mentioned in the previous section, we will relate the creativity that can be shown when solving the decision problem (1)-(2) to the ability to vary the decision model. In this way, a new decision admitted is better than any optimal or best-compromise decision under the previous model, subject to new quality criteria, however. A necessary condition to modify the problem is the ability to adapt the decision-making process according to external information inflow.

Thus, we can formulate the following:

Definition 2. A *creative extension* of problem (1)-(2) will be defined as a morphism $\varphi=(\varphi_1,\varphi_2,\varphi_3,\varphi_4)$ transforming the set of admissible alternatives U, the vector criterion F, the preference structure P, and the set of decision rules S into a new problem

$$(\varphi_1(F):\varphi_2(U)\rightarrow E_1)\rightarrow min(\varphi_3(P)) \tag{3}$$

with the new set of decision rules S_1 induced by $\varphi_4(K)$, i.e. if $s_1\in S_1$ then

$$s_1:2^{\varphi_1(F)(\varphi_2(U))}\times\varphi_4(K)\rightarrow 2^{\varphi_2(U)} \tag{4}$$

A sequence of creative extensions $\varphi_{(1)}\rightarrow\varphi_{(2)}\rightarrow\ldots\rightarrow\varphi_{(n)}$ such that the output problem of the morphism $\varphi_{(k-1)}$ *is an input problem for* $\varphi_{(k)}$ *, for* $k=2,\ldots,n,$ will be called the *creative decision process.*

A scheme representing the creative process that allows one to extend a multicriteria decision problem while taking into account the above situations is presented in Fig.1.

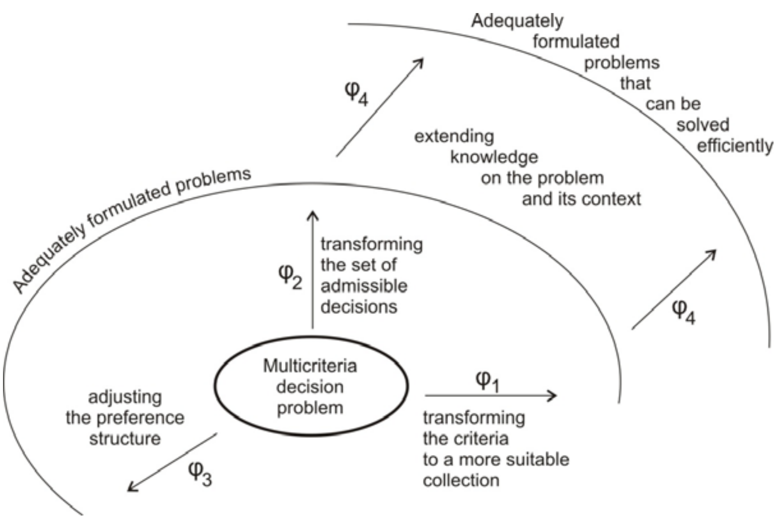


Fig. 1. A scheme of the creative analysis of an MCDM problem

Problem-solving approaches that explore the above-defined creativity concept and lead to a more penetrative analysis of decision problems have already been published [9], [10]. These include:

- The problem of replacing the search for a single best-compromise decision by selecting a subset of admissible decisions that is evaluated by new aggregate criteria such as the average values of original criteria calculated on a subset;

- Considering the problem in a temporal context, taking into account a.o. the choice of an optimal time to make the decision (or announce it), the variable values of attributes that characterize decision alternatives, the possibility of replacing a decision made by another choice etc.;
- Creative elimination of redundant objectives and elements of a preference structure performing an analysis of the previous decision maker's choices and using essential feature-detection algorithms;
- The replacement of a single utility value by a multi-level model of the consequences of a decision made, which can lead to an anticipatory analysis of the problem;
- Creative exploration of the set of decision rules, taking into account cognitive aspects of the decision-maker's behavior.

We will refer to the above-mentioned examples of creative problem extensions in the next sections of this paper. Moreover, creativity can contribute to overcoming the following drawbacks of general approaches to multicriteria problem solving, decision models and support systems:

- Decision support systems are often confused with data storage, retrieval and visualization systems. This corresponds to the first basic problem of decision-making, i.e. decisions made under uncertainty and lack of information. The awareness of the second kind of decision-making problem, i.e. problems where the decision maker's preference structure is non-compatible with the data available or must be elicited, is considerably lower. The knowledge of cognitive decision-making mechanisms that can increase the efficiency and adequacy of this class of decision support systems is even lower.
- There exist a variety of so-called interactive decision-making algorithms (cf. [6]), where the plausibility of information required during the dialogue procedures is not checked, and its further processing does not guarantee final success in form of a convergence to a satisfying compromise decision. Mathematical convergence conditions need not be compatible with the cognitive process of decision making, do not take into account discoveries other than the identification of the next candidate for a compromise solution, and even a simple 'change of mind' by the decision maker during the formal procedure can result in a 'convergence disaster'.
- Contemporary DSS are rarely individualized, and their ability to learn based on previous choices is restricted to new instances of the same problem, if at all.

In spite of the above-mentioned drawbacks, the appearance of computer-aided decision support has contributed to a remarkable change in the practice as well as the philosophy of decision making. The traditional value and role played by the intuition and experience of the decision maker remain relevant, but they should be accompanied by an interactive computer-aided process consisting of

- systematic analysis of available data
- data acquisition and representation.

This should be followed by formulating or re-formulating a mathematical model of the particular decision situation with a representation of knowledge about the preference structure.

While it is difficult to judge the effectiveness of a formal model at this stage, the degree of decision maker satisfaction with the decisions generated using the data and mathematical model can provide information on the model quality. The final phase of the decision process may involve heuristic and non-formalized stages in which initial expectations, intuition, and experience are confronted with the results of machine analysis.

It is commonly accepted that the decisions made should either be optimal, sub-optimal, or compromise. This assumption requires a set of optimality criteria to be defined for each admissible alternative. The notion of optimality has passed an instructive evolution aimed at reaching a higher-level of FOC: from the global minimum of a scalar objective function (usually no freedom of choice), through the notion of a vector minimum in partially ordered criteria space (allowing a choice from among non-comparable non-dominated solutions), to the idea of the extended optimum, as considered in subset selection or dynamical extension problems [10]. According to (3)-(4), these belong to creative decision problem extensions and to the identification of search criteria in a cognitive process [8], the latter leading to third order FOC.

An approach to embedding creativity into decision analysis and support can be based on several key principles:

- the idea of excluding any loss of information about the decision to be made, even inconsistent and contradictory information, which should first be judged as to whether it touches upon the same objects and time instances; pre-defined verification rules can turn out unsatisfactory, which creates a playground for creativity to cope with additional information acquisition and consistency checking
- most decisions should be considered in an *open information space*, i.e. a situation in which the inflow of information in real-time is relevant to the decision making; here, creativity can manifest in the appropriate exploration of this information.

The first principle is based on the following three underlying assumptions, which originate from real-life experience:

- first, we assume that all information available should be applied to the decision process as soon as possible (pre-classification or elimination should be used only exceptionally)
- second, it is assumed that mutually inconsistent or contradictory information received from the decision-maker should be treated as a result of a cognitive process; using a reconstruction of the latter, it should be re-defined or aggregated, but not neglected during the decision-making process
- third, we assume that reference sets and bicriteria trade-offs are the most suitable tools to sufficiently describe many real-life preference structures.

With the usual approach used in interactive decision support, partial information is used there to generate a decision and further information is processed only if the solution was not generated to the decision maker's satisfaction. The final decision depends on the subjective sequence in which the additional information was processed. Moreover, such procedures prolong the decision-making process unnecessarily, and may lead to a loopy inconsistent process, since the convergence conditions for interactive procedures are usually based on overoptimistic assumptions concerning human rationality. Despite the above disadvantages, the success of this philosophy may be explained by the well-known paradigm which states that even an incorrect decision is sometimes better than no decision. In addition, some decision makers, especially those acting as managers or politicians, need an alibi or some kind of decision support to justify a voluntarily or randomly chosen decision.

The first assumption implies that the above approach should be replaced by more complete and thorough analysis of data available.

The second assumption implies the need to understand the human cognitive process that accompanies the decision making. Incorrect or inconsistent statements are often caused by the shortcomings of human perception of decision object features. Instead of forcing the DSS users to give another reply, an intelligent DSS should trace the replies, find and indicate the possible source of inconsistencies using a cognitive perception model [1]. For instance, as a cognitive and temporal extension of the well-known AHP method [2], we propose a procedure which measures the user's reply time in pairwise comparison queries, then relates it to the reliability of scores entered using the observation that the faster the reply is generated, the higher the probability of getting an extremal value of the scale. Another example of using cognitive models to improve well-established interactive procedures is to replace the presentation of one element at a time (classical multicriteria interactive methods) or two elements at a time (pairwise-comparison-based methods) to the user, by presenting arbitrary sets of alternatives at one time and using different types of questions that may be easier to answer, depending on the perception profile of this particular user.

The third assumption refers to the *bicriteria trade-off hypothesis* [13], which states that irrespective of how many criteria are used to make a decision, human decision-makers intuitively try to group them into two aggregated groups and then solve them as a bicriteria problem. This hypothesis, which has recently received new justification coming from brain research, will be further investigated. In particular, one should attempt to find the factors that influence the sequence of aggregation, relating them to feature perception and selection processes.

The second principle mentioned above is based on manifold evidence from experiments performed with web information sources. For instance, when using price comparison engines and recommenders to support goods selection in e-commerce systems, the availability of price offers was changing in time lapses comparable with those needed to make a decision, so no classical optimization problem could be formulated (cf. experiments with air ticket availability [1] and electronic auctions). Apart from internet-based search, similar effects can be observed in personnel or apartment choice, where optimal stopping-based approaches have been applied. The expected rising information flows on the internet will accelerate this process, and create the need for more adequate decision support, using cognitive process modeling.

The notion of creativity proposed here allows us to extend the interactive problem solving process to interactive problem formulation and data acquisition. Thus the problem statement and solution processes are combined in one interactive procedure with several levels of interaction. We expect that this novelty will become a standard in future decision support systems, and its first successful implementation will decide its dominance of the intelligent recommender market.

3 Multicriteria Decision Problem Formulation by Autonomous Systems Based on Reference Sets and Feature Selection

A further extension of common decision analysis approaches that can be attributed to creativity is based on finding relevant features of objects using so-called reference sets. As the features of objects to be chosen can be regarded as pre-criteria, the detection of features by an artificial autonomous decision system (AADS) can allow an automatic formulation of criteria of choice. This can, in turn, give artificial autonomous systems second and higher order freedom of choice. With this approach, the decision maker is able to formulate and process several classes of relevant reference values. However, its scope of subjective handling is restricted, thus the FOC is also restricted.

The foundations of feature subset selection have been given e.g. in [7]. Referring to the reference set approach in multicriteria analysis [12], it should be noted that presenting typical representatives of each class is a common approach in classification procedures. Further steps involve finding characteristic features of such representatives [5]. An intrinsic part of this procedure is learning the relevance of features, the preferences, and, finally, the relevance of objects described by the features. For classification purposes, this approach is often called 'learning from examples'. The approach presented here differs from that used for classification and recognition in the instrumental role of 'cut-off' points used during the procedure: the reference points which form reference sets need not be elements of the database to be researched, and, with the exception of status quo solutions (class A_2 defined later on), do not even need to exist. The essential information is the preference relationship between the elements indicated by the user, who defines the reference points. The examples point out how such a relation acts, then the retrieval algorithm learns how the preferences can be expressed by comparing individual features, and applies them to select alternatives automatically.

Reference sets as defined in the theory of MCDM [12] constitute a tool originally designed to support portfolio selection and design decisions. They are defined as sets of characteristic points in the criteria space with similar levels of utility. There are four basic types of reference sets:

- A_0 – boundaries of optimality – upper (in the case of criteria maximization) boundaries of an area where optimization of criteria makes sense
- A_1 – *target points* –optimization goals
- A_2 – *status quo solutions* – existing solutions which should be improved in the optimization process or lower boundaries of the set of satisfactory solutions
- A_3 – *anti-ideal point* – solutions to avoid.

These set types can be further split into subclasses. All, or only a few, classes of reference sets may occur in a decision problem, while the consistency of problem formulation imposes a number of conditions to be fulfilled by the reference sets. In the context of making a compromise choice, the reference sets can be interpreted as follows [8]:

- A_0 is defined by a set of queries provided by the decision-maker. We assume that the goal of the decision-making process is to find a solution, which is most similar to one of his queries. When no such query can be provided then $A_0=\varnothing$,

- A_1 is a set of criteria reference values ranked by the decision-maker as *most relevant*,
- A_2 is a set of points in the criteria space ranked by the decision-maker as *relevant*,
- A_3 is a set of points in the criteria space ranked by the user as *irrelevant*.

In addition, we can define elements of set A_4 as points ranked by the user as *anti-relevant*, i.e. characterized by attribute values opposite to those sought after. We assume that explicit user preferences constitute the primary background information. We also assume that the criteria can be constructed gradually using the preference information elicited during the interactive search process. Thus, even the number of relevance criteria cannot be assumed to be known a priori as various classes of potential solutions may be characterized by different sets of features and coefficients. The usual application of reference sets is an interactive decision process, where the originally specified targets (set A_1) converge to more realistic, yet acceptable values, the status-quo solutions are updated during the procedure to include newly acquired information about the results of similar choice problems, and so can vary the other classes of reference sets. A scheme of selecting a decision when different classes of direct information about the values of criteria have been defined as reference sets is presented in Fig.2 below.

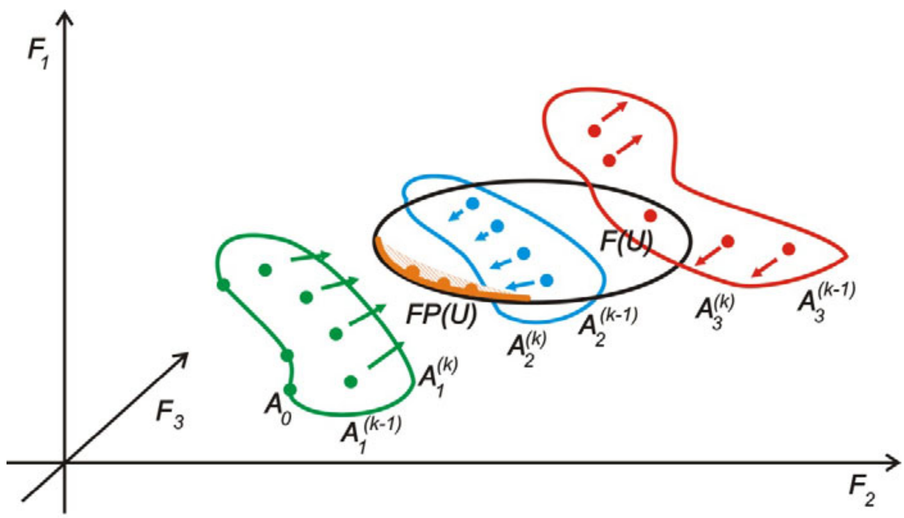


Fig. 2. An example of reference sets used for selecting a compromise solution of a multicriteria decision problem. Arrows denote the update directions during an interactive procedure.

The experiments made on image databases [8] illustrate the considerable potential of the above approaches, which can increase the feasibility and efficiency of Internet-wide graphical recommenders, bringing them to the level of creativity.

Thus a major improvement of Content-Based Information Retrieval (CBIR) can be achieved by a combined application of classification methods and multicriteria preference elicitation. This research direction promises further progress, with the main targets being:

- an optimal man-machine interface when designing a query and providing feedback information
- optimized navigation in the feature space and adaptive feature space contraction/expansion.

An autonomous Internet search engine based on reference sets, with a number of predefined high-level features incorporated is now being implemented [8] to allow a reliable efficiency assessment of the proposed method for this type of application. Its performance will depend on the quality of high-level feature extraction, on the elaboration of a reliable feature extraction algorithm, and on the design of an adaptive decision-making engine, which is, in itself, one of the most challenging problems of creative decision making related to information retrieval.

4 The Design of Creative Decision Support Procedures

The implementation of the reference set method outlined in the previous section can be designed as a user-friendly decision support system allowing users to consider different classes of reference sets, criteria-space constraints, and trade-offs in one model. In addition, it can easily incorporate the decision freedom features, such as changing the set of criteria (FOC-level 3), variable constraint set (FOC-level 2), freedom to choose from a set of nondominated points by selecting an appropriate compromise solution selection method (FOC-level 1). Taking into account the creativity as it has been defined by (3)-(4), one can conclude that the FOC of an appropriate level implies the ability of the decision-maker to apply the corresponding components of the creativity morphism $(\varphi_1,\dots,\varphi_4)$. In real-life DSS applications that would be capable to support creativity in multicriteria decision problem solving, one would require that such systems can support problem formulation as well as its solution within an integrated process. In addition, smart and *creative* acquisition of additional information that can facilitate problem solving should be supported.

The scope of potential applications of this methodology has not been a priori restricted, and it should provide well-suited decision support for a large class of underlying multicriteria decision problems. For instance, specialized decision support procedures can be designed for solving industrial robot diagnostics, financial decision making and planning, or group decision and negotiation problems. Social decision processes can be modeled in a similar way as the decisions made in controlled discrete event systems [10].

Besides its impact on real-life-oriented decision support methodology, the notion of creativity can contribute to the solution of several unresolved theoretical problems, such as the estimation of the value function in image retrieval decision-making problems and the simultaneous analysis of additional preference information given in form of multiple classes of reference points and trade-off constraints [10],[12].

Based on theoretical background provided in [9],[10],[11],[14] one can propose a scheme of such creative decision support system that integrates multicriteria problem re-formulation, enables its embedding in more general problems, allows for additional information acquisition during the decision-making process and takes into account the temporal aspects of decision-making. A graphical scheme is shown in Fig. 3.

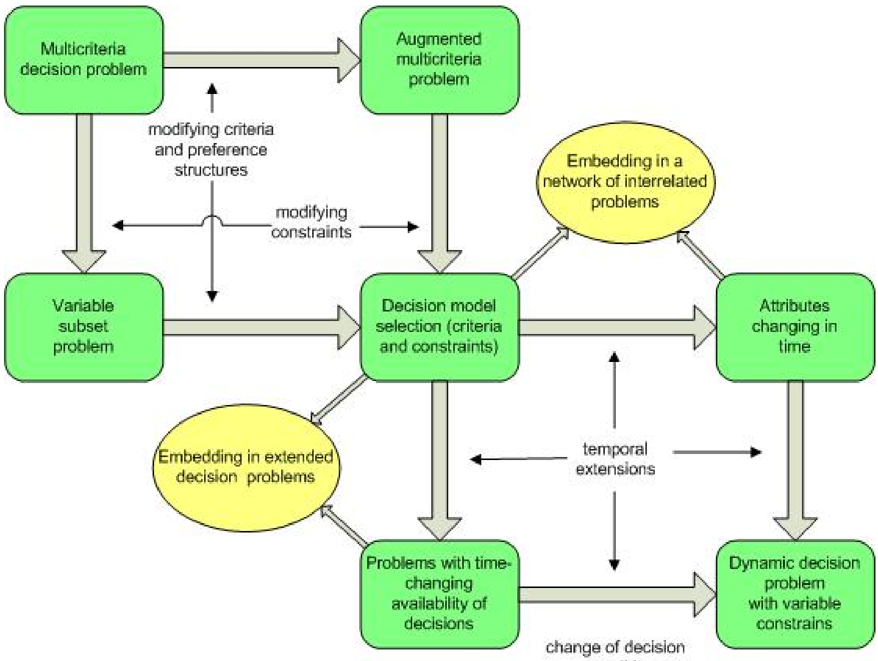


Fig. 3. A scheme of multicriteria decision problem formulation that can be supported by a creative decision support system

The scope of potential applications of this methodology has not been a priori restricted, and it should provide well-suited decision support for a large class of underlying multicriteria decision problems. For instance, specialized decision support procedures can be designed for solving industrial robot diagnostics, financial decision making and planning, or group decision and negotiation problems. Social decision processes can be modeled in a similar way as the decisions made in controlled discrete event systems [10].

Besides its impact on real-life-oriented decision support methodology, the notion of creativity can contribute to the solution of several unresolved theoretical problems: the estimation of the value function in image retrieval decision-making problems and the simultaneous analysis of additional preference information given in form of multiple classes of reference points and trade-off constraints [10],[12].

The common principles of computer decision support tools can be listed as follows:

- Irrespective of how advanced mathematical methods are used to process the underlying information, the sophisticated decision support procedures should not be visible to the decision-maker (*mathematical ignorance assumption*)
- It is impossible to avoid checking the consistency of information supplied by the decision maker, including the preference statements. If necessary, they can be redefined in the background so that the consistency conditions are fulfilled.

Nervous, tired or irrational decision makers will supply responses that can lead to chaotically generated dominant alternatives, so that the conditions for terminating the decision-making process were never fulfilled. Nevertheless, in many situations the same nervous or tired decision makers may urgently need a quick and efficient decision aid. To cope with such situations, decision support systems for common use, such as recommenders, should feature an autonomous mechanism (*automatic decision pilot*) that makes it possible to either defer the decision or make a *cautious decision* autonomously. The latter should be understood as the selection of an alternative which conforms maximally to a decision made previously by this system's user and to the individual cognitive decision model.

Having formulated a decision problem, the computer-aided decision support should require no specialized computer or mathematical knowledge. The use of additional preference information in form of reference sets and bicriteria trade-offs to choose a compromise decision, supplemented by extensive visualization and guideline procedures, will allow the system to provide the resulting compromise decision in annotated graphical form. External experts may only provide assistance to apply more penetrative consequence analysis techniques or to retrieve missing data. When artificial systems are given second or third-order FOC it should be noted that formulating the optimization problem with a preference modeling mechanism is just a part of the overall problem formulation, where the representations of external knowledge [3] and the models of objects to be chosen can be more sophisticated than the decision problem itself. However, attempts to construct automatic models of control systems exist [10], so further research on this problem will follow.

5 Final Remarks and Conclusions

Intelligent autonomous systems, such as forthcoming cognitive online recommenders constitute a new market and social challenge. Their implementation horizon, from the current stage of development, product search and price comparison machines, seems to be on the same level as the expected starting point of implementing new classes of creative DSS outlined in this paper that can create a new market trend. Similar expectations can be formulated as regards graphical and hybrid search engines as well as cooperating mobile robots, clinical support systems and other medical and healthcare applications.

New decision-making concepts and methods which will be developed and covered in further research will make it possible to design intelligent mobile artificial systems that will mainly act autonomously, make discoveries, anticipate the consequences of a decision made (cf. [9], [15]) and enhance the quality of interaction with, possibly remote, humans. As robot decision making will be strongly based on robot vision, merging visual information processing with creative decision support procedures is indispensable. The specific applications of creative robot decision making include mobile space, volcano, and submarine research robots, fire and flood rescue robots, etc. One potential application of the latter class of systems is a decision pilot able to act autonomously where there is a lack of human guidance, as in a flood emergency situation.

The development of computer-aided DSS to date has assumed that users behave in a rational, idealistic way, possessing only limited capacity to check consistency and correct choice. In contrast, the class of intelligent and creative DSS presented here will be able to act autonomously when necessary. Coupled with scene understanding capabilities, when embedded in a mobile system, they will be able to select targets and paths leading to them based on different images compared and analyzed with multicriteria optimization algorithms.

References

1. Etzioni, O., Tuchinda, R., Knoblock, C.A., Yates, A.: To buy or not to buy: mining airfare data to minimize ticket purchase price. In: Proceedings of KDD 2003, pp. 119–128. ACM Press (2003)
2. Ishizaka, A., Lusti, M.: How to Program a Domain Independent Tracer for Explanations. J. Inter. Learning Res. 17(1), 57–69 (2006)
3. Hatzilygeroudis, I., Prentzas, J.: Neuro-symbolic Approaches for Knowledge Representation in Expert Systems. Int. Hybrid J. Intelligent Systems 1(3-4), 111–126 (2004)
4. Kim, S.H.: A formal model of creative decision making. Robotics and Computer-Integrated Manufacturing 8(1), 53–65 (1991)
5. Kohavi, R., John, G.: Wrappers for feature subset selection. Artificial Intelligence 97, 273–324 (1997)
6. Miettinen, K.: Nonlinear Multiobjective Optimization. Kluwer, Dordrecht (1999)
7. Molina, L.C., Belanche, L., Nebot, A.: Feature selection algorithms: A survey and experimental evaluation. In: Proc. of the 2002 IEEE Int. Conf. on Data Mining, pp. 306–313. IEEE Computer Society (2002)
8. Rotter, P., Skulimowski, A.M.J.: Preference extraction in image retrieval. In: Ma, Z. (ed.) Artificial Intelligence for Maximizing Content-based Image Retrieval, pp. 237–262. IGI Global, Hershey (2009)
9. Skulimowski, A.M.J.: Solving Vector Optimization Problems via Multilevel Analysis of Foreseen Consequences. Found. Control Engrg. 10(1), 25–38 (1985)
10. Skulimowski, A.M.J.: Decision Support Systems Based on Reference Sets, Monographs, vol. (40), p. 167. AGH University Publishers, Kraków (1996)
11. Skulimowski, A.M.J.: An Interactive Modification of the Decision Set to Attain a Target Point in Vector Optimization Problems. In: Sawaragi, Y., Inoue, K., Nakayama, H. (eds.) Toward Interactive and Intelligent Decision Support Systems. LNEMS, vol. 285, pp. 142–153. Springer, Heidelberg (1987)
12. Skulimowski, A.M.J.: Methods of Multi-criteria Decision Support Based on Reference Sets. In: Caballero, R., Ruiz, F., Steuer, R.E. (eds.) Advances in Multiple Objective and Goal Programming. LNEMS, vol. 455, pp. 282–290. Springer, Heidelberg (1997)
13. Skulimowski, A.M.J.: Formal models of freedom of choice and cognitive aspects of multicriteria decision support. In: Kłosiński, K.A., Biela, A. (eds.) Human Being and His Decisions, pp. 47–59. KUL Scientific Publishers, Lublin (2009)
14. Skulimowski, A.M.J.: Anticipatory Network Models of Multicriteria Decision-Making Process. In: Proc. of the Intern. Conf. URPDM 2010, Coimbra, p. 11 (April 2010)
15. Sloman, S.A., Hagmayer, Y.: The causal psycho-logic of choice. Trends in Cognitive Sciences 10(9), 407–412 (2006)

A Rule-Based Approach to Activity Recognition

Pitchakan Theekakul[1], Surapa Thiemjarus[1], Ekawit Nantajeewarawat[1], Thepchai Supnithi[2], and Kaoru Hirota[3]

[1] School of Information, Computer, and Communication Technology, Sirindhorn International Institute of Technology, Thammasat University
[2] National Electronics and Computer Technology Center
[3] The Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology

pitchakan.theekakul@student.siit.tu.ac.th, {surapa,ekawit}@siit.tu.ac.th
thepchai.supnithi@nectec.or.th
hirota@hrt.dis.titech.ac.jp

Abstract. This paper presents a rule-based framework for activity classification and illustrates how domain-specific expert knowledge and observation of data in its feature space can be used for rule construction. To demonstrate its practical value, the framework is applied on datasets collected during an orientation-independent activity recognition experiment. Through an implementation based on the Java Expert System Shell (JESS), two types of rules are compared: rules that are specifically constructed for each individual device orientation and those constructed without assuming any prior knowledge on device orientations. Overall accuracy improvements of 7.97% and 9.25% are observed on training and test datasets when orientation-specific rules are used.

Keywords: Rule-based reasoning, body sensor network, activity recognition, device-orientation detection, rule learning.

1 Introduction

Being developed based on innovative wireless wearable and implantable biosensors, Body Sensor Networks (BSNs) [12][4] have found several potential practical applications, particularly in monitoring human health and well-being. A record of daily living activities is important contextual information for indicating a general status of a person. Over the past decade, a success in automatic recognition of activities based on acceleration data has been witnessed in many studies [7][9][5][11][6][10]. Recently, there is an increasing trend of embedding motion sensors into mobile devices such as smart phones. These devices are carried by users on a daily basis, in different locations and orientations, and are thus promising for continuous monitoring of users' daily activities [8]. A gigantic volume of data generated from these sensors raises the need for an automatic activity recognition system.

Rule-based and statistical methods are two main approaches to pattern recognition. With sufficient amount of representative training data, the statistical

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 204–215, 2011.

methods can accurately reflect its true nature. Learning of an accurate recognition model normally involves a laborious and time-consuming task of label tagging. In many BSN-based applications, training data may not be sufficiently available in advance. Some adverse events may rarely occur and labeling of all the collected data often requires an excessive resource-consuming effort. Expert knowledge provides an alternative useful source of information for model construction. The two approaches to pattern recognition are actually closely interrelated in as much as valuable pieces of knowledge are usually formed based on previous experience of events in the past.

This paper presents a rule-based framework for human activity recognition. Based on an experimental study, we illustrate how expert knowledge concerning characteristics of activities under examination and domain-specific constraints can be used along with observation of data in its feature space for construction of orientation-specific rules for classifying activities such as lying, sitting, standing, walking, running, and jumping. The framework is applied on datasets collected during an orientation-independent activity recognition experiment. Using a prototype system implemented based on the Java Expert System Shell (JESS), the obtained orientation-specific rule set is evaluated against a baseline rule set constructed without using device-orientation information.

The paper is organized as follows: Section 2 reviews some related works. Section 3 describes the proposed framework along with our experimental scheme. Section 4 details the construction of rules for orientation detection and that of orientation-specific rules for activity classification. Section 5 presents evaluation results. Section 6 draws conclusions.

2 Related Works

There exist several previous studies on motion-sensor-based activity analysis. Bao and Intille [1] conducted an experiment on activity recognition of users under their semi-natural conditions, where the subjects were asked to perform a sequence of daily living tasks without being told specifically where and how the activities should be performed. Based on mean, energy, frequency-domain entropy, and correlation of the acceleration data, the classification results based on a set of classifiers, namely decision table, instance-based learning, decision tree (C4.5), and naive Bayes classifiers, were compared. In [7], Ravi et al. reported the results of their study on activity recognition using a single tri-axial accelerometer worn near the pelvic region. Four features extracted from the accelerometer data were mean, standard deviation, energy, and correlation. The results showed that combining classifiers with plurality voting provided the best result and activities such as standing, walking, and vacuuming can be recognized with fairly high accuracy using a single tri-axial accelerometer. Pirttikangas et al. [6] performed activity recognition using coin-sized sensor devices attached to four different parts of a human body. They calculated several features from acceleration and heart rate data, then per-formed feature selection using forward-backward search. For classification, kNN and multilayer perceptron were compared and kNN yielded better performance.

There exist several previous studies on ontology and rules. Hatala et al. [3] presented a design and implementation of an augmented audio reality system for museum visitors, called ec(h)o. They focused on the practical evaluation of challenges and capabilities of combination of ontologies and rules in the context of real time ubiquitous application. The ec(h)o system is a proof that ontologies and rules provide an excellent platform for building a highly-responsive context-aware interactive application. Yang et al. [13] developed a JESS-enabled context elicitation system featuring an ontology-based con-textual model that formally described and acquired contextual information pertaining to service requesters and Web services. Gomez and Laube [2] proposed a contextual ontology for enhancing common middleware processing. The purpose of this ontology is to provide necessary information that matches an interest of a business application and to dynamically process information from wireless sensor network in order to provide higher semantic-level information for the application.

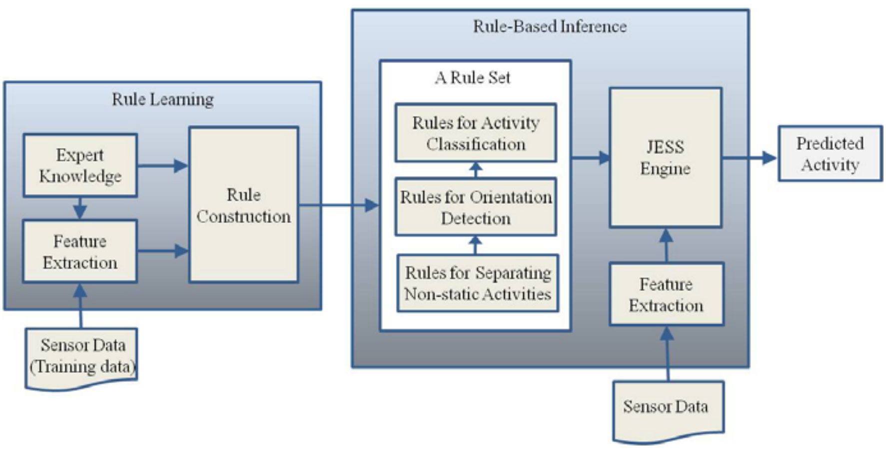


Fig. 1. An overview of our rule-based activity classification framework

3 Framework and Experimental Scheme

As outlined in Fig. 1, our proposed rule-based activity classification framework consists of two main parts: rule learning and rule-based inference. Domain-specific expert knowledge is used along with observation of training sensor data for construction of classification rules. Such knowledge includes assumptions on characteristics of activities under examination and constraints on device packaging design and on the way a device is worn by a user. For example, with the assumption that most non-static activities under consideration are generated by bipedal locomotion (e.g., standing, walking, running, jumping, and transitional activities such as sit-to-stand), which does not involve significant change of a device orientation, it is anticipated that signals from non-static activity portions provide appropriate features for prediction of device orientations [8]. Likewise, when a device orientation is known, the assumption that most activities in this study involve motion mostly in the upward and forward directions is useful for selecting an appropriate axis along which features should be extracted for classification of certain activities. A constraint on sensor packaging may also be used; for example, providing that a device can be rotated only around a certain axis, the information along that axis can be known to be irrelevant to a certain classification task. An obtained rule set is implemented in the JESS rule language, and the JESS rule engine, developed at Sandia National Laboratories, is employed.

In our experimental study, a 3D accelerometer placed on a belt-clip Pulse Oximetry (SpO2) sensor-an integrated sensing device developed by Imperial College based on the Body Sensor Network (BSN) [12] [4] platform was employed at a sampling rate of 50 Hz. Six activities, i.e., lying, sitting, standing, walking, running, and jumping, were performed by thirteen subjects (aged between 16-23 years old). Each subject performed each activity twice for each of the four sensor device orientations (i.e., up, down, left, and right). A total of 104 data segments were obtained, each of which contains approximately 5 seconds for each activity. An observer annotated the data in real-time by clicking a button on the PC-side of the data collection program to record the timestamps of activity transitions. To compensate for potential response delays, the timestamps were shifted forward by 0.6 second.

Fig. 2 provides the domain specification for the experiment. Activities are divided into static activities, i.e., lying, sitting, and standing, and non-static activities, which include dynamic activities, i.e., walking, running, and jumping, and transition activities. Fig. 3 illustrates the coordinate systems for the four device orientations, referred to as O1 to O4. The orientation O1 was used as the reference coordinate system (i.e., $0\,^\circ$). The orientations O2 to O4 result from the clockwise rotation of the device by $90\,^\circ$, $180\,^\circ$, and $270\,^\circ$, respectively.

4 Rule Learning Experiment

Orientation Detection. Rules for predicting device orientations are constructed based on the orientation detection method presented in [8]. First, normalization by mean and standard deviation is applied for preprocessing. Non-static signal portions are then separated from static signal portions based on the deviation magnitude ($|\sigma(i)|$), which is calculated over a shifted window of fixed sized, i.e.,

$$|\sigma(i)| = \sqrt{\sigma_x^2(i) + \sigma_y^2(i) + \sigma_z^2(i)}. \qquad (1)$$

where, $\sigma_x(i)$, $\sigma_y(i)$, $\sigma_z(i)$ are the standard deviations of the acceleration signals along the x-, y-, and z-axes, respectively, of a window i. A fixed window size of 1 second and a shifted window size of 0.5 second are used. If the obtained deviation magnitude value is greater than a threshold of 0.25 times the overall deviation magnitude of the signals calculated over the entire data file, then the signal window is classified as non-static; otherwise it is classified as static. Referring to Fig. 3, based on the sensor packaging design, a sensor device in

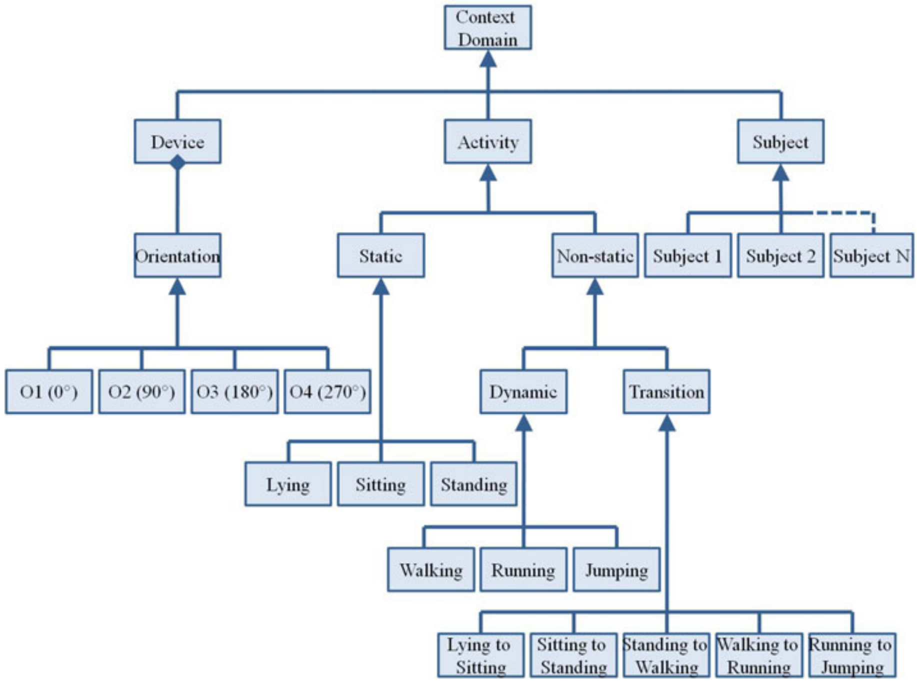


Fig. 2. A domain specification

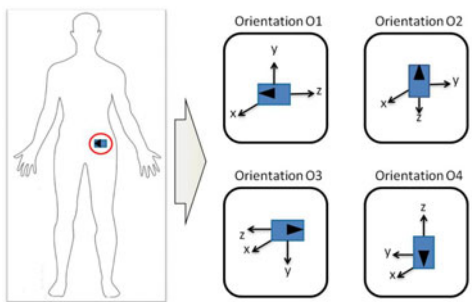


Fig. 3. The coordinate systems for the four device orientations O1-O4

use is always rotated around the x-axis regardless of its orientation. Acceleration information along the x-axis is therefore independent of how the device is moved. Consequently, normalized signal values along the y- and z-axes are used for classifying device orientations. The mean values over a combined non-static signal segment along these two axes are calculated.

Fig. 4 shows a 2D plot of obtained mean values. As indicated by different shapes and colors of the data points, these features can clearly separate the four device orientations, using the two linear equations:

$$z = -0.5y \tag{2}$$

$$z = y \tag{3}$$

Fig. 5 shows the resulting rules for separating non-static activities from static ones and those for predicting device orientations, where $|\sigma|$ is the acceleration deviation magnitude of a window, $|\sigma_{total}|$ is the total deviation magnitude, and μ_y, μ_z are the mean acceleration values along the y- and z-axes.

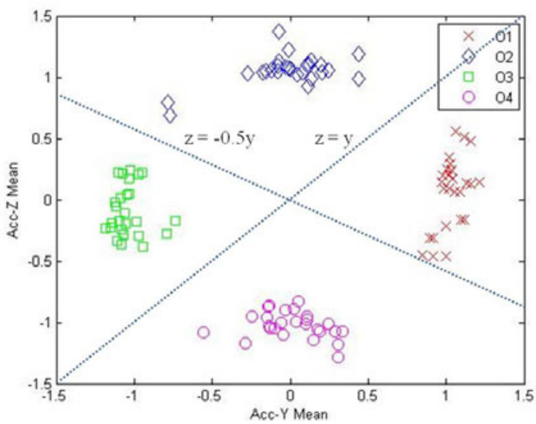


Fig. 4. A 2D plot of calculated mean values and linear equations for orientation classification

Activity Classification. Mean values of normalized signals along the x-, y-, and z-axes over a fix-sized window are used as features for classifying static activities (i.e., lying, sitting, and standing), while standard deviation values of those along the three axes are used for classifying dynamic activities (i.e., walking, running, and jumping). With the expert knowledge that most activities in our experiments involve motion in only the upward and forward directions (e.g., when a subject changes his/her activity from lying to sitting, the signals in the sideways direction do not change), signal values in these two directions will be employed for activity classification. Consequently, the choice of the axes along which features should be extracted is influenced by the orientation of a device. From the coordinate systems in Figure 3, the upward direction corresponds to

$$(|\sigma| \leq 0.25|\sigma_{total}|) \Rightarrow (action = static)$$
$$(|\sigma| > 0.25|\sigma_{total}|) \Rightarrow (action = non-static)$$
$$(action = non-static), (\mu_z > -0.5\mu_y), (\mu_z \leq \mu_y) \Rightarrow (orientation = O1)$$
$$(action = non-static), (\mu_z > -0.5\mu_y), (\mu_z > \mu_y) \Rightarrow (orientation = O2)$$
$$(action = non-static), (\mu_z \leq -0.5\mu_y), (\mu_z > \mu_y) \Rightarrow (orientation = O3)$$
$$(action = non-static), (\mu_z \leq -0.5\mu_y), (\mu_z \leq \mu_y) \Rightarrow (orientation = O4)$$

Fig. 5. Rules for separating non-static activities from static activities and rules for predicting device orientations

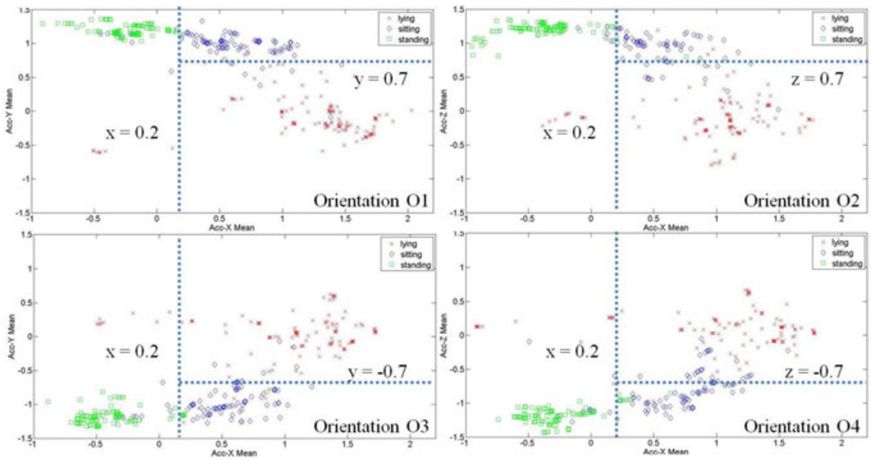


Fig. 6. Mean values of the static activities lying, sitting, and standing in the orientations O1, O2, O3, and O4

the y-axis for the orientations O1 and O3 and corresponds to the z-axis for the orientations O2 and O4, whereas the forward direction corresponds to the x-axis for each of the four orientations.

Fig. 6 shows the selected features (i.e., mean values) for each static activity in each orientation. Based on the figure, thresholds for classifying static activities are determined. For example, for orientation O1, the signal values are classified as lying if the mean along the x-axis is greater than 0.2 and that along the y-axis is less than 0.7; they are classified as standing if the mean along the x-axis is not greater than 0.2; and they are classified as sitting otherwise. Fig. 7 shows the selected features (i.e., standard deviation values) for each dynamic activity in each orientation. Again, based on this figure, thresholds for classifying dynamic activities are determined. Using the obtained thresholds, orientation-specific rules for activity classification are constructed. The resulting rules are shown in Fig. 8, where μ_x, μ_y, μ_z are the acceleration means along the x-, y-, and z-axes, and σ_x, σ_y, σ_x are the acceleration standard deviations along the x-, y-, and z-axes.

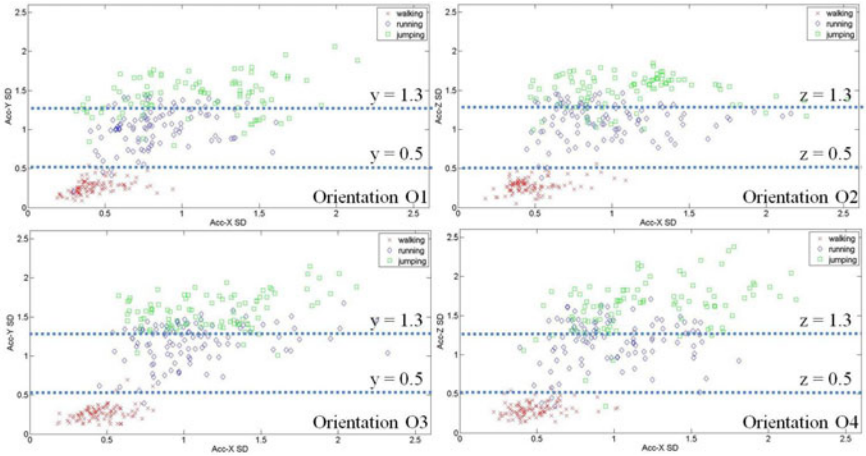


Fig. 7. Standard deviation values of the dynamic activities walking, running, and jumping in the orientations O1, O2, O3, and O4

5 Evaluation

As a baseline for evaluating the rules obtained in the previous section, rules for activity classification without additional knowledge about device orientations are constructed. Fig. 9 shows mean values along the x- and y-axes, which are used to determine thresholds for classifying static activities. In the absence of prior information on device orientations, deviation magnitudes, which combine standard deviation along the three axes, are used for classifying dynamic activities. Fig. 10 shows deviation magnitude values and selected thresholds for the classification. Using these thresholds, baseline activity-classification rules are obtained and are shown in Fig. 11, where μ_x is the acceleration mean along the x-axis, $|\sigma|$ is the acceleration deviation magnitude of a window, and $|\sigma_{total}|$ is the total deviation magnitude.

The orientation-specific rules in Fig. 8 and the baseline rules in Fig. 11 are applied on the training and test datasets. Table 1 and Table 2 show the confusion metric of the resulting classification using the baseline rules and that of the resulting classification using orientation-specific rules, respectively. Table 3 presents overall activity-classification accuracy on the training and test datasets using the two rule sets. The overall accuracy in each column is calculated as the average of the class-specific accuracy to eliminate the problem of unbalanced amount of data in different specific activities. The overall accuracy values obtained by applying the baseline rules to the training and test datasets are 61.50% and 61.79%, respectively, while those obtained by applying the orientation-specific rules to the two datasets are 76.43% and 74.46%, respectively.

$(action = static), (orientation = O1), (\mu_x > 0.2), (\mu_y \leq 0.7) \Rightarrow (action = lying)$
$(action = static), (orientation = O2), (\mu_x > 0.2), (\mu_z \leq 0.7) \Rightarrow (action = lying)$
$(action = static), (orientation = O3), (\mu_x > 0.2), (\mu_y > -0.7) \Rightarrow (action = lying)$
$(action = static), (orientation = O4), (\mu_x > 0.2), (\mu_z > -0.7) \Rightarrow (action = lying)$
$(action = static), (orientation = O1), (\mu_x > 0.2), (\mu_y > 0.7) \Rightarrow (action = sitting)$
$(action = static), (orientation = O2), (\mu_x > 0.2), (\mu_z > 0.7) \Rightarrow (action = sitting)$
$(action = static), (orientation = O3), (\mu_x > 0.2), (\mu_y \leq -0.7) \Rightarrow (action = sitting)$
$(action = static), (orientation = O4), (\mu_x > 0.2), (\mu_z \leq -0.7) \Rightarrow (action = sitting)$
$(action = static), (\mu_x \leq 0.2) \Rightarrow (action = standing)$
$(action = non - static), (orientation = O1), (\sigma_y \leq 0.5) \Rightarrow (action = walking)$
$(action = non - static), (orientation = O2), (\sigma_z \leq 0.5) \Rightarrow (action = walking)$
$(action = non - static), (orientation = O3), (\sigma_y \leq 0.5) \Rightarrow (action = walking)$
$(action = non - static), (orientation = O4), (\sigma_z \leq 0.5) \Rightarrow (action = walking)$
$(action = non - static), (orientation = O1), (0.5 < \sigma_y \leq 1.3) \Rightarrow (action = running)$
$(action = non - static), (orientation = O2), (0.5 < \sigma_z \leq 1.3) \Rightarrow (action = running)$
$(action = non - static), (orientation = O3), (0.5 < \sigma_y \leq 1.3) \Rightarrow (action = running)$
$(action = non - static), (orientation = O4), (0.5 < \sigma_z \leq 1.3) \Rightarrow (action = running)$
$(action = non - static), (orientation = O1), (\sigma_y > 1.3) \Rightarrow (action = jumping)$
$(action = non - static), (orientation = O2), (\sigma_z > 1.3) \Rightarrow (action = jumping)$
$(action = non - static), (orientation = O3), (\sigma_y > 1.3) \Rightarrow (action = jumping)$
$(action = non - static), (orientation = O4), (\sigma_z > 1.3) \Rightarrow (action = jumping)$

Fig. 8. Orientation-specific rules for activity classification

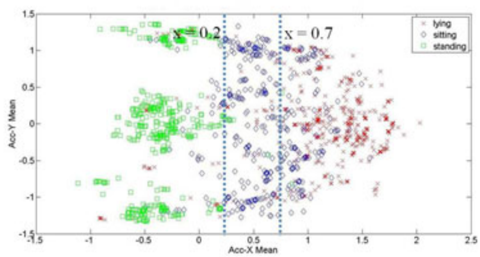


Fig. 9. Mean values for classifying lying, sitting, and standing activities

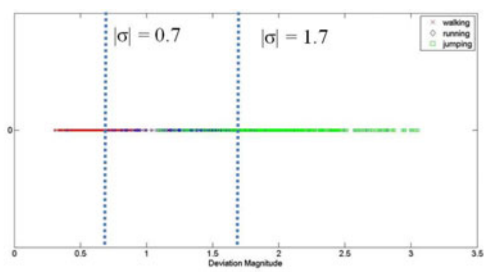


Fig. 10. Deviation magnitude values for classifying walking, running, and jumping activities

$(|\sigma| \leq 0.25|\sigma_{total}|), (\mu_x > 0.7) \Rightarrow (action = lying)$
$(|\sigma| \leq 0.25|\sigma_{total}|), (0.2 < \mu_x \leq 0.7) \Rightarrow (action = sitting)$
$(|\sigma| \leq 0.25|\sigma_{total}|), (\mu_x \leq 0.2) \Rightarrow (action = standing)$
$(|\sigma| > 0.25|\sigma_{total}|), (|\sigma| \leq 0.7) \Rightarrow (action = walking)$
$(|\sigma| > 0.25|\sigma_{total}|), (0.7 < |\sigma| \leq 1.7) \Rightarrow (action = running)$
$(|\sigma| > 0.25|\sigma_{total}|), (|\sigma| > 1.7) \Rightarrow (action = jumping)$

Fig. 11. Baseline rules for activity classification

Table 1. The confusion metric of activity classification using baseline rules

	Training Dataset						
	Predicted						
	Class	Lying	Sitting	Standing	Walking	Running	Jumping
Actual	Lying	279	14	37	129	37	0
	Sitting	96	166	20	79	11	0
	Standing	0	8	358	25	6	0
	Walking	0	0	8	298	100	0
	Running	0	0	0	2	255	151
	Jumping	0	0	0	0	60	311

	Test Dataset						
	Predicted						
	Class	Lying	Sitting	Standing	Walking	Running	Jumping
Actual	Lying	256	62	21	119	36	0
	Sitting	120	127	45	85	7	0
	Standing	0	15	352	24	5	0
	Walking	0	0	8	297	95	0
	Running	0	0	0	2	241	155
	Jumping	0	0	0	0	65	302

Table 2. The confusion metric of activity classification using orientation-specific rules

	Training Dataset						
	Predicted						
	Class	Lying	Sitting	Standing	Walking	Running	Jumping
Actual	Lying	292	1	37	156	10	0
	Sitting	37	225	20	84	6	0
	Standing	0	8	358	31	0	0
	Walking	0	0	8	386	12	0
	Running	0	0	0	9	283	116
	Jumping	0	0	0	1	56	314

	Test Dataset						
	Predicted						
	Class	Lying	Sitting	Standing	Walking	Running	Jumping
Actual	Lying	315	3	21	140	15	0
	Sitting	50	197	45	90	2	0
	Standing	2	13	352	27	2	0
	Walking	0	0	6	379	13	0
	Running	0	0	0	8	275	115
	Jumping	0	0	0	2	77	288

Table 3. Overall activity-classification accuracy on the training and test datasets using the two schemes of rule construction

Class	Baseline Rules		Orientation-Specific Rules	
	Training	Test	Training	Test
Lying	56.25%	51.82%	58.87%	63.77%
Sitting	44.62%	33.07%	60.48%	51.30%
Standing	90.18%	88.89%	90.18%	88.89%
Walking	73.40%	74.62%	95.07%	95.23%
Running	62.50%	60.55%	69.36%	69.10%
Jumping	83.83%	82.29%	84.64%	78.47%
Overall	68.46%	65.21%	76.43%	74.46%

6 Conclusions

This paper demonstrates a rule-based approach to activity recognition for a single 3D accelerometer-based motion analysis. Training and test datasets were collected from thirteen subjects, performing six activities with four different device orientations. The proposed method is based on the assumption that device-orientation detection is useful for activity classification. Using domain-specific expert knowledge, construction of rules for detection of device orientations along with orientation-specific rules for activity classification is described. The obtained rules are implemented in the JESS rule language. With derived information on device orientations, the overall accuracy of 76.43% and 74.46% is achieved on the training and test datasets, which is approximately 7.97% and 9.25% higher, respectively, compared to classification without orientation detection. Further work includes integration of other kinds of expert knowledge, e.g., subject-specific information and constraints on possible time periods in which activity changes may take place, for enhancement of prediction accuracy, and also application of the framework to classification of other kinds of contextual information.

Acknowledgments. This research is financially supported by Thailand Advanced Institute of Science and Technology (TAIST), National Science and Technology Development Agency (NSTDA), Tokyo Institute of Technology and Sirindhorn International Institute of Technology (SIIT), Thammasat University (TU). It is also supported in part by the Thailand Research Fund (TRF). The SpO2 sensor used in this study are funded by Telecommunication Research and Industrial Development Institute (TRIDI) and the National Telecommunications Commission (NTC). Thanks to Apiwat Henprasettae for the graphical illustration.

References

1. Bao, L., Intille, S.S.: Activity recognition from user-annotated acceleration data. In: Ferscha, A., Mattern, F. (eds.) PERVASIVE 2004. LNCS, vol. 3001, pp. 1–17. Springer, Heidelberg (2004)
2. Gomez, L., Laube, A.: Ontological middleware for dynamic wireless sensor data processing. In: 3rd International Conference on Sensor Technologies and Applications, pp. 145–151 (2009)
3. Hatala, M., Wakkary, R., Kalantari, L.: Rules and ontologies in support of real-time ubiquitous application. Web Semantics: Science, Services and Agents on the World Wide Web 3, 5–22 (2005)
4. Lo, B.P.L., Thiemjarus, S., King, A., Yang, G.Z.: Body sensor network - a wireless sensor platform for pervasive healthcare monitoring. In: 3rd International Coference on Pervasive Computing, Munich, Germany, pp. 77–80 (2005)
5. Mantyjarvi, J., Himberg, J., Seppanen, T.: Recognizing human motion with multiple acceleration sensors. In: IEEE International Conference on Systems, Man, and Cybernetics, Hawaii, USA, pp. 747–752 (2001)
6. Pirttikangas, S., Fujinami, K., Nakajima, T.: Feature selection and activity recognition from wearable sensors. In: Youn, H.Y., Kim, M., Morikawa, H. (eds.) UCS 2006. LNCS, vol. 4239, pp. 516–527. Springer, Heidelberg (2006)
7. Ravi, N., Dandekar, N., Mysore, P., Littman, M.L.: Activity recognition from accelerometer data. In: 17th Conference on Innovative Applications of Artificial Intelligence (2005)

8. Thiemjarus, S.: A device-orientation independent method for activity recognition. In: International Conference on Body Sensor Networks, Biopolis, Singapore, pp. 19–23 (2010)
9. Thiemjarus, S., Lo, B.P.L., Yang, G.Z.: A spatio-temporal architecture for context-aware sensing. In: 3rd International Workshop on Wearable and Implantable Body Sensor Networks, Cambridge, MA, pp. 191–194 (2006)
10. Van Laerhoven, K., Cakmakci, O.: What shall we teach our pants? In: 4th IEEE International Symposium on Wearable Computers, Atlanta, GA, USA, pp. 77–83 (2000)
11. Veltink, P.H., Bussmann, H.B., de Vries, W., Martens, W.L., Van Lummel, R.C.: Detection of static and dynamic activities using uniaxial accelerometers. IEEE Transactions on Rehabilitation Engineering 4, 375–385 (1996)
12. Yang, G.Z.: Body Sensor Networks. Springer, London (2006)
13. Yang, S.J.H., Zhang, J., Chen, I.Y.L.: A jess-enabled context elicitation system for providing context-aware web services. Expert Systems with Applications 34, 2254–2266 (2008)

Simultaneous Character-Cluster-Based Word Segmentation and Named Entity Recognition in Thai Language

Nattapong Tongtep and Thanaruk Theeramunkong

School of Information, Computer, and Communication Technology,
Sirindhorn International Institute of Technology, Thammasat University
131 Moo 5, Tiwanont Rd., Bangkadi, Muang, Pathum Thani, Thailand 12000
{nattapong,thanaruk}@siit.tu.ac.th
http://www.siit.tu.ac.th

Abstract. Named entity recognition in inherent-vowel alphabetic languages such as Burmese, Khmer, Lao, Tamil, Telugu, Bali, and Thai, is difficult since there are no explicit boundaries among words or sentences. This paper presents a novel method to exploit the concept of character clusters, a sequence of inseparable characters, to group characters into clusters, utilize statistics among characters and their clusters to extract Thai words and then recognize named entities, simultaneously. Integrated of two phases, the word-segmentation model and the named-entity-recognition model, context features are exploited to learn parameters for these two discriminative probabilistic models, i.e., CRFs, to rank a set of word and named entity candidates generated. The experimental result shows that our method significantly increases the performance of segmenting word and recognizing entities with the F-measure of 96.14% and 83.68%, respectively.

Keywords: Named Entity Recognition, Word Segmentation, Character Cluster, Information Extraction.

1 Introduction

Nowadays, the growing amount of textual information is available in various kinds of formats, especially in digital one. To utilize such information, Information Extraction (IE) plays an important role to classify, categorize, or transform relevant information from unstructured text documents into a proper format. IE typically involves four steps: named entity extraction, relation extraction [12], co-reference resolution, and slot filling. Extracting named entities is recognized as one of the most important tasks of IE, usually called Named Entity Recognition and Classification (NERC) [7]. Normally, in the recognition step, the boundaries of similar entity occurrences are determined while the types of entities are assigned to segmented entities such as person name, date, time, organization, and money expressions in the step of classification.

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 216–225, 2011.

In the past, a number of techniques were developed to identify named entities (e.g., product names, phone numbers, accident types, and terrorist levels) in several written languages (e.g., English [14] and Thai [9]) and various domains (e.g., biomedical [5] and news [13]). These techniques include pattern-based extraction [13], hidden Markov models (HMMs) [14], and conditional random fields (CRFs) [5]. While several algorithms have been proposed for this task in segmental alphabetic languages like English, NER remains a challenging task especially in inherent-vowel alphabetic language such as Burmese, Khmer, Lao, Tamil, Telugu, Bali, and Thai. In these languages, NER is particularly difficult since there is no explicit word boundary and words are formed by a sequence of contiguous characters. More seriously, some of these languages, such as Thai, have no sentence boundary. Since there are no explicit boundaries among words or sentences in Thai, word segmentation is an important preprocess. In the past, most NER methods utilized word segmentation to transform a running text to a sequence of words before detecting which words are likely to be an NE (word-based model). Due to this characteristic, the performance of NER strongly depends on the quality of word segmentation. As more recent works, there have been a number of attempts to detect NEs from a running text directly without segmenting into words (character-based model). However, this model may face with performance tradeoff since word information may be important in detecting NEs.

In this paper, we propose a method to utilize TCCs (Thai Character Clusters) which are units of contiguous inseparable characters but mostly smaller than words to detect NEs (character-cluster-based model). In this work, CRFs is occupied as our classification model for word segmentation (WS) and named entity recognition (NER). Besides TCCs, our method uses the type or group of TCCs as additional features for WS and NER. The remaining part of this paper is organized as follows: Section 2 describes our method in detail. Experiment and results are analyzed in Section 3. Finally, conclusion and future work are given in Section 4.

2 The Proposed Method

2.1 Thai Writing System with Hierarchical Structure Representation (HSR)

In this section, we first describe the characteristics of the Thai writing system and represent them in a hierarchical structure. As an alphabetic language, graphemes (characters or letters) in Thai represent the words (phonemes). Known as an inherent-vowel alphabetic language (also called syllabic alphabet, alphasyllabary or abugida [1]), Thai consists of symbols for consonants and vowels, the same as Burmese, Khmer, Lao, Tamil, Telugu, and Bali. Unlike English (segmental alphabetic language) and Arabic or Hebrew (consonantal alphabetic language or abjad [1]), vowels and consonants in Thai are treated separately, each Thai consonant has an inherent vowel which can be changed to another vowel or muted by means of diacritics and Thai vowel marking is almost mandatory.

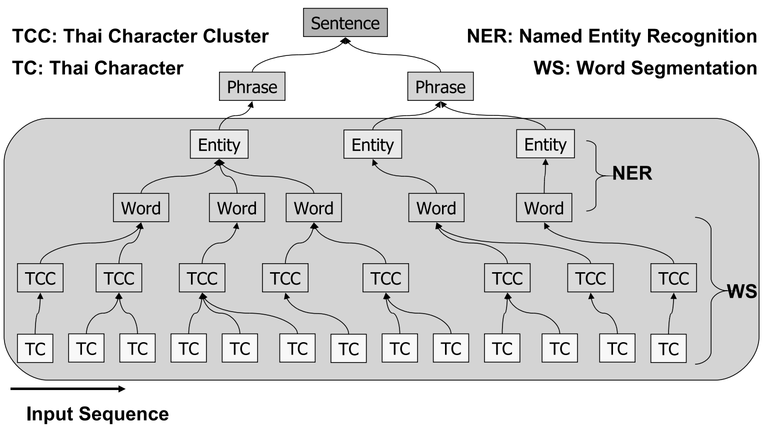


Fig. 1. Graphical structures of hierarchical structure representation (HSR)

The Thai language consists of 44 consonants, 21 vowel symbols, 4 tone markers for its 5 tonal levels, and a number of punctuation marks. Thai writing system is left-to-right direction, without spaces between words and no uppercase and lowercase characters. Vowels can be written before, after, above, or below consonants, while all tone marks, and diacritics are written above and below the main character. A Thai word is typically formed by the combination of one or more consonants, one vowel, one tone mark, and one or more final consonants to make one syllable. Thai verbs do not change their forms as with tense, gender, and singular or plural form, but there are additional words to help with the meaning. Thai has no distinct boundary maker between words and sentences, like space and a full stop in English. Koanantakool et al. [3] elucidated the history of Thai language and analyzed characteristics of Thai characters involved in Thai text processing in detail.

Analyzing the formation of word, phrase, and sentence in Thai language, we found that relations among them can be illustrated by a hierarchical structure representation (HSR) shown in Figure 1.

2.2 Character Cluster and Its Type

This section presents a character cluster which is an unambiguous unit smaller than or equals to a word, larger than or equals to a character, and cannot be further divided. First introduced by Theeramunkong et al. [10], Thai Character Clusters (TCCs) are used to improve searching accuracy by stating which character cannot be separated from another character. Later the concept of character clusters was also adapted to use in several languages such as Burmese [11] and Khmer [2].

The Thai character clusters can easily be recognized by applying a set of rules. For example, based on Thai language spelling features, a front vowel and the next character have to be grouped into a same unit. A tonal mark is always

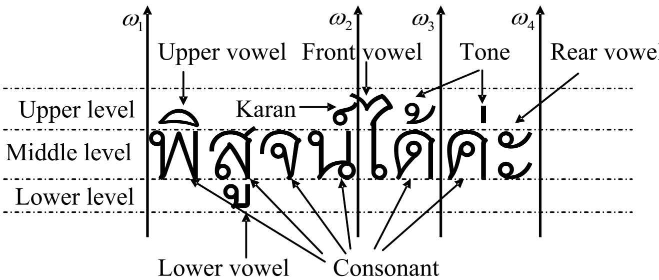


Fig. 2. An example of Thai characters

located above a consonant and cannot be separated from the consonant. A rear vowel and the previous character have to be grouped into a same unit. Besides the character cluster, this work originally proposes a so-called Thai character cluster type (TCCT). An internal structure of a TCC is used to identify the type of that TCC, such as the type of C+VF means Consonant + Front Vowel. This work develops 26 types of Thai character cluster (TCCT) as shown in Table 1.

Figure 2 illustrates an example of a Thai phrase with three words. In this example, there are three levels and seven types of characters: upper/lower/front/rear vowels, consonant, tone, and karan (a pronunciation deletion character). ω_1, ω_2, ω_3, and ω_4 indicate word boundaries. TCCs and TCCTs are |พิ|สู|จ|น์|ได้|ค่ะ| and |T3|T4|C|TK|T1T|T2T|, respectively.

Table 1. Type of character cluster

Symbol	**Type**	**Example**	**Symbol**	**Type**	**Example**
T1	C+VF	ใอ ไอ เอ แอ โอ	T1T	C+VF+T	T1+T
T2	C+VB	อะ อำ อา	T2T	C+VB+T	T2+T
T3	C+VU	อั- อ็ อํ อิ อึ อี อื	T3T	C+VU+T	T3+T
T4	C+VL	อุ, อู	T4T	C+VL+T	T4+T
TT	C+T	อ่ อ้- อ๊- อ๋-	TK	C+Karan	อ์
T6	C+VF+VB	เอะ แอะ โอะ เอาะ	T6T	C+VF+VB+T	T6+T
T7	C+VF+VU	เอิ-	T7T	C+VF+VU+T	T7+T
T8	C+VB+VU	อัวะ, อัว	T8T	C+VB+VU+T	T8+T
T9	C+VF+VB+VU	เอียะ เอือะ เอีย	T9T	C+VF+VB+VU+T	T9+T
S	Space	<space>	V	Vowel	เ แ โ ใ ไ
D	Digit	๐-๙, 0-1	P	Punctuation	@#$%&*ฯๆ
C	Consonant	ก-ฮ	T	Tone	อ่ อ้ อ๊ อ๋
E	English	a-z, A-Z	O	Other	

2.3 Phase-Merging Discriminative Probabilistic Model

Since the input of hierarchical structure representation (HSR) is a sequence, this paper utilizes a discriminative model so called conditional random fields (CRFs). CRFs is the state-of-the-art algorithm in sequence labeling and has shown good performance in a natural language processing task, firstly introduced by Lafferty et al [4]. In NER task, a linear-chain CRFs is often used based on an undirected graph $G = (V, E)$, where V is the set of states $Y = \{y_t | 1 \leq t \leq T\}$ for a given length-T input sequence $X = \{x_1, ..., x_t\}$ and $E = \{(y_{t-1}, y_t) | 1 \leq t \leq T\}$ is the set of $T - 1$ edges in linear chain. The conditional probability of the state sequence Y for a given input sequence X is

$$P(Y|X) = \frac{1}{Z(X)} \exp\left\{ \sum_t \sum_k \lambda_k f_k(y_{t-1}, y_t, X, t) \right\}, \tag{1}$$

$$Z(X) = \sum_Y \exp\left\{ \sum_t \sum_k \lambda_k f_k(y_{t-1}, y_t, X, t) \right\}, \tag{2}$$

where $Z(X)$ in Equation (2) is a normalization factor of all state sequences, $f_k(y_{t-1}, y_t, X, t)$ is the feature function, and λ_k is the feature's weight, y_t and y_{t-1} refer to the current state and the previous state, respectively. The training process is to find the weights that maximize the log likelihood of all instances in training data.

Given formula (1), we adopt L-BFGS [6] (a limited-memory quasi-Newton code for unconstrained optimization) which is a numerical optimization technique to find its maximum value. Once CRFs model has been trained, Viterbi or forward-backward algorithm can be derived to do inference for a test sequence.

In this paper, we propose phase-merging discriminative probabilistic model for segmenting word and recognizing named entities simultaneously.

Phase-Merging (PM). The conditional probability of the state sequence Y for a given input sequence X is equal to Equation (1), where $X = \{x_1, x_2, ..., x_T\}$ is a sequence of TCC inputs and $Y = \{y_1, y_2, ..., y_T\}$ is a sequence of corresponding word and named entity predictions for simultaneous word segmentation (WS) and named entity recognition (NER).

CRFSuite implemented by Okazaki [8] is exploited in this work.

2.4 Labeling

Word segmentation based on hierarchical structure representation (HSR) uses BIES labeling. 'B' is the beginning of character cluster of a word, 'I' is the inside of character cluster of a word, 'E' is the ending of character cluster of a word, and 'S' is a single-character-cluster word. BIES-NE labeling is used for named entity recognition. 'B-NE' is the beginning of character cluster of a named entity, 'I-NE' is the inside of character cluster of a named entity, 'E-NE' is the ending of character cluster of a named entity, and 'S-NE' is a single-character-cluster named entity.

Table 2. Feature

Feature	TCC	TCCT
Unigram	tcc_{t-2}	$tcct_{t-2}$
	tcc_{t-1}	$tcct_{t-1}$
	tcc_t	$tcct_t$
	tcc_{t+1}	$tcct_{t+1}$
	tcc_{t+2}	$tcct_{t+2}$
Bigram	tcc_{t-2}/tcc_{t-1}	$tcct_{t-2}/tcct_{t-1}$
	tcc_{t-1}/tcc_t	$tcct_{t-1}/tcct_t$
	tcc_t/tcc_{t+1}	$tcct_t/tcct_{t+1}$
	tcc_{t+1}/tcc_{t+2}	$tcct_{t+1}/tcct_{t+2}$
Trigram	$tcc_{t-2}/tcc_{t-1}/tcc_t$	$tcct_{t-2}/tcct_{t-1}/tcct_t$
	$tcc_{t-1}/tcc_t/tcc_{t+1}$	$tcct_{t-1}/tcct_t/tcct_{t+1}$
	$tcc_t/tcc_{t+1}/tcc_{t+2}$	$tcct_t/tcct_{t+1}/tcct_{t+2}$

2.5 Feature

In this work, three sets of features are investigated: character cluster (TCC), character cluster type (TCCT), and combination of character and its type (TCC+TCCT). Unigram, bigram and trigram in each type of feature listed in Table 2.

3 Experiment

3.1 Corpus

Our experiments are all based on a dataset from Benchmark for Enhancing the Standard of Thai language processing (BEST2010)[1]. The BEST2010 dataset consists of five million words in four categories: article, encyclopedia, news, and novel. Since the sentences from news are closer to Thai common usage, only category of news is focused in this work.

The BEST2010 dataset was firstly used in Thai Word Segmentation Software Contest, which is a special topic in the National Software Contest 2010 (NSC2010)[2]. Competitors received the BEST2010 dataset which consists of seven million words in eight categories: article, encyclopedia, news, novel, Buddhism, law, talk, and Wikipedia. The BEST2010 data set is annotated manually based on the BEST2010 guidelines. The basic statistics information of the data set used in our work is summarized in Table 3.

Table 3. Composition of BEST2010 data set

BEST2010	Category	#Words	#TCCs	#NEs
data set#1	news	1,426,677	3,102,897	65,760
data set#2	news	282,334	610,816	13,525
data set#3	news, article, novel, encyclopedia	≈100,000	264,530	N/A

[1] http://www.hlt.nectec.or.th/best/?q=node/3

[2] http://fic.nectec.or.th/nsc12

Table 4. Experimental setting

Task	Setting	Training data	Testing data
NER	L-Train	data set#1	data set#2
WS	L-Train	data set#1	data set#2
	WS-1	data set#1	data set#3

3.2 Experimental Setting and Evaluation Criteria

The standard F_β measure is used to evaluate our method, can be calculated as follows

$$F_\beta = \frac{(\beta^2 + 1) \times Precision \times Recall}{\beta^2 \times (Precision + Recall)}, \tag{3}$$

where

$$Precision = \frac{Corr}{OutputWNE}, \tag{4}$$

and

$$Recall = \frac{Corr}{RefWNE}. \tag{5}$$

In word segmentation, $Corr$ is the number of words in the output that is correctly segmented, $OutputWNE$ is the total number of words in the output, and $RefWNE$ is the total number of words in the reference. In named entity recognition, $Corr$ is the number of named entities in the output that is correctly segmented, $OutputWNE$ is the total number of named entities in the output, and $RefWNE$ is the total number of named entities in the reference. In this paper, we use $F = F_{\beta=1}$.

Table 4 shows our experimental setting. In WS and NER tasks with L-Train setting, the training set (data set#1) and the testing set (data set#2) are from the news category. In WS task with WS-1 setting, the training set is from the news category, but the testing set (data set#3) which consists of one hundred thousand words approximately, is obtained from four categories (news, article, novel, and encyclopedia). The scoring program provided on the BEST2010 website[3] will report the overall F-measure calculated from the entire test set (data set#3) for word segmentation only.

3.3 Experimental Result

The performance of our method for word segmentation task including the effect of different feature sets are shown in Table 5 and Table 6. The performance of named entity recognition task and the effect of different feature sets are compared as shown in Table 7.

[3] http://www.hlt.nectec.or.th/Evaluation

Table 5. Performance of the word segmentation (WS) task with L-Train setting

Feature	Precision(%)	Recall(%)	$F_{\beta=1}$(%)
TCC	96.34	95.96	96.14
TCCT	75.67	62.78	68.38
TCC+TCCT	96.32	95.77	96.04

Table 6. Performance of the word segmentation (WS) task with WS-1 setting

Feature	Precision(%)	Recall(%)	$F_{\beta=1}$(%)
TCC	92.76	90.19	91.46
TCCT	65.59	58.42	61.80
TCC+TCCT	92.71	90.33	91.51

From Table 5, our word segmentation system achieves the F-measure of 96.14% using only TCC. Using TCC+TCCT is slightly reduces the F-measure 0.10%, 0.19% for recall, and 0.02% for precision. From Table 6, our word segmentation system achieves the F-measure of 91.51% using TCC and TCCT. TCC+TCCT slightly increases the F-measure 0.05% from 90.19% when using only TCC, 0.14% of recall increased, and 0.05% of precision reduced. The performance of proposed method for word segmentation is dramatically dropped when using only TCCT. We conclude that using character clusters as the input sequence for word segmentation achieves the F-measure more than 90%.

From Table 7, our named entity recognition system achieves up to the F-measure of 83.68% when using TCC only. TCC+TCCT decreases 0.38% of the F-measure, 0.41% of recall, and 0.36% of precision when comparing with TCC. The performance of proposed method for named entity recognition is dramatically dropped when using only TCCT. We summarize that using character clusters as the input sequence for named entity recognition achieves the F-measure more than 80%.

We compare our word segmentation method with other works from NSC2010 as shown in Table 8. The WS systems from 12P34C001, 12P34C002 and 12P34C003 used all eight categories[4] as the training set, while our work used only one category from news. 12P34C001 proposed dictionary-based and rule-based methods using longest matching techniques. 12P34C002 proposed N-gram language model with phrase table construction techniques, and using word lists as the post-processing. 12P34C003 proposed stacked generalization using decision tree. From Table 8, our proposed phase-merging model (PM) for WS outperformed dictionary-based and rule-based methods. The performance of our proposed method is slightly lower than the system 12P34C002 and 12P34C003, since our method used a smaller training set and a single learning model without any post-processing step.

[4] Article, encyclopedia, news, novel, Buddhism, law, talk, and Wikipedia

Table 7. Performance of the named entity recognition (NER) task with L-Train setting

Feature	**Precision(%)**	**Recall(%)**	$\mathbf{F}_{\beta=1}$**(%)**
TCC	82.71	84.68	83.68
TCCT	31.75	35.47	33.51
TCC+TCCT	82.35	84.27	83.30

Table 8. Comparison of word segmentation with other works from National Software Contest 2010 (NSC2010)

System	**Precision(%)**	**Recall(%)**	$\mathbf{F}_{\beta=1}$**(%)**
12P34C001	85.10	87.65	86.36
12P34C002	92.70	96.10	94.37
12P34C003	93.55	94.13	93.84
Our PM	92.71	90.33	91.51

4 Conclusion and Future Work

This paper presents discriminative probabilistic model for simultaneous word segmentation and named entity recognition using the character cluster concept. We define type of character cluster and demonstrate the effectiveness of our method. The experimental result shows that our method significantly increases the performance of segmenting word and recognizing entities with the F-measure of 96.14% and 83.68%, respectively. Our proposed work including hierarchical structure representation (HSR) can be applied to any inherent-vowel alphabetic language for natural language processing. For future work, ranking of error types, size of training data and other semantic features such as part-of-speech will be focused.

Acknowledgments. This work has been supported by the Thailand Research Fund (TRF), under Grant No. BRG50800013, a TRF Royal Golden Jubilee Ph.D. Program, under Grant No. PHD/0034/2550, the National Electronics and Computer Technology Center, Thailand, under Grant No. NT-B-22-KE-38-54-01, Government Research Fund via Thammasat University, Thailand, and the National Research University Project of Thailand Office of Higher Education Commission.

References

1. Daniels, P.T., Bright, W.: The World's Writing Systems. Oxford University Press (1996)
2. Huor, C.S., Rithy, T., Hemy, R.P., Navy, V.: Detection and correction of homophonous error word for khmer language (2006), http://www.panl10n.net/english/outputs/Working%20Papers/Cambodia/Microsoft%20Word%20-%204_E_N_248.pdf

3. Koanantakool, T.H., Karoonboonyanan, T., Wutiwiwatchai, C.: Computers and the thai language. IEEE Ann. Hist. Comput. 31(1), 46–61 (2009)
4. Lafferty, J.D., McCallum, A., Pereira, F.C.N.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: ICML 2001: Proceedings of the 18th Int'l Conference on Machine Learning, pp. 282–289. Morgan Kaufmann Publishers Inc., San Francisco (2001)
5. Li, L., Zhou, R., Huang, D.: Two-phase biomedical named entity recognition using crfs. Computational Biology and Chemistry 33(4), 334–338 (2009), http://www.sciencedirect.com/science/article/B73G2-4WRD3DM-1/2/c56643f1540c4e379d80119594d7a799
6. Liu, D.C., Nocedal, J.: On the limited memory bfgs method for large scale optimization. Math. Program. 45(3), 503–528 (1989)
7. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Linguisticae Investigationes 30(1), 3–26 (2007), http://www.ingentaconnect.com/content/jbp/li/2007/00000030/00000001/art00002
8. Okazaki, N.: Crfsuite: A fast implementation of conditional random fields (crfs) (2007), http://www.chokkan.org/software/crfsuite/
9. Theeramunkong, T., Boriboon, M., Haruechaiyasak, C., Kittiphattanabawon, N., Kosawat, K., Onsuwan, C., Siriwat, I., Suwanapong, T., Tongtep, N.: THAI-NEST: A Framework for Thai Named Entity Tagging Specification and Tools. In: CILC 2010: Proceedings of the 2nd Int'l Conference on Corpus Linguistics (CILC 2010), May 13-15, pp. 895–908 (2010)
10. Theeramunkong, T., Sornlertlamvanich, V., Tanhermhong, T., Chinnan, W.: Character cluster based thai information retrieval. In: IRAL 2000: Proceedings of the 5th Int'l Workshop on Information Retrieval with Asian Languages, pp. 75–80. ACM, New York (2000)
11. Thet, T.T., Jin-Cheon, N., Ko, W.K.: Word segmentation for the myanmar language. J. Inf. Sci. 34(5), 688–704 (2008)
12. Tongtep, N., Theeramunkong, T.: A Feature-Based Approach for Relation Extraction from Thai News Documents. In: Chen, H., Yang, C.C., Chau, M., Li, S.-H. (eds.) PAISI 2009. LNCS, vol. 5477, pp. 149–154. Springer, Heidelberg (2009)
13. Tongtep, N., Theeramunkong, T.: Pattern-based Extraction of Named Entities in Thai News Documents. Thammasat International Journal of Science and Technology 15(1), 70–81 (2010)
14. Whitelaw, C., Patrick, J.: Named entity recognition using a character-based probabilistic approach. In: Proceedings of the 7th Conference on Natural Language Learning at HLT-NAACL 2003, pp. 196–199. ACL, Morristown (2003)

A Comparison of Four Association Engines in Divergent Thinking Support Systems on Wikipedia

Kobkrit Viriyayudhakorn, Susumu Kunifuji, and Mizuhito Ogawa

Japan Advanced Institute of Science and Technology, Japan
{kobkrit,kuni,mizuhito}@jaist.ac.jp

Abstract. Associative information, e.g., associated documents, associated keywords, freelinks, and categories are potential sources for divergent thinking support. This paper compares four divergent thinking support engines using associative information extracted from Wikipedia. The first two engines adopt the association search engine GETA , and the last two engines find associations by using document structure. Their quality is compared by experiments with both quantitative and qualitative evaluations, using Eureka! interface, which was inspired by "Hasso-Tobi 2" divergent thinking support groupware.

1 Introduction

Divergent and convergent thinking are defined by J.P. Guilford [1], drawing distinctions between two characteristics of creativity. The brainstorming process proposed by Osborn [2], which is one of the most popular methodologies for enhancing creative thinking ability, consists of two major phases. The divergent thinking phase generates as many ideas as possible with the following rules [3].

1. Produce a large quantity of ideas, without any criticism.
2. Unusual ideas are highly welcome.
3. Adapting or modifying previously suggested ideas is encouraged.

The convergent thinking phase filters, fuses, and derives ideas to make them concrete.

A divergent thinking support system assists users to archive the goals of divergent thinking. There are three levels of divergent thinking support systems, defined by Young [4].

1. The secretariat level, a system only stores and displays the log of user's thoughts, such as a word processor.
2. The framework-paradigm level, a system has secretariat level ability and also provides a user with appropriate paradigms or suggestions to expand user's thoughts.
3. The generative level, a system has framework-paradigm level ability, and automatically constructs and displays new ideas corresponding to previous ideas.

At the framework-paradigm level, a divergent thinking support system also can be classified into one of three categories (from naïve to sophisticated) by the characteristics of generated suggestions [5].

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 226–237, 2011.

1. Free association
2. Forced association
3. Analogy conception

The first category is an almost random association, which is believed to be useless.

There are several examples for the second category. Watanabe [6] proposed the "Idea Editor" that extracts associated keywords, which are moderately unusual to the user. Associative keywords are also used in "Group Idea Processing System" proposed by Kohda, et al. [7]. In both systems, associations among keywords are statistically pre-computed by using the structure of documents.

Kawaji and Kunifuji [5] proposed "Hasso-Tobi 2" divergent thinking support groupware, which extracts freelinks from Wikipedia (in Japanese) after analyzing an input sentence using a morphological analyzer. Wang, et.al. [8] proposed the "Idea Expander", which extracts images by querying an input sentence on Microsoft Bing Image. In both systems, association owes to existing structures on the Internet.

The third category is the most difficult, and only a few examples exist. Young [9] proposed the "metaphor" machine, which can construct metaphors of noun entities by querying those that use the same predicate in a corpus.

This paper compares four methodologies of forced association in divergent thinking by experiments on Wikipedia. The first two adopt *association search* (the second category), which dynamically finds quantitative associative relations among documents by statistical computation. The later two adopt more conventional *informative entity extraction*, which finds a maximal matching between an input sentence and entities of Wikipedia.

We prepare an interface, Eureka!, inspired by "Hasso-Tobi 2" [5], and implement our system by using the association search engine GETA [10]. The quality of four methodologies is compared by experiments on Wikipedia, through both quantitative and qualitative evaluations.

Section 2 explains association search and informative entity extraction. Section 3 briefly reviews the association search engine GETA [10]. Section 4 presents our divergent thinking support system including Eureka! interface. Section 5 shows the four divergent thinking support engines, and the settings of experiments. Section 6 gives experimental results and our observations, and Section 7 concludes the paper.

2 Structured Documents

2.1 Structure of Documents and Information Sources

For divergent thinking support systems, we focus on multi-layered structured documents (e.g., Fig. 1) as a knowledge-base. Each node of a structure is a non-empty sequence of tokens. Typically, a token is a word, and a node is either a category, a title, or a document. A child node may contain hyperlinks that point to a node in a higher layer. For instance, Fig. 1 describes the structure of Wikipedia, such that "Category" is a

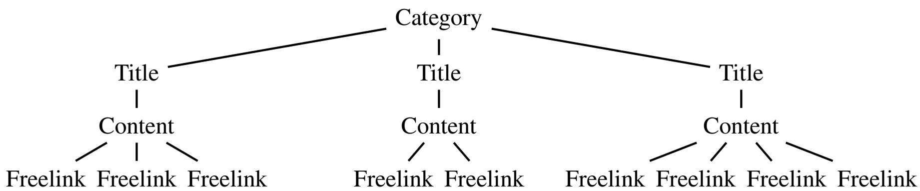
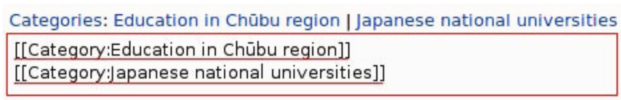


Fig. 1. An example structure of target knowledge-bases

(a) Freelinks in the JAIST Wikipedia page



(b) Categories labeled at the bottom of JAIST Wikipedia page

Fig. 2. The associative information in a Wikipedia article and its source code

parent of “Title”, and “Title” is a parent of “Content”, and a “Freelink”[1] points to a “Title”.

Our instance of a structured document is Wikipedia (English version), which is one of the largest collective intelligence knowledge-bases, and which contains a huge collection of encyclopedia articles. Each article in Wikipedia consists of a title, content, freelinks (see Fig. 2a), and category labels (see Fig. 2b). We use the whole English Wikipedia website[2] and deployed it into our local knowledge-base, which was used in all experiments.

2.2 Association Search

In an association search, an article is regarded as a multiset [11] of tokens (typically, an article is a document and a token is a word). A query is a (multi)set of tokens, and the searched result is a ranking among tokens with respect to a given similarity measure.

Let ID_1 be a set of articles, and let ID_2 be a set of tokens.

[1] http://en.wikipedia.org/wiki/Wikipedia:Free_links
[2] http://download.wikimedia.org/enwiki/20100730/

Definition 1. *An* association system *is a quadruplet* $\mathcal{A} = (ID_1, ID_2, a, \mathrm{SIM})$ *where* a *is an association function and* SIM *is a similarity function such that*

$$\begin{cases} a : & ID_1 \times ID_2 \to \mathbf{N} \\ \mathrm{SIM} : & ID_2 \times MP(ID_1) \to \mathbf{R}_{\geq 0} \end{cases}$$

where $\mathbf{N}$ *is the set of natural numbers,* $\mathbf{R}_{\geq 0}$ *is the set of non-negative real numbers, and* $MP(X)$ *is the multiset consisting of non-empty subsets of* X. *We say that* $\mathcal{A}^t = (ID_2, ID_1, a^t, \mathrm{SIM}^t)$ *is the transpose of* $\mathcal{A}$ *where* $a^t(y, x) = a(x, y)$ *and a given* $\mathrm{SIM}^t : ID_1 \times MP(ID_2) \to \mathbf{R}_{\geq 0}$.

For $X \subseteq ID_1$ *(resp.* ID_2*) and* $n \in \mathbf{N}$*, let* $A(X, n)$ *(resp.* $A^t(X, n)$*) be the function collecting the top* n*-elements in* ID_2 *(resp.* ID_1*) with respect to similarity* $\mathrm{SIM}(y, X)$ *(resp.* $\mathrm{SIM}^t(y, X)$*) for* $y \in ID_2$. *An* association search *is*

$$A^t(\{y \mid (y, v) \in A(X, m)\}, n)$$

for $m \in \mathbf{N}$.

In the definition of an association search, the number m is not specified. From empirical study, GETA (see Section 3) sets m to 200 as default by its developers, balancing efficiency and precision. Note that during an association search, we first compute $A(X, m)$. This result is regarded as a *summary* that characterizes X.

Typical examples of association searches are:

- ID_1 is the set of documents, ID_2 is the set of words, and $a(d, w)$ is the number of occurrences of a word w in a document d. In this case, an association search is documents-to-documents.
- ID_1 is the set of words, ID_2 is the set of documents, and $a^t(w, d) = a(d, w)$. In this case, an association search is words-to-words.

GETA permits a similarity function SIM of the form

$$\mathrm{SIM}(d, q) = \sum_{t \in q} \frac{\mathrm{wq}(t, q) \cdot \mathrm{wd}(t, d)}{\mathrm{norm}(d, q)}$$

with the assumptions that $\mathrm{wd}(t, d) = 0$ if $t \notin d$ and $\mathrm{wq}(t, q) = 0$ if $t \notin q$. Typically,

- the value of $\mathrm{norm}(d, q)$ is dependent only on d. (In such cases, SIM^t is obtained by simply swapping wq and wd.)
- both wd and wq are defined dependent on the association function a.

For an efficient association search implementation (e.g., GETA), we assume $\mathrm{SIM}(y, X) = 0$ if $a(x, y) = 0$ for each $x \in X \subseteq ID_1$ and $y \in ID_2$. Note that an association search does not require structured documents. The key observation is a dual relationship between words and documents. ID_1 and ID_2 are swapped by regarding a document as a multiset of words, and a word as a multiset of documents that contain it with multiplicity. Thus the association search ignores the ordering of words; it does not distinguish, for example, "Weather is not always fine" and "Weather is always not fine".

2.3 Informative Entity Extraction

Informative entity extraction assumes a structured document, in which a token is a word. A query is a phrase (i.e., a sequence of tokens), and the searched result is the set of titles (i.e., sequences of tokens) that contain the phrase as its subsequence. We call such titles *informative entities*.

Definition 2. *Let $\mathcal{W}$ be a set of words, and let $\mathcal{T}(\subseteq \mathcal{W}^*)$ be a set of titles (sequences of words). For an input word sequence $\psi \in \mathcal{W}^*$, we define*

$$Subseq(\psi) = \{\psi' \neq \epsilon \mid \exists \psi_1, \psi_2 \,.\, \psi_1 \psi' \psi_2 = \psi\}$$
$$\mathcal{A}(\psi) = Subseq(\psi) \cap \mathcal{T}$$

$Subseq(\psi)$ is the set of non-empty subsequences of ψ (which are sorted in decreasing order of length) and $\mathcal{A}(\psi)$ is the set of informative entities. For example, from "Reduce electricity usage in Japan Advanced Institute of Science and Technology", "Japan Advanced Institute of Science and Technology" is extracted, which is an item[3] in Wikipedia. Since $Subseq(\psi)$ is sorted, the extraction reports the longest informative entity first, which we can expect a more specific matching.

3 GETA - Generic Engine for Transposable Association Computation

The Generic Engine for Transposable Association Computation[4] (GETA) is an association search engine developed at NII [10]. We use a variant GETAssoc, which intends personal use and easy installation. A key feature of GETA is its scalability; it quickly handles dynamic association search on more than ten million documents, such as on Webcat Plus[5], Imagine[6], Cultural Heritage Online[7], and Pictopic[8].

The key data structure of GETA is a WAM (Word Article Matrix), which represents an association function in Definition 1. WAM is usually a huge sparse matrix of which columns are indexed by words and rows are indexed by names of documents. When ID_1 is a set of words and ID_2 is a set of documents, the intersection of the row of a word w and the column of a document d is $a(w, d)$, which is the number of occurrences of a word w in a document d. Then, the transpose $a^t(w, d)$ is obtained as a transposed WAM. In GETA implementation, a huge and sparse WAM is compressed either vertically or horizontally. These two compressed matrices enable us to compute association functions a and a^t, respectively.

For example, "twitter dollar" is a query of the sample WAM in Fig. 3a. This query is regarded as a two-word document (Fig. 3b). If we adopt the inner product of two

[3] http://en.wikipedia.org/wiki/Japan_Advanced_Institute_of_Science_and_Technology
[4] http://getassoc.cs.nii.ac.jp
[5] http://webcatplus.nii.ac.jp
[6] http://imagine.bookmap.info
[7] http://bunka.nii.ac.jp
[8] http://photobank.pictopic.info/

	"twitter"	"tennis"	"dollar"	"facebook"
IT news	2	0	1	4
Sport news	0	2	1	0
Economic news	0	0	2	0

(a) A sample WAM

	"twitter"	"dollar"
Query	1	1

(b) A sample query vector

	Similarity Score
IT news	3
Sport news	1
Economic news	2

(c) The summary (associated documents) of query "twitter dollar"

	Similarity Score
"facebook"	12
"dollar"	8
"twitter"	6
"tennis"	2

(d) The associated keywords of query "twitter dollar"

Fig. 3. An example of WAM and sample results

column vectors as a similarity function, the result, which is a summary of the query, is shown in Fig. 3c. This summary is queried to the transposed WAM. A similarity function is again computed to yield the associated keywords as Fig. 3d.

The default similarity measures prepared by GETA are Smart measure proposed by Singhal, et al. [12], Okapi BM25, Cosine, and Dot Product. We adopt Smart measure during experiments.

To improve retrieval accuracy, stemming, which reduces inflected words to their root forms (for example, "fishing" and "fisher" to just "fish"), is applied as preprocessing on Wikipedia. We use Snowball[9] English stemmer.

4 Divergent Thinking Support System

4.1 Structure of System

A divergent thinking support system consists of two main components, an association engine and an interface. An engine constructs suggestions for a user. An interface manages the interaction between a user and an engine.

4.2 Workflow of System

First, a user inputs a topic sentence to start the system as in Fig. 4. Next, an association engine reads it, produces a list of associative information by consulting a knowledge-base, and forwards the list as a suggestion list to the interface. Then, the interface displays the list to users. Based on the suggestions, the user inputs the next query sentence. The input sentence is forwarded to the engine to obtain the next suggestion list. This process is repeated until a user is satisfied with the discovered ideas.

[9] http://snowball.tartarus.org/

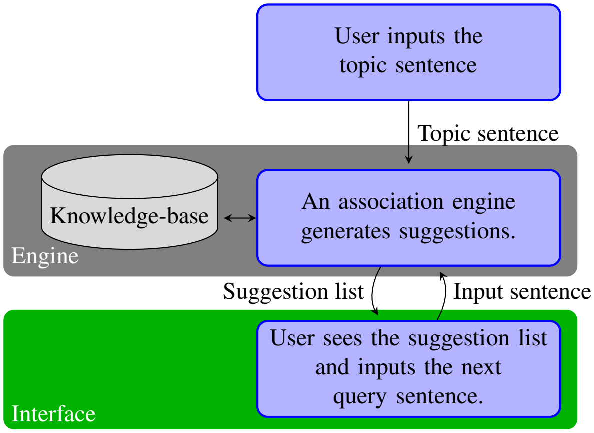


Fig. 4. The workflow of a divergent thinking support system

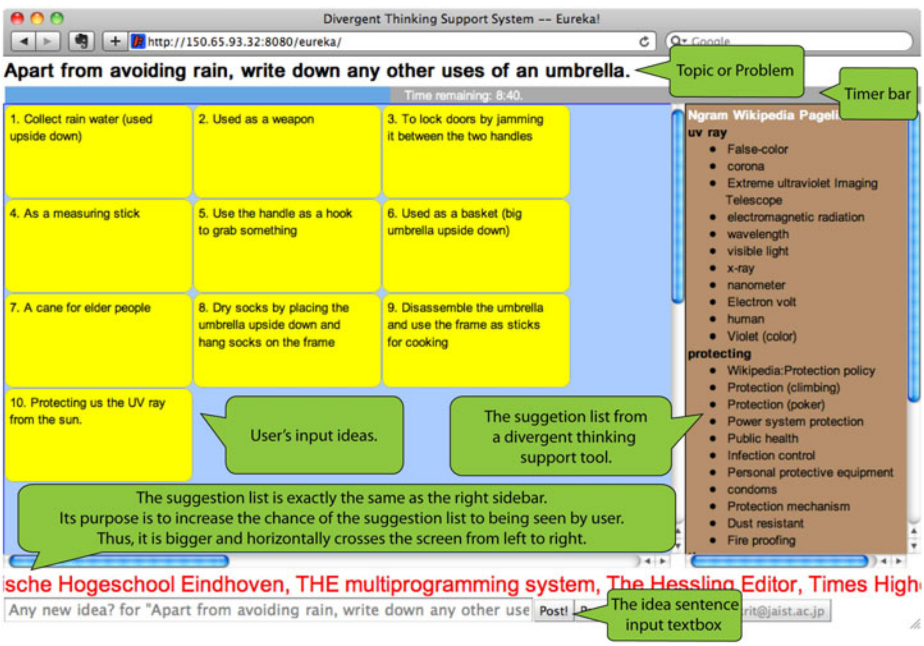


Fig. 5. Eureka! screenshot and functions

4.3 Eureka! Interface

For experiments, we prepare an interface, "Eureka!" (Fig. 5). Eureka! is mainly inspired from the "Hasso-Tobi 2" [5], which is the collaborative divergent thinking support system. Our focus is to evaluate the performance of different association engines, rather than the effect and the influence of social interactions. Thus, we restrict "Eureka!" to be a single-user interface.

The topic is alway shown at the top of the screen in large font to prevent off-topic thinking. At the bottom right sidebars, suggestions are displayed. The right sidebar displays the suggestions in small fonts to allow quick glances by a user. In the bottom sidebar, the suggestions in large font move horizontally in the right-to-left manner (like headline news). Top-ranked suggestions are shown first, intended for easy finding by users. All previously input ideas (query sentences) are displayed as rectangle labels in the central scrollable area. They are displayed without overlaps between labels.

Such a layout intends two effects. First, a user may easily notice previous ideas, which would prevent duplication of ideas. Second, a user can stimulate idea recycling. Whenever a previous idea label is clicked, the suggestion list produced from that idea is loaded and displayed on the bottom right sidebars.

5 Experimental Setting

5.1 Four Divergent Thinking Support Engines

Following are the association engines used during experiments. All are used under Eureka! interface, and only visible differences produced suggestion lists.

- **GD: GETA's most related Documents.** A user input sentence is a query for retrieving associated documents by GETA. The titles of the search results are suggestions, which are sorted by their similarity scores.
- **GK: GETA's most related Keywords.** A user input sentence is a query for retrieving the associated keywords of the query by GETA. The resulting associated keywords are suggestions, which are sorted by their similarity scores.
- **WF: Wikipedia's Freelinks.** A user input sentence is a query for extracting informative entities. The freelinks in their contents are suggestions, which are sorted by the length of entities. That is, a freelink with a longer parent entity comes first. In our current implementation, if their parent entities have the same length, a freelink with a parent entity that appears earlier in an input sentence comes first.
- **WC: Wikipedia's Categories.** A user input sentence is a query for extracting informative entities. All titles in the same category as extracted entities are suggestions, which are sorted by the length of entities. That is, a category with a longer child entity comes first. In our current implementation, if entities have the same length, the ordering among categories and titles in categories obeys to that of Wikipedia.
- **NE: No engines.** No associative information is supplied (no suggestion lists).

5.2 Evaluation Methods

The quality of an engine is measured by the degree of creativity that users generate. As proposed by Guilford[13], the degree of creativity can be measured by fluency, flexibility, and originality of ideas. Neupane, et al. proposed the following three measures [14].

- **Fluency of ideas.** The total number of input ideas, excluding those that are judged to be off-topic, redundant, impossible, and/or useless.

- **Flexibility of ideas.** The total number of viewpoints in input ideas. Viewpoints are defined prior to experiments.
- **Originality of ideas.** The total number of distinct ideas. When ideas are very close or identical, they are grouped together.

For example, consider a problem "Apart from avoiding rain, write down any other uses of an umbrella". Table 1 shows the ideas generated by a participant. Table 2 shows that there are four viewpoints. The third and fifth ideas in Table 1 are grouped. These classifications are performed after experiments by three human evaluators.

5.3 Procedure of Experiment

The experiments are conducted by five groups of users, and each group consists of two users (ten users in total). All participants are graduate students in knowledge science. Before the experiments, every participant is informed:

1. System procedure and usage.
2. Divergent thinking, its rules and examples.
3. Q&A

The five following topics are assigned to all participants. In each topic, ten viewpoints are prepared in advance by three human evaluators, and all resulting ideas are classified into the most related viewpoints, though there are a few difficult cases.

1. Apart from avoiding rain, write down any other uses of an umbrella.
2. If all human beings had a third hand on their back, write down the advantages of that hand.
3. What steps should the concerned authorities take to increase the number of foreign tourists in Japan?

Table 1. A log of ideas during Eureka! interaction

Idea No.	Generated Idea
1	Avoiding the sun
2	Collect rain water (upside down)
3	Used as a ruler
4	To lock doors by jamming the umbrella between the two handles
5	Used as a measuring stick
6	Use the handle to grab something
7	Used as a basket (upside down)
8	A cane for elderly people
9	Dry socks by putting the umbrella upside down and hanging socks on the frame
10	Disassemble the umbrella and use the frame as sticks for cooking

Table 2. A list of viewpoints during Eureka! interaction

Idea Viewpoint	Idea No.
Furniture	9
Tool	1,3,4,5,6,8
Recycle	10
Accessories	
Interior	
Plaything	
Container	2,7
Using Materials	
Clothes	
Social	

Table 3. Group assignment

Topic	A	B	C	D	E
1	GD	GK	WF	WC	NE
2	GK	WF	WC	NE	GD
3	WF	WC	NE	GD	GK
4	WC	NE	GD	GK	WF
5	NE	GD	GK	WF	WC

Table 4. Fluency of ideas

Topic	GD	GK	WF	WC	NE
1	**46(0.31)**	30(0.20)	12(0.08)	22(0.15)	39(0.26)
2	21(0.17)	24(0.19)	**38(0.30)**	7(0.06)	36(0.29)
3	19(0.13)	**43(0.30)**	30(0.21)	41(0.28)	12(0.08)
4	**28(0.26)**	10(0.09)	17(0.16)	25(0.24)	26(0.25)
5	11(0.09)	**43(0.37)**	15(0.13)	25(0.21)	23(0.20)
Total	125(0.96)	**150(1.15)**	112(0.88)	120(0.94)	136(1.08)

Table 5. Flexibility of ideas

Topic	GD	GK	WF	WC	NE
1	**8(0.25)**	7(0.22)	2(0.06)	7(0.22)	**8(0.25)**
2	**8(0.27)**	6(0.20)	**8(0.27)**	2(0.07)	6(0.20)
3	7(0.21)	7(0.21)	8(0.24)	**9(0.26)**	3(0.09)
4	**7(0.23)**	4(0.13)	5(0.17)	**7(0.23)**	**7(0.23)**
5	4(0.13)	**9(0.29)**	6(0.19)	6(0.19)	6(0.19)
Total	**34(1.09)**	33(1.05)	29(0.93)	31(0.97)	30(0.96)

4. How could you contribute to power saving at your school?
5. Imagine that you are a product designer, please design new products likely to be sold to teenagers.

Experiments use each four engines and no engine (for comparison), as described in Section 5.1, and the maximum number of suggestions is limited to 30 for each engine.

To avoid the effect of tool experiences, the topics and the engines are assigned to groups in different orders, as shown in Table 3. The timeout for each topic is set at 15 minutes.

Three human evaluators, who did not participate in the experiments, judged the fluency, the flexibility, and the originality of ideas. Majority vote is taken if conflict occurs.

Table 6. Originality of ideas

Topic	GD	GK	WF	WC	NE
1	**39(0.31)**	26(0.21)	8(0.06)	21(0.17)	31(0.25)
2	19(0.17)	20(0.18)	**35(0.31)**	6(0.05)	34(0.30)
3	17(0.15)	24(0.22)	27(0.25)	**35(0.32)**	7(0.06)
4	**28(0.26)**	10(0.09)	17(0.16)	25(0.24)	26(0.25)
5	8(0.08)	**37(0.36)**	15(0.15)	24(0.24)	18(0.18)
Total	111(0.97)	**117(1.06)**	102(0.93)	111(1.02)	116(1.04)

Table 7. Average of total standard values

Measures	Average
GD	1.01
GK	**1.09**
WF	0.91
WC	0.98
NE	1.02

6 Experimental Results and Observation

6.1 Quantitative Results

The quality of divergent thinking engines is estimated by four measures in Section 5.2. Their scores, which are the sum scores of two sessions by the same group, are shown in Table 4, Table 5, and Table 6. The number in parenthesis is standard value, which is the ratio of a score to the total score of all engines on the same topic.

In the results, fluency of ideas is the highest when supported by GK. Original ideas are also most often discovered when supported by GK. GD yields the highest flexibility.

Table 7 shows the averages of total standard values yielded by each engine. GK yields the highest average score. Thus, we conclude that GK is the best among the four divergent thinking support engines described in Section 5.1.

Our observation on the advantages of GK is, first, that GETA can accurately summarize keywords. Second, each item of the summary is just a single word, which requires little time to understand and is loosely interpreted in various ways. Therefore, users most fluently generate ideas when GK is used.

The second best is NE. After observation, participants do not waste time waiting for a engine's response, or reading the resulting suggestions. Each engine requires approximately three seconds to generate a suggestion list. The NE is better than most engines because participants have more time than others.

6.2 Qualitative Results

The qualitative evaluation is the survey answered by participants. The question is multiple choice asked immediately after finishing each topic, i.e., "Please evaluate the usefulness of this divergent thinking support engine on a scale from 1(Poor) to 5(Excellent)".

GD	GK	WF	WC	NE
3.0	**3.9**	2.6	2.7	1.0

According to the survey, GK obtains the most satisfactory evaluation from users, which is consistent with the quantitative results. After analysis, we observed that most users avoid inputting lengthy informative entities. Instead, they input sample words or abbreviations. For example, instead of inputting "Japan Advanced Institute of Science and Technology", they input "University" or just "JAIST". Thus, informative entity extraction (see Section 2.3), which reports the longest informative entity first, fails in most cases. It is a major difficulty that reduces the quality of both WF and WC.

7 Conclusion

This paper compared four divergent thinking engines based on forced association. GD and GK adopt the association search engine GETA, while WF and WC use the more conventional informative entity extraction. The quality of the four engines is evaluated by experiments on Wikipedia using the Eureka! interface. GK (the most related keywords by GETA, using associated keywords of an input sentence) is the most effective and gives the highest satisfaction to users.

We think that dynamic association creation on a large scale database (e.g., GETA) is useful for divergent thinking support, whereas previous systems used either statistically precomputed (relatively few) associations, or manually-prepared (huge number of) freelinks on the Internet. We believe that empirical statistical computation would find unexpected associations beyond biased human thinking.

Acknowledgment. This research is partially supported by Grant-in-Aid for Scientific Research (20680036) from the Ministry of Education, Culture, Sport, Science and Technology of Japan and by the Japan Advanced Institute of Science and Technology (JAIST) under the JAIST overseas training program. We thank Prof. Akihiko Takano and Prof. Shingo Nishioka at the National Institute of Informatics (NII) for fruitful suggestions.

References

1. Guilford, J.P.: The Nature of Human Intelligence. McGraw-Hill, New York (1967)
2. Osborn, A.: Your Creative Power. Charles Scribner's Son, New York (1948)
3. Kunifuji, S., Kato, N., Wierzbicki, A.: Creativity support in brainstorming. In: Wierzbicki, A., Nakamori, Y. (eds.) Creative Environments. SCI, vol. 59, pp. 93–126. Springer, Heidelberg (2007)
4. Young, L.F.: Idea Processing Support: Definitions and Concepts. In: Decision Support and Idea Processing Systems, ch. 8, pp. 243–268 (1988)
5. Kawaji, T., Kunifuji, S.: Divergent thinking supporting groupware by using collective intelligence. In: Proceedings of The Third International Conference on Knowledge, Information and Creativity Support Systems (KICSS 2008), pp. 192–199 (2008)
6. Watanabe, I.: Idea Editor: Overview. Technical Report Research Report IIAS-RR-90-2E, Fujitsu Laboratories Limited (October 1990)
7. Kohda, Y., Watanabe, I., Misue, K., Hiraiwa, S., Masui, M.: Group Idea Processing System: GrIPS. Transactions of the Japanese Society for Artificial Intelligence 8, 601 (1993)
8. Wang, H.-C., Cosley, D., Fussell, S.R.: Idea Expander: Supporting group brainstorming with conversationally triggered visual thinking stimuli. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work, pp. 103–106. ACM (2010)
9. Young, L.F.: The metaphor machine: A database method for creativity support. Decision Support Systems 3(4), 309–317 (1987)
10. Takano, A.: Association Computation for Information Access. In: Grieser, G., Tanaka, Y., Yamamoto, A. (eds.) DS 2003. LNCS (LNAI), vol. 2843, pp. 33–44. Springer, Heidelberg (2003)
11. Dershowitz, N., Manna, Z.: Proving termination with multiset orderings. Communication of the ACM 22(8), 465–476 (1979)
12. Singhal, A., Buckley, C., Mitra, M., Mitra, A.: Pivoted document length normalization, pp. 21–29. ACM Press (1996)
13. Guilford, J.P.: Creativity and its Cultivation. In: Traits of Creativity, pp. 142–161, Harper (1959)
14. Neupane, U., Miura, M., Hayama, T., Kunifuji, S.: Distributed environment for brain writing support groupware and its evaluation, tree and sequences. Journal of Japanese Creativity Society 10, 74–86 (2006) (in Japanese)

An Ontology-Based Method for Measurement of Transferability and Complexity of Knowledge in Multi-site Software Development Environment

Pornpit Wongthongtham and Natsuda Kasisopha

DEBII, Curtin University, GPO Box U1987, Perth, WA, 6845, Australia
P.Wongthongtham@cbs.curtin.edu.au,
Natsuda.Kasisopha@postgrad.curtin.edu.au

Abstract. Challenges over multi-site software development are on working in virtual teams and sharing knowledge. It is quite normal that software engineers working in a virtual team have never met face to face in multi-site environments. In addition they have different educational backgrounds and interpret methods in different ways. Software engineering education, training, and practice are different between universities, cities, and countries. As a result, it is difficult to share a piece of knowledge between distributed teams and among remote team members. There are a number of standards that different teams could be referring to. Remote software engineers use a particular standard as their own individual guide and when they share their own knowledge base and terminology is different from those of others. Most issues raised are related to inconsistency in understanding software engineering theories and practice. Therefore sharing knowledge is the challenge and to resolve the differences between the distributed teams we need to understand its key variables of knowledge sharing. In this paper we propose an ontology-based approach for knowledge sharing measurement. Particularly in the approach, we look into measurement of transferability and complexity of knowledge. The impact of nature of knowledge on knowledge sharing is focused. A prototype is developed taking Software Engineering Ontology as example.

Keywords: Ontology, Knowledge Sharing Measurement, Transferability of Knowledge, Complexity of Knowledge, Multi-site Software Development, Software Engineering.

1 Introduction

Moving the software development to countries where employees are on comparatively lower wages is becoming the norm in our modern and economy world. It is the imperative of financial gain that drives businesses to multi-site development. There are more advantages than disadvantages of multi-site software development for example economic and labour benefits, overcome skill shortages, productivity through 24 hour work, just to name a few. Global corporations have moved to multi-site software development. With high technologies in information and communication it

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 238–252, 2011.

eliminates the perception of distance. However, challenges are rather on working in virtual teams and sharing knowledge. It is quite normal that software engineers working in a virtual team have never met face to face in multi-site environments. In addition as they are scatter around the world, they have different educational backgrounds and interpret methods in different ways. Software engineering education, training, and practice are different between universities, cities, and countries. As a result, it is difficult to share a piece of knowledge between distributed teams and among remote team members. For example, the inconsistency in presentation, documentation, design, and diagrams could be barriers to other teams or members in understanding a message or a piece of work. Often issues on diagrams which do not follow the same notation standard are disregarded causing misunderstanding and can even cause project significantly delay. There are a number of standards that different teams could be referring to for example AS4651 Australian standard metamodel for software development methodologies [1], ECSS-E-40 software engineering standard [2], etc. Remote software engineers use a particular standard as their own individual guide and when they share their own knowledge base and terminology is different from those of others. Most issues raised are related to inconsistency in understanding software engineering theories and practice. Therefore sharing knowledge is the challenge and to resolve the differences between the distributed teams we need to understand its key variables of knowledge sharing.

In this paper we propose an ontology-based approach for knowledge sharing measurement particularly in multi-site software development environment. We first look into collaboration issues in multi-site software development in the next section. We then form the key variables. The impact of nature of knowledge on knowledge sharing is focused. Next we explain the approach which is based on ontology (complexity of ontology and transferability of ontology). We have developed a prototype which is explained in the experimental studies. Before we conclude our work we review existing approach in related work section.

2 Collaboration in Multi-site Software Development Environment

Activities in multi-site software development are performed by team members who work remotely in a distributed environment and who interact with parties that may be people or systems. These activities can be viewed into four categories based on issues arose in multi-site software development i.e. teamwork, project management, technical knowledge, and project complexity. These four issues form into four key variables of knowledge sharing measurement. Table 1 shows four categories of collaboration issues in multi-site software development environment.

Table 1. Collaboration issues in multi-site software development environment

Collaboration Issues	
Social Captital	Knowledge Capital
Teamwork	Technical Knowledge
Project Managment	Project Complexity

The first two categories i.e. teamwork and project management are considered under the general notion of social capital while the other two categories i.e. technical knowledge and project complexity are included in the notion of knowledge capital.

2.1 Social Capital

Teamwork and project management significantly depend on the relationships between team members (that is on social capital). Team members include customers, users, software/system developers, and stakeholders. Ineffective sharing of knowledge among team members is the most critical for project failure.

Teamwork. Teamwork issues refer to issues related to team member development, communication between members, and team management. Factors caused project failure include ineffective communication and coordination among project teams, inexperienced project manager, lack of specialized skills, low confidence in team members, insufficient support from manager, inadequately train of team members. DeMarco and Lister argued that aspect of the skills and interactions of software team is most critical and hard to overcome [3].

Project Management. Project management issues refer to issues related to project plan and schedule, budget, assessment, control, quality assurance. This includes uncertainty of project milestones, change management, progress report, and project management methodology.

2.2 Knowledge Capital

Technical knowledge and project complexity contributes to the knowledge capital which is all about knowledge work. Knowledge capital is accumulated by various knowledge creating processes.

Technical Knowledge. Technical knowledge issues refer to issues related to software process activities including requirement engineering, design, implementation, testing, validation and verification, development methodology. It could cause by ambiguous system requirement, incorrect system requirement, wrong development strategies, inappropriate software design, inadequate testing, lack of reusable support of data, code, component, document, etc. However, McCreery and Moranta believe that project challenges were more behavioural and interpersonal than technical [4]. Issues related to communication, collaboration, and team connectedness are more critical.

Project Complexity. Project complexity issues refer to issues related to the complexity of project requirements. This includes the projects utilizing cutting edge technology and that require high level of technical complexity.

3 Key Variables of Knowledge Sharing

Wang and Yang define knowledge sharing as the action in which people dispread relevant information to others across the organization [3]. Melnik and Maurer divide knowledge sharing into two perspectives i.e. codification approach and personalization approach [5]. Codification approach is based on the notion of knowledge as object [6-9]which can be created, collected, stored, and reused [5]. Personalization approach is based on the notion of knowledge as relationship [10-13] which is uncertain, soft, and embedded in work practices and social relationships.

Knowledge sharing in software development can be defined as activities between team members in spreading project data/information/agreement. Knowledge sharing includes communication, updates, advice, problem solving, decision making, issue raising, discussion, etc. over project data/information/agreement. Knowledge sharing in software development situation enables team members to enhance their competency and mutually generate new knowledge [5].

From the above four collaboration issues given in previous section, it forms into four key variables of knowledge sharing measurement which is shown in Table 2.

Table 2. Four key variables for knowledge sharing measurement

Knowledge Sharing	
Trust	Knowledge
Benevolence	Transferability of Knowledge
Competency	Complexity of new Knowledge

3.1 Trust

The concept of trust is related to different and various fields including philosophy, sociology, business, computing. There are number of trust definitions. Mayer et al. define trust as “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trust or, irrespective of the ability to monitor or control that other party” [14]. Moe and Smite define trust as “the shared perception by the majority of team members that individuals in the team will perform particular actions important to its members and that the individuals will recognize and protect the rights and interests of all the team members engaged in their joint endeavour” [15]. Jarvenpaa et al believe that trust has direct positive effect on cooperation and performance and an increase in trust is likely to have a direct, positive impact on team members’ attitudes and perceived outcomes [16]. Giddens [17] sees trust in different view and says that there would be no need of trust if the activities were clearly visible and easy to understand. Hence from his view the prime condition for lack of trust is lack of full information or ambiguous information. As a result trust requires a good knowledge sharing.

Trust can be founded in different ways. The most common way is a direct relationship. In vertical view trust is important to leadership while in horizontal view trust is important for knowledge sharing and team working. In relation to teamwork the two most important dimensions of trust that should be focused are benevolence and competency.

Benevolence Trust and Competency Trust. Benevolence trust is related to willingness within teamwork based on the idea that members will not intentionally harm another when given the opportunity to do so. This kind of trust can be positive or negative which members in the team may believe on others willingness to share knowledge and trust level can be in highest level. On the other hand, they may refuse to others willingness and trust can be negative. Competency trust refers to trusting agent's believe on trusted agent's competency. It describes a relationship in which a member believes that another member is knowledgeable about a given subject area. Competence-based trust also can be negative or positive and members can believe on others ability or they completely refuse others ability in a given subject area.

3.2 Knowledge

Knowledge is a combination of data and information being produced by human thought processes. Knowledge management is the process by which organizations generate value from their intellectual and knowledge-based assets [18]. Drucker defines knowledge as an input resource that will have a greater impact than will physical capital in the future [19]. Knowledge can be categorized in two different classes i.e. explicit and tacit knowledge. Explicit knowledge can relatively easily be formulated by means of symbols and can be transferred to others easily [20]. Tacit knowledge is defined as non-codified, disembodied know-how that is acquired via the informal take-up of learned behaviour and procedures [21].

It is necessary to understand the nature of knowledge in order to analyse the process of knowledge sharing between and within organizations or individuals. The characteristics of knowledge influence the outcome of knowledge sharing [20]. The impact of the nature of knowledge on knowledge sharing is our focus in this paper. The nature of the knowledge also affects the importance of trust in knowledge sharing.

Complexity and Transferability of Knowledge. When the knowledge seems simple, competence-based trust is not necessarily important and in this case, people care more about benevolence-based trust. On the other hand, when the knowledge is complex and professional, people care more about competency-based trust. We divide knowledge type into easy or complex knowledge (complexity of knowledge) and easy or hard transferable knowledge (transferability of knowledge). The complex knowledge and/or long knowledge transfer chain suffer from information distortion and loss which could lead to inefficient knowledge sharing.

4 Ontology-Based Approach for Measurement of Transferability and Complexity of Knowledge

How well people share their knowledge is the subject of this paper. There are two key variables involved for knowledge sharing measurement i.e. transferability and complexity of knowledge. For people who have different information domains we firstly measure their background similarity and find the different knowledge. Then we measure complexity of the different part of knowledge. If both have very similar background of knowledge, both will share the knowledge well. If both have similar background of knowledge and new knowledge is not complicated, both can share the knowledge at some value. In worst case if both come from totally different background of knowledge and new knowledge is very complicated then both will not be able to share knowledge. The impact of nature of knowledge on knowledge sharing is focused in this paper. We propose an ontology-based approach for knowledge sharing measurement. Particularly in the approach, we look into measurement of transferability and complexity of knowledge.

4.1 Ontology

Ontology is an explicit specification of a conceptualisation [22] enabling underlying knowledge representation. Ontologies are used in widespread application areas e.g. semantic web, medical informatics, e-commerce, etc. Mainly ontologies are used to provide a shared semantically domain knowledge in a declarative formalism. Ontology specifies consensual knowledge accepted by a community. However meaning and understanding of concepts in ontologies vary in different sites. Determining the semantic similarity or difference of two ontologies at two sites is vital to knowledge sharing. In this paper we represent multi-site software development knowledge with ontology which can be distinguished into general knowledge and specific knowledge. General knowledge is explicit and is easily understood by locals and neighbours since both their ontologies are similar. Specific knowledge is more technical and difficult to understand and depends on a team member's background and knowledge level and their ontologies are different. Typically it is assumed in this paper that the ontologies are described within the same knowledge representation language i.e. OWL or OWL2 [23].

4.2 Transferability of Ontology

To measure the transferability of two knowledge backgrounds, ontology similarity is considered and calculated. In the means of obtaining the senses and hyponyms of the each concept in the ontologies and based on the structure of the ontologies, the similarity of two ontologies can be calculated. Precisely said ontology transferability is signified by ontology similarity. Nevertheless, there may be more than one sense for each concept. The senses of subclasses of ontology can be determined by their ancestors. To which sense from the root of the ontology it is determined by the users.

A numeric measurement of ontology transferability is obtained using ontology similarity formulas. Wang and Ali defined the difference of set of concepts captured in an ontology from set of concepts captured in the ontology [24]. If the two ontologies are totally different, the difference value is given 1 or the similarity value is given 0. On the contrary, if the two ontologies are the same, the difference value is given 0 or the similarity value is given 1. In this paper, the similarity of a set of concepts, S1, captured in ontology 1, O1, from set of concepts, S2, captured in ontology 2, O2, is defined as

$$\text{Similarity of S1 and S2} = \{x | x \in S1 \wedge x \in S2\} \quad (1)$$

The semantic similarity between O1 and O2 or the transferability can be defined by function Trans(S1, S2) in formula (2) as following

$$\text{Trans(S1, S2)} = 1 - \frac{|S1 - S2|}{|S1|} \quad (2)$$

We compare in both directions thus Trans(S1, S2) and Trans(S2, S1) may be given different value. Determining the semantic similarity or difference of two ontologies is vital to ontology transferability. Transfer of knowledge will be efficient and effective if knowledge senders and knowledge receivers having similar understanding of concepts in ontologies.

4.3 Complexity of Ontology

Ontology complexity is related to the complexity of conceptualization of the domain of interest i.e. software engineering domain. It is measured to reflect how easy any ontology is to understand. Definition of ontology complexity is clarified in features that characterize complexity of ontology i.e. (i) usability and usefulness and (ii) maintainability. For example, a more complicated ontology indicates a more specified knowledge. However, it is difficult to comprehend and requires a high value of competence-based trust. Usability and usefulness of the knowledge may be then decreasing which implies a major impact on knowledge sharing. Additionally complicated ontology is hard to maintain.

In order to measure the complexity of ontology, number of ontology classes, datatype properties, object properties, constraints, and hierarchical paths is considered. Number of Ontology Classes (NoOC) is needed to obtain average value. In OWL the ontology classes are represented as *owl:Class*. Number of Datatype Properties (NoDP) illustrates how well concepts are being defined. In OWL the datatype properties are indicated as *owl:dataTypeProperty*. Number of Object Properties (NoOP) illustrates how well spread of concepts within the ontology. In OWL the object properties are indicated as *owl:objectProperty*. Number of Constraints (NoC) illustrates how well relations being restricted. In OWL the constraints are indicated as *owl:allValuesFrom, owl:someValueFrom, owl:hasValue, owl:cardinality, owl:minCardinality,* and *owl:maxCardinality*. Lastly Number of Hierarchical Paths (NoHP) illustrates how fine concepts being presented. In OWL the hierarchical paths are represented as *owl:subClassOf*.

To calculate complexity of an ontology O, a numeric measurement is defined by function Complex(O) using above parameters in formula (3) as following

$$\text{Complex(O)} = \left(\frac{\sum(NoDP+NoOP+NoC+NoHP)}{Max(NoDP)+Max(NoOP)+Max(NoC)+Max(NoHP)}\right)/NoOC \quad (3)$$

Where Max(NoDP) is maximum number of datatype property, Max(NoOP) is maximum number of object property, Max(NoC) is maximum number of constraint, and Max(NoHP) is maximum number of hierarchical path.

The complexity value is ranged between 0 and 1 which 0 means the ontology is not very complicated while 1 means the ontology is very complicated.

5 Experimental Studies

The experimental studies take Software Engineering Ontology [25] as example. The prototype is implemented using JAVA. We use OWL2.0 API to load and manipulate ontologies which are related to the domains of people who are going to share the knowledge. Java WordNet Library (JWNL) is the main API which is used to obtain the semantic meanings of each concept captured in ontologies.

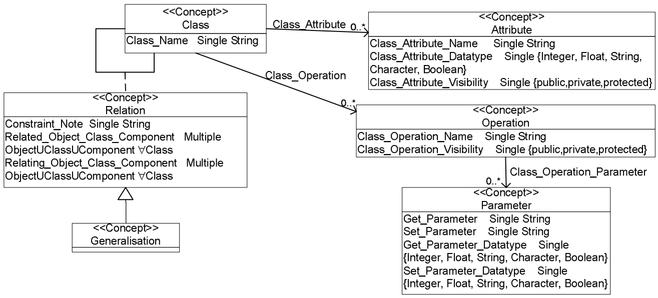


Fig. 1. Classification Models Ontology Relations

Assume remote software engineers having different information domains e.g. a software engineer at site A in Australia having ‘classification model ontology’ and another software engineer at site B in India having ‘class diagram ontology’. Figure 1 shows classification model ontology relations and Figure 2 shows class diagram ontology relations.

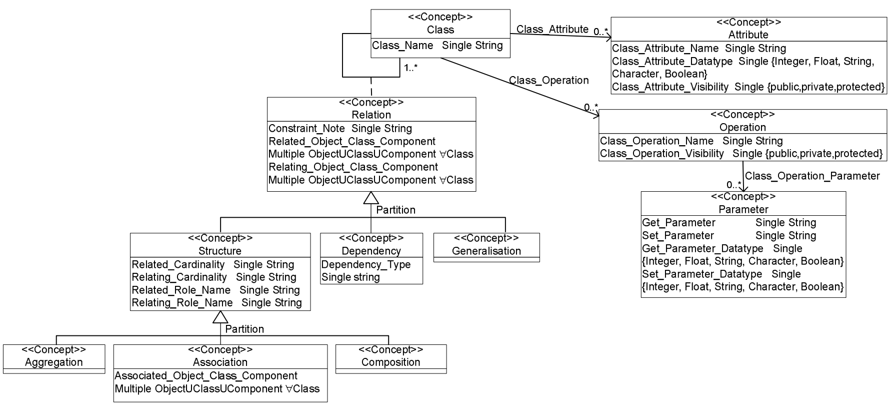


Fig. 2. Class Diagram Ontology Relations

If both software engineers need to share knowledge how well both software engineers share the knowledge is the subject of this paper. There are two key variables i.e. transferability and complexity of knowledge for knowledge sharing measurement. In the experimental studies, we measure the transferability of knowledge indicating by ontology similarity. By comparing both ontologies, the different part will be calculated its complexity.

In order to assess transferability from class diagram ontology to classification model ontology, we need to get sense set of the two ontologies. In other words, we get the concepts and their senses with hypernyms for both class diagram ontology and classification model ontology. In process of getting sense set, users initially choose which sense s/he means at the root concept if there is more than one sense. Senses and its hypernyms are obtained from WordNet. Among those retrieve from WordNet, we also include is-a relationship to differentiate concept from others if there is more than one sense for that particular concept.

Table 3 below shows some of senses and hypernyms from WordNet for class diagram ontology. The highlighted senses are ones being included in the sense set or precisely said they are ones that have meaning within the meant content.

Table 3. Some of senses and hypernyms retrieved from WordNet for class diagram ontology

Concept	Senses	Hypernyms
relation	relation	abstraction
	sexual intercourse, intercourse, sex act, copulation, coitus, coition, sexual congress, congress, sexual relation, relation, carnal knowledge	sexual activity, sexual practice, sex, sex activity
	relative, relation	person, individual, someone, somebody, mortal, soul
	relation, telling, recounting	narration, recital, yarn
	relation back, relation	legal principle, judicial principle, judicial doctrine
	relation	dealings, traffic
Structure	structure, construction	artifact, artefact
	structure	constitution, composition, makeup
	structure	cognition, knowledge, noesis
	structure, anatomical structure, complex body part, bodily structure, body structure	body part
	social organization, social organisation, social structure, social system, structure	system, scheme
dependency	dependence, dependance, dependency	state
	dependence, dependance, dependency	physiological state, physiological condition
	colony, dependency	geographical area, geographic area, geographical region, geographic region

Table 4. (*continued*)

generalisation	abstraction, generalization, generalisation	theorization, theorisation
	generalization, generalisation, induction, inductive reasoning	colligation
	generalization, generalisation, generality	idea, thought
	generalization, generalisation, stimulus generalization, stimulus generalisation	transfer, transfer of training, carry-over
aggregation	collection, aggregation, accumulation, assemblage	group, grouping
	collection, collecting, assembling, aggregation	grouping
association	association	organization, organisation
	association	social activity
	association	union, unification
	affiliation, association, tie, tie-up	relationship
	association, connection, connexion	memory, remembering
	association	relation
	association	chemical process, chemical change, chemical action
	association	group, grouping
composition	composition	mixture
	constitution, composition, makeup	property
	composition, composing	relation
	musical composition, opus, composition, piece, piece of music	music
	composing, composition	creating by mental acts
	writing, authorship, composition, penning	verbal creation
	typography, composition	printing, printing process
	composition, paper, report, theme	essay
	composition	creation

Since concept Relation is the root concept and there are more senses for this particular concept, thus user needs to initially select the sense s/he means. By looking at hypernyms the meant sense is selected (the first one i.e. relation with hypernym of abstraction). There are 8 senses for concept Association shown in Table 3. The sixth sense is selected to be included in the sense set because its hypernyms (relation) matches with the selected root sense. However, we need to differentiate 'association' from other 'association' among other 8 senses by incorporating 'is-a' relationship. In order to identify the 'is-a' relationship, we add '_is-a_' following with parent concept. Thus it becomes 'association_is-a_relation'. Similarly for concept Composition, the third senses i.e. composition and composing are selected to be included in the sense set due to the matched hypernyms with the selected root sense (relation). For this, the sense set includes 'composition_is-a_relation' and 'composing'. For concepts Structure, Dependency, Generalisation, and Aggregation, there is no match sense with parent root sense i.e. relation, we simply include them into the sense set. Therefore, from Table 1, the sense set for class diagram ontology is {relation, structure, dependency, generalization, aggregation, association_is-a_relation, composition_is-a_relation, composing} and the sense set for classification model ontology is {relation, generalisation}. After comparing, there are six distinct senses that appear in the class diagram ontology but do not appear in the classification model ontology. The six senses are "structure, dependency, aggregation, association_is-a_relation, composition_is-a_relation, and composing" The transferability between a set of concepts, S1, captured in class diagram ontology from a set of concepts, S2, captured in classification model ontology is as follow:

$$\text{Trans}(S1, S2) = 1 - (6/8) = 0.25.$$

In conversion, the transferability between a set of concepts, S2, captured in classification model ontology from a set of concepts, S1, captured in class diagram ontology is as follow:

$$\text{Trans}(S2, S1) = 1 - (0/2) = 1$$

Next we calculate complexity of different part in the class diagram ontology. In the different part there are 5 classes i.e. Structure, Dependency, Aggregation, Association, and Composition. Class Structure has 5 datatype properties and 2 object properties, 2 constraints, and no hierarchical path. Class Dependency has 2 datatype properties, 2 object properties, 2 constraints, and no hierarchical path. Class Aggregation has 5 datatype properties, 2 object properties which are all inherited properties, 2 constraints, and 1 hierarchical path. Class Association has 5 datatype properties and 3 object properties, 3 constraints, and 1 hierarchical path. Class Composition has 5 datatype properties, 2 object properties which are all inherited properties, 2 constraints, and 1 hierarchical path. Therefore complexity value of the different path in the class diagram ontology is as follow:

$$\text{Complex}(O) = \left(\frac{9+6+10+12+10}{5+3+3+1}\right)/5 = 0.7833$$

Figure 3 shows transferability value in both directions and complexity value of the new knowledge.

The result shows that the software engineer who has classification model ontology will understand the software engineer who has class diagram ontology to some degree. In contrast, the software engineer who has class diagram ontology will considerably understand the software engineer who has classification model ontology. How well both software engineers to share their knowledge also depends on the complexity of their knowledge background difference.

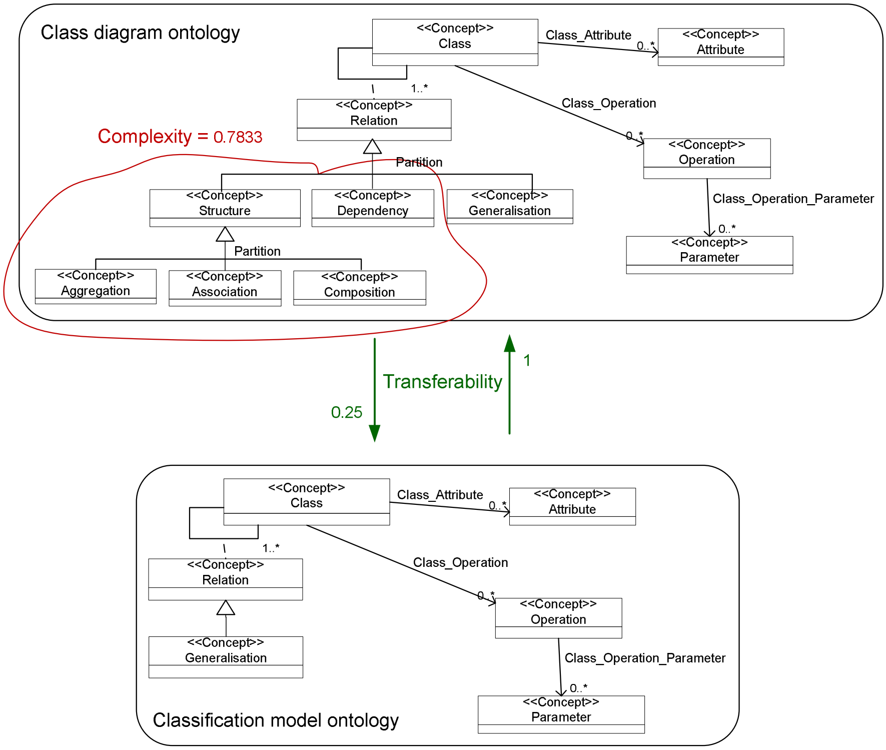


Fig. 3. Complexity and Transferability as Two Keys Variables for Knowledge Sharing

6 Related Work

There are many studies in semantic web applications emphasizing on measuring ontology similarity and difference. A number of approaches have been proposed to deal with the heterogeneity of ontologies [24]. Ontology integration approach maps different ontologies into a generic ontology using vocabulary heterogeneity resolution on various ontologies [26-29]. In this method, the semantic transferability has not been measured before merging into the generic ontology. Measuring the semantic transferability is important in the integrated ontology whether the ontologies should be merged. Suggested Upper Merged Ontology (SUMO) is developed to merge ontologies by sharing ideas from all available ontologies and mapping the entries of merged ontologies with WordNet entries [30]. However this approach does not address the requirement of transferability of two ontologies. One approach creates a computational model to assess semantic similarity among entity classes from different and independent ontologies without having to form a shared ontology [31]. This approach is not practical to measure semantic transferability of two ontologies due to the complexity of matching process. Another approach proposes ontology-based information retrieval model by using domain ontology to extend the original keywords input by users and calculates the concept similarity. Yet this approach does not address the requirement of transferability of two ontologies.

In regard to existing work on ontology complexity, there are existing metrics for analyzing ontology quality but only few of them focus on complexity of ontology. Burton-Jones et al. [32] measure elements of quality i.e. syntactic quality, semantic quality, pragmatic quality, and social quality using a number of attributes. Dazhou et al. [33] present complexity measurement for ontology based on UML. However UML cannot entirely represent semantic richness like ontology does. UML is not a suitable modeling language to represent an ontology, thus, the method cannot measure the structure complexity of ontology objectively. Chris Mungall [34] researched the increased complexity of Gene Ontology which is similar to Dalu et al. method [35]. Anthony et al. [36] also proposed a metric suite to measure the increased complexity of tourism ontologies throughout ontology evolution. However, the metrics in [5], [6], and [7] are evaluating ontology in ontology evolution. Idris [37] proposed conceptual coherence and conceptual complexity metrics based on graph theory. Orme et al. [38] examined coupling between ontologies. Nevertheless, in [5], [6], [7], [8], and [9], complexity is analyzed by the concept structure and does not consider the number of restrictions.

To summary, there is no approach in the literature attempt to measure ontology transferability and existing works on ontology complexity measurement are not enough to cover completed measurement. In this paper we address the ontology transferability and complexity as two key variables for knowledge sharing in multi-site software development.

7 Conclusion

We have addressed knowledge complexity and knowledge transferability as key variables for knowledge sharing in multi-site software development. We then proposed the ontology based approach which measures ontology complexity and transferability to correspond to knowledge complexity and knowledge transferability respectively. The Software Engineering Ontology was used in case studies and a prototype has been developed for proof of concept. WordNet library is used to obtain semantic meaning of concept. For future work we would like to apply our approach to other domain for example e-commerce, health domains.

References

1. Australian Standard metamodel for software development methodologies (2004), `http://www.saiglobal.com/PDFTemp/Previews/OSH/as/as4000/4600/4651-2004.pdf` (cited June 25, 2010)
2. Jones, M., et al.: Introducing ECSS Software-Engineering Standards within ESA. EESA Bulletin 111, 132–139 (2002)
3. Juan-Ru, W., Jin, Y.: Study on Knowledge Sharing Behavior in Software Development Team. In: Fourth International Conference on Wireless Communications, Networking and Mobile Computing (WiCom 2008), Dalian, China (2008)

4. McCreery, J., Moranta, V.: Mapping Team Collaboration in Software Development Projects. In: Portland International Conference on Management of Engineering and Technology (PICMET 2009), Portland, Oregon USA (2009)
5. Melnik, G., Maurer, F.: Direct Verbal Communication as a Catalyst of Agile Knowledge Sharing. In: Proceedings of the Agile Development Conference (ADC 2004), Salt Lake City, UT, USA (2004)
6. Alavi, M., Leidner, D.: Knowledge Management Systems: Issues, Challenges, and Benefits. Communications of the AIS 1(7), 2–36 (1999)
7. Hansen, M., Haas, M.: Competing for Attention in Knowledge Markets: Electronic Document Dissemination in a Management Consulting Company. Administrative Science Quarterly 46(1), 1–28 (2001)
8. Szulanski, G.: The Process of Knowledge Transfer: A Diachronic Analysis of Stickiness. Organizational Behavior and Human Decision Processes 82(1), 9–27 (2000)
9. Zack, M.: Managing Codified knowledge. Sloan Management Review 40(4), 45–58 (1999)
10. Boland, R., Tenkasi, R.: Perspective Making and Perspective Taking in Communities of Knowing. Organization Science 6(4), 350–372 (1995)
11. Brown, J., Duguid, P.: The Social Life of Information. Harvard Business School Press, Boston (2000)
12. Nidumolu, S., Subramani, M., Aldrich, A.: Situated Learning and the Situated Knowledge Web. Journal of Management Information Systems 18(1), 115–150 (2001)
13. Nonaka, I., Konno, N.: The Concept of "Ba": Building a foundation for Knowledge Creation. California Management Review 40(3), 40–55 (1998)
14. Mayer, R.C., Davis, D.J., Shoorman, F.D.: An Integrative Model of Organizational Trust. Academy of Management Review 20(3), 709–734 (1995)
15. Moe, N.B., Šmite, D.: Understanding a lack of trust in Global Software Teams: a multiple-case study. Springer, Heidelberg (2007)
16. Jarvenpaa, S.L., Shaw, T.R., Staples, D.S.: Toward contextualized theories of trust: The role of trust in global virtual teams. Information Systems Research 15(3), 250–267 (2004)
17. Giddens, A.: The Consequences of Modernity. Stanford University Press (1990)
18. Smith, M.: Design and natural science research on information technology. Decision Support Systems 15(4), 251–266 (1995)
19. Drucker, P.F.: Post-capitalist society. Butterworth Heineman, New York (1993)
20. Nonaka, I., Takeuchi, H.: The knowledge creation company: How Japanese companies create the dynamics of innovation. Oxford University Press, New York (1995)
21. Howells, J.: Tacit knowledge, innovation and technology transfer. Technology Analysis & Strategic Management 8(2), 91–105 (1996)
22. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. In: International Workshop on Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic Publishers, Deventer, The Netherlands, Padova, Italy (1993)
23. OWL 2 Web Ontology Language Document Overview (2009), `http://www.w3.org/TR/owl2-overview/` (cited May 31, 2010)
24. Wang, J.Z., Ali, F.: An efficient ontology comparison tool for semantic web applications. In: The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2005). IEEE (2005)
25. Wongthongtham, P., et al.: Development of a software engineering ontology. IEEE Transactions on Knowledge and Data Engineering 21(8), 1205–1217 (2009)
26. Weinstein, P., Birmingham, P.: Comparing Concepts in Differentiated Ontologies. In: 12th Workshop on Knowledge Acquisition, Modeling, and Management, Banff, Canada (1999)

27. Stuckenschmidt, H., Timm, I.J.: Adapting communication vocabularies using shared ontologies. In: The Second International Workshop on Ontologies in Agent Systems, Bologna, Italy (2002)
28. Kashyap, V., Sheth, A.: Semantic Heterogeneity in Global Information Systems: The Role of Metadata, Context, and Ontologies. In: Schlageter, M.P.a.G. (ed.) Cooperative Information Systems: Trends and Directions. Academic Press, London (1998)
29. Mena, E., et al.: OBSERVER: An Approach for Query Processing in Global Information Systems Based on Interoperation across Preexisting Ontologies. Distributed and Parallel Databases 8(2), 223–271 (2000)
30. Pease, A., Niles, I., Li, J.: The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In: The AAAI-2002 Workshop on Ontologies and the Semantic Web, Edmonton, Canada (2002)
31. Rodriguez, M.A., Egenhofer, M.J.: Determining semantic similarity among entity classes from different ontologies. IEEE Transactions on Knowledge and Data Engineering (2003)
32. Andrew, B.-J., et al.: A semiotic metrics suite for assessing the quality of ontologies. Data Knowl. Eng. 55(1), 84–102 (2005)
33. Dazhou, K., et al.: A complexity measure for ontology based on UML. In: Proceedings 10th IEEE International Workshop on Future Trends of Distributed Computing Systems, FTDCS 2004 (2004)
34. Mungall, C.: Increased complexity in the GO (2005) BDGP / GO Consortium
35. Zhang, D.-l., Ye, C., Yang, Z.: An evaluation method for ontology complexity analysis in ontology evolution. In: Staab, S., Svátek, V. (eds.) EKAW 2006. LNCS (LNAI), vol. 4248, pp. 214–221. Springer, Heidelberg (2006)
36. Anthony, M.O., Haining, Y., Letha, H.E.: Indicating ontology data quality, stability, and completeness throughout ontology evolution: Research Articles. J. Softw. Maint. Evol. 19(1), 49–75 (2007)
37. His, I.: Analyzing the Conceptual Coherence of Computing Applications through Ontological Excavation (2004)
38. Anthony, M.O., Haining, Y., Letha, H.E.: Coupling Metrics for Ontology-Based Systems. IEEE Softw. 23(2), 102–108 (2006)

Dogsperate Escape: A Demonstration of Real-Time BSN-Based Game Control with e-AR Sensor

Kalanyu Zintus-art, Supat Saetia,
Varichmaes Pongparnich, and Surapa Thiemjarus

School of Information, Computer, and Communication Technology
Sirindhorn International Institute of Technology, Thammasat University
{5022781066,5022791016,5022780340}@studentmail.siit.tu.ac.th
surapa@siit.tu.ac.th

Abstract. Dogsperate Escape is a visually attractive exercise game that allows the player to perform various activities, provoking physical exercises with a memorable experience. This paper presents an implementation of the motion control module in the game based on a bio-inspired, accelerometer embedded e-AR sensor. The acceleration signals from the device are wirelessly transmitted to the PC for real-time activity analysis. A real-time activity recognition algorithm has been constructed based on pre-collected datasets acquired from ten subjects while performing a simple pre-defined activity routine. The algorithm is validated on eleven different subjects while performing another activity routine. By integrating the algorithm into the game, the recognized activities can be mapped into specific input commands for character control.

Keywords: sport game, Body Sensor Network, activity recognition, real-time motion analysis.

1 Introduction

In recent years, alternative interfaces to the traditional method of game control via a controller are gaining popularity, as can be witnessed from several products released by the world leading game manufacturers such as Sony[1], Microsoft[2], and Nintendo[3]. The underlying mechanism for these novel game interfaces is usually based on vision and/or sensor technology.

In [1], Crampton *et al.* developed a real-time vision-based finger counting system. The extracted number of fingers is used as an input command for game control. Hysniemi *et al.* [2] developed a game control system based on a web camera. In the study, left tilt and right tilt gestures are automatically detected, but

[1] Sony Computer Entertainment America LLC. Playstation move, 2010. http://us.playstation.com/ps3/playstation-move/
[2] Microsoft Corporation. Kinect for Xbox 360, 2010. http://www.xbox.com:80/en-US/kinect/
[3] Nintendo. Wii, 2010. http://www.nintendo.com/wii/console/controllers/

T. Theeramunkong et al. (Eds.): KICSS 2010, LNAI 6746, pp. 253–262, 2011.

(a) The title screen. (b) Obstacle avoidance in the tutorial stage.

(c) A different environment in the first stage. (d) Intense evasion gameplay in final stage.

Fig. 1. Snapshots of the title screen and the three different stages in the game

the punching movements are controlled by the Wizard of Oz approach. Vision-based technology provides a means for object recognition and motion tracking. A vision-based system has an advantage in terms of mobility since it frees users from carrying a controlling device. Well-known limitations of most vision-based systems, however, are the restricted field of view, occlusion and constraints in lighting conditions [3].

Body Sensor Network (BSN) [4,5] is a multidisciplinary research field that is designated to revolutionalize the diagnosis method in the traditional healthcare system towards continuous monitoring of patients' health status under their natural physiological conditions. In addition to healthcare applications, the technology has recently been extended for well-being, military, security, sports and gaming. A wireless BSN provides an alternative and complementary solution to the vision-based approach. It outcomes the constraint earlier described, allowing the user to move freely. As a demonstration of the potential value of BSNs in gaming application, Crampton *et al.* [6] proposed the use of a full-body accelerometer-based input system for game control. It was envisaged by the authors that the technology can be used as an alternative method to the binary foot-pad used in an interactive dance game such as Dance Dance Revolution. In [7], Heinz *et al.* conducted an initial experiment on real-time recognition of martial arts movements based on body-worn gyroscopes and accelerometers. The XBus Master System (XM-B-XB3)[4], developed by XSens, was used as the global sensor control.

[4] XSens. http://www.xsens.com/en/general/xbus-kit/

In this study, we developed a game called Dogsperate Escape using a BSN device attached to the user's ear for game control. Based on the Micro-Electro-Mechanical System (MEMS) technology, the low-power, miniaturized and light-weight sensor device provides a pervasive hardware platform suitable for an interactive game input. The acceleration signals from the device are wirelessly transmitted to the PC, at which real-time analysis is performed. An activity recognition algorithm has been constructed based on a pre-collected acceleration data acquired from ten subjects and validated on twenty different subjects while performing another pre-defined activity routine. The algorithm has been integrated into the game for triggering specific input commands.

This paper is organized as follows: Section 2 introduces the game along with its implementation and the development tools. Wireless BSN-based game control mechanism, training data collection, and algorithm validation are presented in Section 3. This includes the description of the input device, the protocol of training data collection and the real-time activity recognition algorithm. Experiment setup and results are described in Section 4. Section 5 concludes this paper.

2 Dogsperate Escape

Dogsperate Escape is an adventure game in which the player controls a dog character to avoid obstacles that appear along the path while running towards the target destination. Along the journey, the character will be chased by an evil cat. The player can control the character's basic movements such as turn left, turn right, walking, running, and jumping. When bumped against obstacles, penalties will be set upon the dog character, resulting in speed reduction, blurry graphics, or switching between the left and right motion control. In order to win the game, the player has to complete all the stages without getting caught by the evil cat. Figure 1 shows the snapshots of the title screen and the three different stages in the game. Each stage is designed with different environmental scenes and obstacles to make the game more interesting. The game engine is developed in C# programming language using Microsoft's XNA Game Studio[5]. The XNA framework is a set of tools which provide a managed runtime environment for helping developers to avoid repetitive boiler plate coding tasks. The graphical models of the game characters and environmental objects are created using 3Ds Max[6]. The game can be played both via keyboard and e-AR sensor.

3 BSN-Based Game Control

In this section, the e-AR sensor is first introduced, followed by the description of the protocol in which the training datasets are collected. The last subsection describes the learned activity recognition algorithm.

[5] Microsoft Corporation. Beginner's guide to XNA game studio, 2010. http://creators.xna.com/en-US/education/gettingstarted/bgintro/chapter1

[6] Autodesk, Inc. 3Ds Max features, 2010. http://www.autodesk.com/3dsmax

3.1 e-AR Sensor

e-AR sensor, as shown in Figure 2, is a low-power BSN-based sensor device developed by Sensixa[7]. Equipped with a 3D accelerometer, the device is capable of detecting a range of indices: gait cycle, steady/unsteady locomotion, acceleration in three dimensions, and spinal/joint shock wave transmission. The design of the e-AR sensor imitates the rotational and translational sensing functionality of the balancing system in human inner ear. By placing the e-AR sensor on the ear, information regarding postures and activities of the wearer can be analyzed. The e-AR sensor has been successfully deployed in several healthcare-related applications such as behavior profiling [8], gait analysis [9], activity recognition [10], pre- and post-operative monitoring [11], and sport performance assessment [12]. In game control, a miniturized wearable devices such as the e-AR sensor has an advantage over the existing sensor-based systems such as Playstation Move, Nintendo Wii and Apple iPhone since it does not require user to hold the device, thus allowing more natural interaction with the game. Compared to the existing vision-based systems such as Microsoft Kinect, the e-AR sensor has more flexibility in terms of operating conditions. The wireless device has no field of view restriction and is resilient to different lighting conditions.

Fig. 2. e-AR Sensor

3.2 Collection of Training Data

Our goal is to construct a real-time recognition algorithm for classifying the six activities, namely, standing, left tilt, right tilt, walking, running, and jumping. To achieve this, acceleration datasets were collected from 10 subjects, aged between 20 and 25 years old, at sampling rate of 50 Hz. The test subjects were asked to wear the e-AR sensor on the right ear while performing the following activity routine:

[7] Sensixa Ltd. e-AR sensor, 2007. http://sensixa.com/

– stand for 5 seconds
– left tilt for 5 seconds
– right tilt for 5 seconds
– jump 1 time
– walk for 5 seconds
– run for 5 seconds

The collected data were manually annotated and analyzed for appropriate features and threshold values to be used in the recognition algorithm.

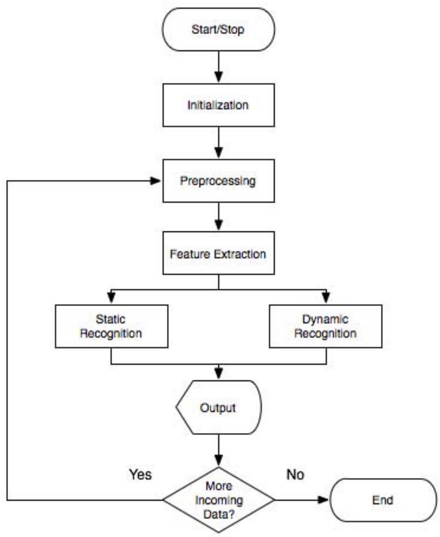


Fig. 3. A flowchart showing different steps of the activity recognition process

3.3 An Activity Recognition Algorithm

Our activity recognition system uses a rule-based approach in which incoming signals are checked against a pre-constructed set of rules in real time to find the best match out of the 6 activities, namely standing(head straight), left tilt, right tilt, walking, running, and jumping. The first three activities are classified as static while that last three are classified as dynamic. The two groups of activities are separately analyzed simultaneously. The static activities are analyzed instantaneously (i.e. every data sample). Temporal information is required in classifying dynamic activities and therefore this group of activities is analyzed at a regular time interval. Figure 3 illustrates the workflow of the proposed activity recognition algorithm.

Calibration and Preprocessing. At the beginning of the game, initial calibration is required since the orientation of the accelerometer can be slightly varied among subjects. At the initialization stage, each user was asked to stand still for approximately 2 seconds when the game start. From the first 100 data points collected from the accelerometer, the standard deviation along x- and y-axes and signal median are calculated. To reduce the effect of different screen positions, along with subject and sensor variations, the incoming signals are normalized as follows:

$$\frac{S(t) - \mu}{max_{i \in \{1,2,3,....,100\}}(S(i))} \tag{1}$$

where S is the signal in time t, and μ is the median of the initial data segment S. Figure 4 shows a plot of acceleration data acquired from a specific subject after normalization.

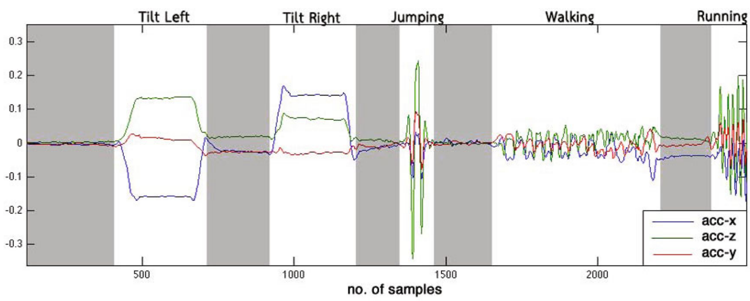


Fig. 4. Normalized acceleration signal along 3 axes

Feature Extraction. Three features are extracted from the acceleration signals: 1) the normalized signal value along x-axis (sideward), $S_x(t)$, 2) the standard deviation along y-axis (frontward), σ_y, and 3) the number of crossing points over a specific threshold along the y-axis , C_y. The crossing point threshold is specified to half of the maximum signal amplitude below the mean within a signal window. $S_x(t)$ is important for classifying static activities such as standing straight, left tilt and right tilt. σ_y and C_y are useful for analyzing dynamic activities such as walking, running, and jumping. The extraction of features used for dynamic activity classification is illustrated in Figure 5.

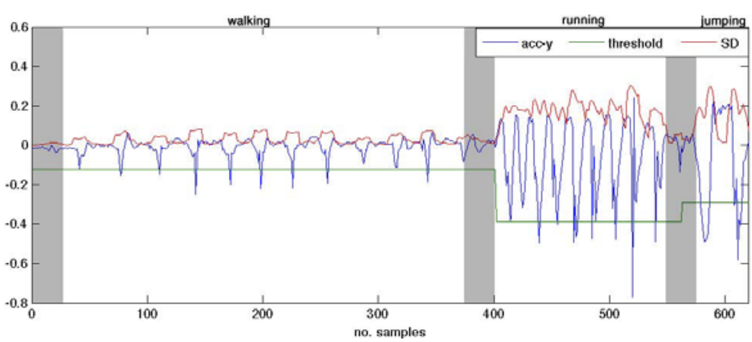


Fig. 5. An illustration of the feature extraction process for dynamic activities

Activity Recognition. In this process, every incoming acceleration signal datapoint is processed by the following mathematical rules.

For standing straight:

$$\sigma_x \times \alpha \leq S_x(t) \leq -\sigma_x \times \alpha \Rightarrow \text{Stand} \tag{2}$$

For left tilt:

$$S_x(t) > \sigma_x \times \alpha \Rightarrow \text{Left tilt} \tag{3}$$

For right tilt:

$$S_x(t) < -\sigma_x \times \alpha \Rightarrow \text{Right tilt} \tag{4}$$

where $S_x(t)$ is the acceleration value along x-axis at time t, σ_x is the standard deviation of the first 100 signal values, and α is the scaling factor. The value of α depends on the sensitivity level of the activity detection algorithm. The higher the value of α implies that the algorithm is less sensitive to the tilt postures and an additional effort is required in performing the posture. When α is too low, it is difficult for the player to maintain straight direction. After multiple testings, we found that α equals to 0.3 is an optimal value.

For dynamic activities, the following set of rules is used to find the best matching activity.

For walking:

$$(\beta_s \leq \sigma_y \leq \beta_w) \wedge (C_y \geq \phi_w) \Rightarrow \text{Walk} \tag{5}$$

For running:

$$(\beta_w \leq \sigma_y \leq \beta_r) \wedge (C_y \geq \phi_r) \Rightarrow \text{Run} \tag{6}$$

For jumping:

$$(\beta_w \leq \sigma_y \leq \beta_j) \wedge (C_y \leq \phi_j) \Rightarrow \text{Jump} \tag{7}$$

where σ_y is the standard deviation of the window along y-axis. $\beta_s, \beta_w, \beta_r$, and β_j are the standard deviation thresholds for standing still, walking, running, and jumping activities, respectively. ϕ_w, ϕ_r, ϕ_j are the crossing point thresholds for walking, running, and jumping activities, respectively.

Based on the training data, the standard deviation and crossing point thresholds were extracted from the maximum values of standard deviation and the minimum values of the crossing points within each activity. Table 1 summarizes the threshold values used in the proposed algorithm.

Table 1. The threshold values used for dynamic activity classification

Activity	**SD(β)**	**Crossing Count(ϕ)**
Standing	0.005	0
Walking	0.034	3
Running	0.154	4
Jumping	0.115	2

To recognize the dynamic activities, a sliding window of size 1 second, with a 0.5 second shift is used. This time interval was estimated from the highest possible speed observed for a jump, which is approximately once per second.

4 Model Validation and Experimental Results

To validate the performance of the proposed activity recognition algorithm, another group of 11 subjects,aged between 5 and 33 years old, were asked to wear the e-AR sensor on any ear while performing the following activity routine:

stand - run - tilt left - tilt right - walk - jump - run - stand - tilt left - walk - tilt right - jump - tilt left - walk - tilt right - jump - stand - run

The activity recognition algorithm has been implemented in a testing software which shows the recognition results on the screen monitor in real-time. During the experiment, the results displayed on the monitor were recorded by an observer, while another observer kept the record of the actual activities performed by the test subjects. The two results are then compared to assess the accuracy of the algorithm.

Table 2. Confusion matrix shows the performance of the proposed real-time activity recognition algorithm when applied on 11 test subjects

	Standing	**Left Tilt**	**Right Tilt**	**Walking**	**Running**	**Jumping**
Standing	33	0	0	0	0	0
Left Tilt	0	33	0	0	0	0
Right Tilt	0	0	33	0	0	0
Walking	0	0	0	32	0	0
Running	0	0	0	1	33	1
Jumping	0	0	0	0	0	32
Accuracy	100%	100%	100%	96.96%	100%	96.96%

Table 2 illustrates the confusion matrix of the obtained results. It can be observed that the static activities can classified with almost perfect accuracy. One instance of the walking activity were misclassified as running, which can be because of the close relationship between the two activities. Different people have different walking and running speeds. Fast walking performed by a subject could be considered as running in another. A misclassification from jumping to running can be due to the bouncing effect after the jump. In average, an overall accuracy of 98.987% was achieved.

The activity recognition algorithm was then integrated into Dogsperate Escape for usability testing. Several subjects were asked to play the game freely. The recognition time of activities was observed by comparing the actions performed by the game character with the actual activities performed by the player. Since the game allows player to only jump or move in left and right directions,

Table 3. Recognition time for activities used in the game

Activity	Recognition Time
Left Tilt	$\leq$ 0.3 second
Right Tilt	$\leq$ 0.3 second
Jumping	$\leq$ 1 second

walking and running are not tested in the game. It has been observed that switching the sensor location from right to left ear does not affect the recognition accuracy of the proposed algorithm. After the subjects finished playing the game, comments and suggestions were collected for further improvement of the system. Table 3 shows the recognition time for activities used in the game.

5 Conclusions

This paper presents an interactive game control based on a wireless BSN device equipped with a 3D accelerometer. A real-time activity recognition algorithm for detecting basic movements of the user has been proposed. The use of temporal features implies that there still exists some delay in the game character's response to user input. However, from user feedback, once the window size is reduced to below 1 second, the response time can satisfy most players who played the game. Further reduction of window size can lead towards faster response time in a trade of recognition accuracy.

In general, the experimental results show that the algorithm performs well in a real-time gaming environment. The players can control the character in the game with acceptable accuracy regardless of their location. While playing the game, the players need to perform multiple activities such as tilt left, tilt right, and jumping continuously, giving them a good exercise while having a fun and exciting experience through the gameplay.

Acknowledgements. Thanks to the Sirindhorn International Institute of Technology students and staffs for their support during data collection. The e-AR sensor used in this study is funded by Telecommunication Research and Industrial Development Institute (TRIDI) and the National Telecommunications Commission (NTC), Thailand. This research is also supported in part by the Thailand Research Fund (TRF).

References

1. Cramption, S., Betke, M.: Counting fingers in real time: a webcam-based human-computer interface with game applications. In: The Conference on Universal Access in Human-Computer Interaction, pp. 1357–1361 (2003)
2. Höysniemi, J., Auvinen, P.: Shadow Boxer: a physically interactive fitness game. In: The Third Nordic Conference on Human-Computer Interaction, Tampere, Finland, pp. 389–392 (2004)

3. Jaijongrak, V., Thiemjarus, S., Kayasith, P., Kumazawa, I.: American sign language fingerspelling gesture recognition: developmental progress of a hand gesture interface for speech-impaired disabilities using body sensor networks. In: The International Conference on Information and Communication Technology for Embedded Systems, Pathumthani, Thailand (2010)
4. Yang, G.-Z.: Body Sensor Networks. Springer, London (2006)
5. Lo, B.P.L., Thiemjarus, S., King, R., Yang, G.-Z.: Body sensor network a wireless sensor platform for pervasive healthcare monitoring. In: The Third International Conference on Pervasive Computing, Munich, Germany, pp. 77–80 (2005)
6. Crampton, N., Fox, K., Johnston, H., Whitehead, A.: Dance, Dance Evolution: accelerometer sensor networks as input to video games. In: IEEE International Workshop on Haptic, Audio and Visual Environments and Games, Ottawa, Ontario, pp. 107–112 (2007)
7. Heinz, E., Kunze, K., Gruber, M., Bannach, D., Lukowicz, P.: Using wearable sensors for real-time recognition tasks in games of martial arts: an initial experiment. In: The IEEE Symposium on Computational Intelligence and Games, Reno, Nevada, pp. 198–102 (2006)
8. Ali, R., ElHelw, M., Atallah, L., Lo, B., Yang, G.: Pattern mining for routine behavior discovery in pervasive healthcare environments. In: The Fifth International Conference on Information Technology and Applications in Biomedicine, Shenzhen, China, pp. 241–244 (2008)
9. Atallah, L., Aziz, O., Lo, B., Yang, G.: Detecting walking gait impairment with an ear-worn sensor. In: The Sixth International Workshop on Wearable and Implantable Body Sensor Networks, Berkeley, CA, USA, pp. 175–180 (2009)
10. Ali, R., Atallah, L., Lo, B., Yang, G.: Transitional activity recognition with manifold embedding. In: The Sixth International Workshop on Wearable and Implantable Body Sensor Networks, Berkeley, CA, USA, pp. 98–102 (2009)
11. Lo, B., Atallah, L., Aziz, O., Elhelw, M., Darzi, A., Yang, G.: Real-time pervasive monitoring for post-operative care. In: The Forth International Workshop on Wearable and Implantable Body Sensor Networks, Aachen, Germany, pp. 122–127 (2007)
12. Pansiot, J., King, R.C., McIlwraith, D., Lo, B., Yang, G.: ClimBSN: climber performance monitoring with BSN. In: The Fifth International Workshop on Wearable and Implantable Body Sensor Networks, Hongkong, China, pp. 33–36 (2008)

Author Index