지능정보연구 제18권 제2호 2012년 6월(pp.1~17)

Mobile Device and Virtual Storage-Based Approach to Automatically and Pervasively Acquire Knowledge in Dialogues

Keedong Yoo
Division of Management, Dankook University
(kdyoo@dankook.ac.kr)

The Smartphone, one of essential mobile devices widely used recently, can be very effectively applied to capture knowledge on the spot by jointly applying the pervasive functionality of cloud computing. The process of knowledge capturing can be also effectively automated if the topic of knowledge is automatically identified. Therefore, this paper suggests an interdisciplinary approach to automatically acquire knowledge on the spot by combining technologies of text mining-based topic identification and cloud computing-based Smartphone. The Smartphone is used not only as the recorder to record knowledge possessor's dialogue which plays the role of the knowledge source, but also as the sensor to collect knowledge possessor's context data which characterize specific situations surrounding him or her. The support vector machine, one of well-known outperforming text mining algorithms, is applied to extract the topic of knowledge. By relating the topic and context data, a business rule can be formulated, and by aggregating the rule, the topic, context data, and the dictated dialogue, a set of knowledge is automatically acquired.

Received : April 06, 2012 Revision : April 12, 2012 Accepted : April 18, 2012
Type of Submission : English Fast-track Corresponding author : Keedong Yoo

1. Introduction

To capture knowledge on the spot with its related context data, the subject who possesses and utilizes knowledge must be monitored continuously, nearly around the clock. A knowledge worker, or a knowledge possessor, who has the knowledge to be captured performs his or her

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(No.H00021).

jobs using special, and so preserving-worthy, knowledge. His or her dialogues, writings, and activities exhibited during performing jobs are the live sources to be managed with care. Applying proper technologies to each source, fully-automated, and so autonomous, as well as pervasive acquisition of knowledge can be achieved (Park et al., 2005; Yoo, 2011). The context data which plays the role of the meta-knowledge can also be captured by monitoring the knowledge worker.

The context data can be defined as any information characterizing the situation of a task session or interaction between a user and his or her service world (Zhou et al., 2011). The context data as the meta-knowledge can be understood as the data explaining the situation of the knowledge worker, or the knowledge possessor, who is possessing knowledge and therefore is utilizing it in performing his or her ordinary job; therefore it is a kind of the user context. User context can be summarized into three categories (Schmidt et al., 1998) : information of the user (e.g. knowledge of habits and emotional state), the user's social environment (e.g. co-location of others, social interaction, and group dynamics), and the user's tasks (e.g. spontaneous activity, engaged tasks, and general goals). To characterize the given knowledge, several types of context data, such as the identity that explains who the knowledge worker is as well as which area he or she specialized in, the time and location he or she uses knowledge, the event (schedule) the given knowledge is to be applied, and the topic (keyword) the exhibited knowledge is about, etc., are combined together.

Mobile devices, such as Smartphone and Smartpad, are widely used nowadays because of their multiple functionalities based on wireless data communications and open APIs (application programming interfaces). As developers of mobile devices design them to be more light, small, and versatile, mobile devices tend to omnipotent enough to be a mandatory item of human life : every individual carries and uses them anytime and anywhere. Even in performing ordinary jobs, individuals carry one or more mobile devices in their hands or pockets, therefore mobile devices can be used as a sensor to monitor knowledge workers' dialogues, writings, and activities in a real-time basis (Bae, 2010). Once a mobile device monitors and records what a knowledge worker talks, writes, and acts, then it can also transmit them to a server whose address has been designated in advance by the user. Using the paradigm of cloud computing, a networking model for enabling convenient on-demand network access to a shared pool of configurable computing resources (Pallis, 2010), transmitted data can be analyzed so that the meaning of each, which means the subject or topic of them, can be identified. Context data can be also captured and transmitted to a server using mobiles devices in a very convenient way, because most of context data have been preset and stored in the mobile device using a user profile and a scheduler. By relating the topic and context data, a business rule can be formulated, and by aggregating the rule, the topic, context data, and the transmitted data, a set of knowledge is concluded.

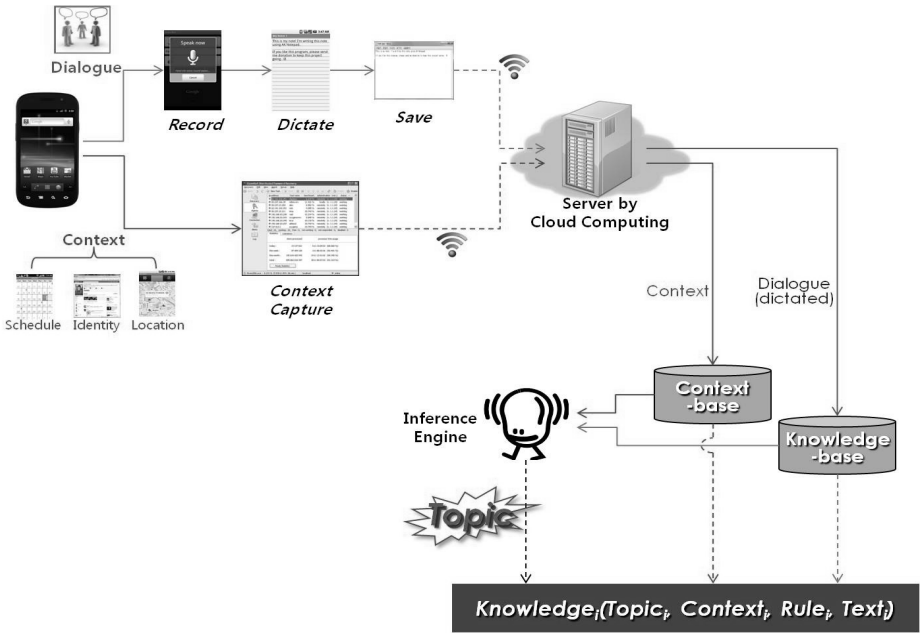
The objective of this paper is to propose

a framework to capture knowledge in ordinary dialogues immediately when and where it is exhibited in an automated and pervasive manner. To pervasively capture knowledge, the Smartphone which is operated on cloud computing and plays the role of a sensor to monitor and gather dialogue-based knowledge and its context data is applied. Also, to automatically conclude the meaning of knowledge, the support vector machine (SVM), one of well-known outperforming text mining algorithm, is applied.

2. Research Framework

2.1 Overview

In this research, the Smartphone acts as a sensor to record conversations, the source of knowledge, in a dialogue as well as to capture context data, the meta-knowledge. The STT (Speech-to-Text) application installed in the Smartphone initiates its function of dictation by transmitting the recorded data to the vendor-providing server for speech recognition. Dictated results are again sent to the Smartphone and displayed on the screen to wait user's confirmation. Time for completing dictation can be various depending on the condition of network traffic, however in most cases dictation can be completed within a few seconds averagely. Dictated results are usually saved in text format, and they are transmitted to a pre-designated server operated on the cloud computing environment. Context data are also transmitted to the server and temporarily saved in the server. At the server, the text-based electronic document containing the



<Figure 1> Research Framework

contents of dialogues is analyzed by SVM classifier to extract its topic. Once the topic of the document, that is the topic of the dialogue, is identified, the system stores the texts in the document with respect to the topic and the context data. If any business rules related with the topic exist, then they are extracted from a knowledge base to be stored together with the topic, the context data, and the text as a set of knowledge. <Figure 1> shows the proposed framework of this research.

2.2 Business Rule-based Knowledge

Once the dictated dialogue and context data captured by the Smartphone are transmitted to a server operated on a cloud computing basis, further treatment to extract the topic (keyword) of the transmitted data is processed. By extracting topics of the transmitted text-based data, each document can be classified in terms of the extracted topics or keywords. The document can be deemed as knowledge itself, because it may contain various kinds of knowledge applicable to other similar situations. Because the topic of the document is the topic of knowledge, knowledge possessors' knowledge can be categorized according to the topic. Besides the topic resulted from analyzing the document, knowledge possessors' context data which have been identified and transmitted from the context acquisition subsystem must be applied to conclude the relationship between the knowledge and the situation that the knowledge is deployed. Context data, such as knowledge possessors' identities, location, time, and schedule, can play the role of the meta-knowledge by uniquely depicting the knowledge used in the given situation. By associating the document and context data, knowledge can be stored in the knowledge base with the form of a business rule.

Suppose a case that 'Professor A' is discussing about how to make payment for a purchased book that is expensive but is necessary for his research. General payment method is known as one of 'research fund', 'personal expense', and 'mileage points.' If the theme (topic) of the book is related with his research then he can use the 'research fund'; otherwise payment must be made through one of 'personal expense' and 'mileage points.' In this case, the most important information that determines the method of payment is the theme of the book. Next important information is the balance of the research fund, because either of the rest, personal expense and mileage point, must be unavoidably considered if the balance is not enough. Possible business rules for this situation are illustrated in <Figure 2>.

```
Rule1:
IF Theme_of_Book = Theme_of_Research THEN Method_of_Payment = 'Research_Fund'
                ELSE Method_of_Payment = ['Personal_Expense' | 'Mileage_Point']
Rule2:
IF Balance_of_Research_Fund >= Price_of_Book THEN Method_of_Payment = 'Research_Fund'
                ELSE Method_of_Payment = 'Research_Fund' + ['Personal_Expense' | 'Mileage_Point']

Theme_of_Book = [Title_of_Book | Keyword_of_Book]
Theme_of_Research = [Title_of_Research | Keyword_of_Research]
Method_of_Payment = ['Research_Fund' | 'Personal_Expense' | 'Mileage_Point']
Title_of_Book = "The title of the book revealed in the cover page"
Keyword_of_Book = "The keywords of the book defined by the authors and revealed in the bibliographic page"
Title_of_Research = "The title of research as an instance of the class 'Research' stored in a database"
Keyword_of_Research = "The keywords of the research as an instance of the class 'Research' stored in a database"
Balance_of_Research_Fund = "Amount of research fund remained and stored in a database"
Price_of_Book = "The price of the book revealed in the bibliographic page"
```

<Figure 2> Example Rules

Information included in these rules cannot be directly determined by simply monitoring the dialogue. The context data, such as the identity of the participant and the title of the book, which can be directly obtained by monitoring the dialogue, can constitute required information to form rules through the causal chain relationship. For example, the data of 'Professor A', the identity of the participant, can determine the title and keyword of the research by querying corresponding research he is responsible for. The title of the book also underpins the business rule, because it addresses the keywords and the price of the book. Causal chain relationships between the context data and the rule-constituting information can be identified as <Figure 3> shows.

```
Relationship1:
'Identity' determines 'Title_of_Research' AND 'Keywork_of_Research' AND 'Balance_of_Research_Fund'
Relationship2:
'Title_of_Book' determines 'Keyword_of_Book' AND 'Price_of_Book'

Identity = "Name and ID number of the participant"
```

<Figure 3> Relationship between Context Data and Rule-Constituting Information

Meanwhile, the topic of the dialogue characterizes the business rule by identically depicting which rules can be alternatively related with the event. The topic of the dialogue cannot be regarded as the context data, because it cannot be directly obtained by monitoring the dialogue but because can be identified by further analysis. Context data are used in triggering the exact business rules by defining the causal chain relationships between rule-related information.

Once the topic of the dialogue is identified, a set of knowledge can be automatically formulated by combining the topic, context data, and the text-based document containing the contents of the dialogue. If the topic of the dialogue is concluded as the 'Book Payment', then a set of knowledge can be expressed and stored as <Figure 4>. In this case, the identity of the discussion participant and the title of the book play the role of context data, because they define one of the most important information to trigger rules. The context data-based rule is used to enhance the understandability of the resultant knowledge, because future users may have difficulty in understanding the knowledge by reading the contents of the document only.

```
Identity[i] = 'Professor A'
Title_of_Book = 'XXXXXXX'
Rule[i] = 'Rule1' + 'Rule2'
Document[i] = 'Document001'
"text-based document containing the contents"
Rule1 = IF Theme_of_Book = Theme_of_Research THEN Method_of_Payment = 'Research_Fund'
                    ELSE Method_of_Payment = ['Personal_Expense' | 'Mileage_Point']
Rule2 = IF Balance_of_Research_Fund >= Price_of_Book THEN Method_of_Payment = 'Research_Fund'
                    ELSE Method_of_Payment = 'Research_Fund' +['Personal_Expense' | 'Mileage_Point']

IF Topic[i] = 'Book_Payment' AND Context[i] = Identity[i] + Title_of_Book[i]
THEN INSERT INTO Knowledge[i] = Knowledge(Topic[i], Context[i], Rule[i], Document[i])
```

<Figure 4> Rule to Store Resultant Knowledge

2.3 SVM as the Classifier

The indexing and retrieval algorithm of information retrieval technology can be applied to automatically identify the topic of each electronic document and classify it according to predefined knowledge categories. Identifying the topic of knowledge is important in that the topic (or the keyword) indicates the subject of knowledge embedded in the document. To extract the topic of knowledge based on predefined knowledge categories, text mining techniques can be employed (Kwon and Lee, 2003; Kim and Ahn,

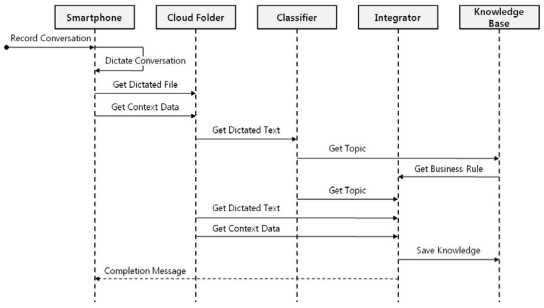
2010).

The topic of an electronic document is used further for storing the document in a knowledge repository. This topic can indicate the situation (or the context) of the knowledge holder and play a role as meta-knowledge that can directly explain the content of the knowledge embedded in the document. By combining the topic with other context data, a specific piece of knowledge can be identically defined. In other words, contextual knowledge can be obtained by defining knowledge according to related context data. Based on such context data which define the context of knowledge use, knowledge can be accurately selected and automatically serviced to the user if the user's context data explaining the situation that surrounds the user have been captured and concluded, that is, an autonomous process for the knowledge service, from the automated acquisition of knowledge to its proactive dissemination, can be realized.

Among various algorithms of text mining, this research deploys SVM for its adequacy of application. SVM try to find the hyperplane which produces the greatest possible margin among the boundary points to separate positive and negative training samples. SVM is adequately applicable to the topic (or the keyword) classification for several reasons : a high dimensional input space, few irrelevant features, sparse document vectors, and linearly separable characteristics of text categorization problems (Joachims, 1998). Based on these characteristics, several researches have proved SVM outperforms other machine learning algorithms with respect to efficiency as well as accuracy (Meyer et al., 2003; Basu et al., 2002; Rennie and Rifkin, 2001; Joachims, 1998; Dumais et al., 1998).

3. Prototype Implementation

The prototype is developed using JDK v1.5.0_ 06 under Java2 runtime environment to enhance the high interoperability and the ease of implementation. It initiates knowledge acquisition by recording and dictating knowledge possessors' dialogues using STT module of Android v2.3.3 (Gingerbread) keyboard. <Figure 5> shows how the prototype works.



<Figure 5> Sequence Diagram of the Prototype

For the effectiveness of explaining how the prototype system works, suppose an example dialogue about a research proposal between two research associates, as follows :

The Prototype initiates its process by providing 'Recording Standby' signal to users, and begins recording as users start speaking. When speaking is finished or paused, dictated results

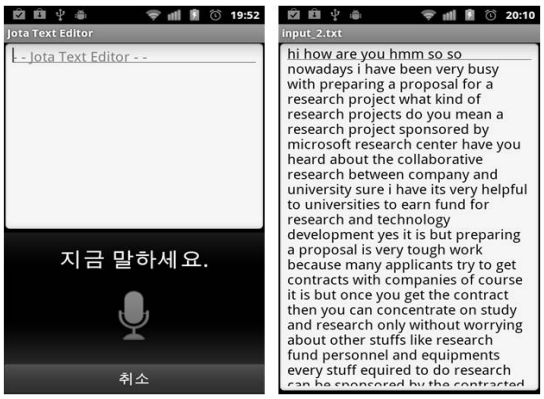
Example Dialogue Concerning How to Make a Research Proposal

Emily :	Hi, how are you?
Betty :	Hmm, so so, nowadays I have been very busy with preparing a proposal for a research project.
Emily :	What kind of research projects do you mean?
Betty :	A research project sponsored by the Microsoft research center. Have you heard about the collaborative research between a company and a university?
Emily :	Sure I have. It's very helpful to universities to earn fund for research and technology development.
Betty :	Yes it is. But preparing a proposal is very tough work, because many applicants also try to get contracts with companies.
Emily :	Of course it is. But once you get the contract then you can concentrate on study and research only without worrying about other stuffs, like research fund, personnel, and equipments. Every stuff required to do research can be sponsored by the contracted company.
Betty :	Then do you have an idea about an attractive proposal?
Emily :	Let me see. First, you must concretely specify the goal or objective of your research project by listening target company's current situation.
Emily :	What the company aims to get from the research project, must what problems the company want to solve through the collaboration with you might be one of the key considerations, I think.
Betty :	You got it. Then what about the research fund? I mean, how I can estimate the amount of fund sponsored by the company.
Emily :	Just follow the directions provided by the company.
Betty :	As I know, the Microsoft estimates the amount of research fund based on the number of people who
Emily :	participate in the research project and the amount of equipments that are required to be set. Of course the lower the better, I mean too much amount of research fund can be doubted. Sharing research resources like people and equipments is one important reason for collaborative research project. I can give you a sample of the research fund estimation once I submitted to the cMicrosoft. It will be helpful to you.
	Yes please, it must be very helpful to prepare my research proposal. Thank you very much.
	You're welcome. Good luck!

are displayed and the mode of 'Recording Standby' is again activated. When the user terminates the process of recording, it automatically transmits the dictated data to the cloud folder of the designated server operated on the cloud computing environment. The cloud folder, or the cloud storage, is one of widely known cloud computing technologies, and initiates its function by providing the Internet-based data storage as a service. One of the biggest merits of cloud storage is that users can access data in a cloud anytime and anywhere, using any device (Liu et al., 2011). Typical examples of cloud storage services are Amazon S3[1], Mosso[2], Wuala[3], or uCloud[4]; All of these services offer users clean and simple storage interfaces, hiding the details of the actual location and management of resources (Pamies-Juarez, 2011). Once a document to be archived is stored in a cloud folder, users can access and download it anytime and anywhere if the right to access has been granted. Also in terms of processing capability, the data stored in a cloud folder can be more efficiently accessed and analyzed by servers, because most of mobile devices do not equip sufficient computing capacity enough to extract outputs in a real time basis. Because of such advantage in utilizing organizational information resources, more companies and organizations are implementing the online storage under the cloud computing

1) http://aws.amazon.com/s3.
2) http://www.rackspacecloud.com.
3) http://www.wuala.com.
4) http://www.ucloud.com.

environment. <Figure 6> shows the recording and dictation process.

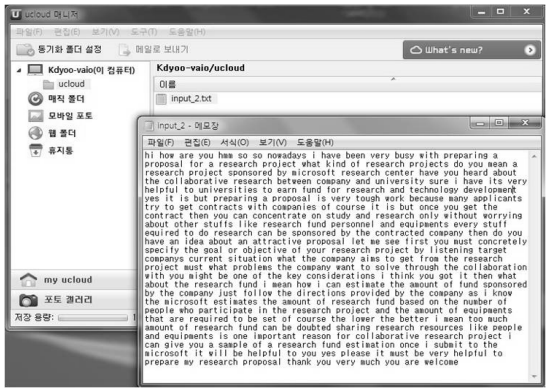


<Figure 6> Recording standby(left) and Dictation Completed(right)

When transmitting the dictated text file to the server, the Smartphone also simultaneously transmits corresponding context data which depict when and where the file has been made as well as whom the file has created by. <Figure 7> shows the context data stored and obtained by the Smartphone.

<Figure 7> Context Data Stored in Smartphone (user profile(left), Location(middle), and Schedule(right))

Once the text file and context data are transmitted and saved in the cloud folder of the server, these data can be synchronously accessed using a laptop or a desktop and the Smartphone. Analyzing the text file to extract the topic must be processed at the server-side, because currently the capacity of Smartphone is not enough to perform the job. <Figure 8> shows the view of the text file transmitted to and stored in the server.

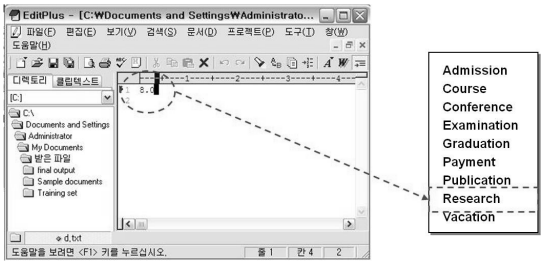


<Figure 8> Text File Stored in the Server

The text file stored in the server is to be accessed via a laptop or a desktop to perform the topic identification by sequentially conducting it into the word stemmer, the word vector tool, and the SVM-based classifier. This research used the LibSVM v2.81 (Chang and Lin, 2001) to implement the SVM classification. Word stems and vectors must be provided before the LibSVM (the classifier) performs its job. Therefore the word stemming and vector creating tool of Yale (Mierswa et al., 2006), a free open-source environment for KDD and machine learning, was employed. The result from the SVM module (<Figure 9>) is extracted as the number that stands for each category, which denotes the correspond-

ing topic. There exist 9 categories in this research, and the result indicates the numbers of each category. In the figure, '1 8.0' means 'document no.1 belongs to the eighth category', in other words, the topic of document no.1 is 'research.'

Finally, a set of knowledge must be stored in a knowledge base according to the identified topic and related context data. Since the topic plays role of the specific category to store knowledge, the context data, the document containing dictated conversations, and business rules (if any) must be moved under the topic.



<Figure 9> Resultant Topic of the Example Dialogue

Finally, a set of knowledge must be stored in a knowledge base according to the identified topic and related context data. Since the topic plays role of the specific category to store knowledge, the context data, the document containing dictated conversations, and business rules (if any) must be moved under the topic. '<Figure 10>' shows the programming codes to move the document under the resultant topic category. After completing storing the document under the destination category, the input document temporarily stored in the cloud folder needs to be deleted for the efficient management of the server.

```
// move input_document(dictated conversation) to resultant category--------------------

            // read the result category no
            BufferedReader result = new BufferedReader(new FileReader("temp_for_libsvm/KM_OutputFile"));
            String resultCategoryNo = result.readline();
            String resultCategoryTitle = null ;

            // find destination directory
            for ( int at = 0 ; at < kmCategoryInfo.size() && resultCategoryNo != null ; at++ ){
                        categoryInfo tempcate = (categoryInfo)kmCategoryInfo.elementAt(at);
                        if ( resultCategoryNo.equals(tempcate.getCategoryNo() ) ){
                                    resultCategoryTitle = tempcate.getCategoryTitle() ;
                                    break ;
                        }
            }

            // copy file to destination category
            int tempi, len=0;
            InputStream in = null;
            OutputStream out=null;

            try{
                        in= new FileInputStream(new File("input_document/"+argv[0]));
                        out= new FileOutputStream("classified_documents/"+resultCategoryTitle+"/"+argv[0],false);
                        while((tempi=in.read())!= -1){
                        out.write(tempi);
                        len++;
            }

    in.close();
    out.close();

            //  delete input file
            File delFile = new File ("input_document/"+argv[0]);
            System.out.println(delFile.getAbsoluteFile()  );
            if( !delFile.delete() ) System.out.println("file is not deleted");
            }catch(Exception ee){System.out.println(ee.getMessage() );}
```

<Figure 10> Programming Codes to Store Dictated Conversation('input_document') under the Identified Topic

4. Performance

The performance of the prototype system needs to be evaluated with respect to two factors: the extent of automation and the accuracy of acquisition. To examine the performance with considering these factors, the focus group interview as the qualitative approach and the statistical analysis as the quantitative approach were performed.

4.1 Focus Group Interview to Measure the Extent of Automation

The primary purpose of this research is to fully automate the process of knowledge acquisition. Therefore examining the extent of automation in the entire process for acquiring knowledge is necessary. The extent of automation can be regarded as a qualitative measure, because the extent can be decided by users and because it can be differently felt from user to user. Of course it is clearly observed that the entire

<Table 1> Evaluation Results from Focus Group Interview

Member	# of Trials	Rating Pts	Features		
			Fully-automated	Semi-automated	None-automated
1	3	6	• STT App-based dictation • Topic identification • Knowledge storing	• Data transmission	
2	3	6	• STT App-based dictation • Topic identification	• Knowledge storing	
3	5	7	• STT App-based dictation • Data transmission • Topic identification • Knowledge storing		
4	5	6	• STT App-based dictation • Data transmission • Topic identification	• Knowledge storing	
5	2	5	• STT App-based dictation • Data transmission • Topic identification	• Classifier training	• Business rule extraction
6	3	6	• STT App-based dictation • Topic identification • Knowledge storing		
7	5	6	• STT App-based dictation • Topic identification	• Knowledge storing	
8	3	6	• STT App-based dictation • Topic identification		
9	5	5	• STT App-based dictation	• Topic identification • Knowledge storing • Data transmission	• Business rule extraction • User confirmation
10	4	7	• STT App-based dictation • Topic identification • Data transmission		

process of knowledge acquisition has been automated through the prototype system. However, this fact cannot sufficiently explain how much a user satisfies and trusts the automated function of the prototype system. Therefore, a 10-member group of experts was organized and questioned.

At first, each member was asked to experience the prototype system by loudly reading a script selected from a newspaper article. Most of members repeated the experiment until they had examined the functions and identified the features. After finishing the experiment, at second, each of them was asked to rate the level of automation using the 7-point Likert scales, from ‘none-automated’ to ‘fully-automated’. Based on the marked point, finally, each member was asked to explain the reason why he or she determined the point by specifying features that

are fully-, semi-, and none-automated. <Table 1> shows the results from the focus group interview.

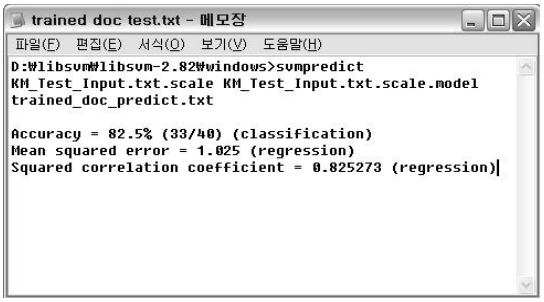
Most of members of the focus group evaluated the prototype system highly automated the process of knowledge acquisition, as the rating points in <Table 1> show. Based on the results in the Table, an interesting but very important implication can be derived : Key requirements to make the knowledge acquisition process be fully-automated can be identified. 'STT App-based dictation' and 'SVM-based topic identification' are the key features enabling the prototype system to be regarded as a fully-automated system. The feature of 'STT App-based dictation' can give an answer to the question of how knowledge in ordinary conversations and meetings can be acquired. Knowledge in conversations is too various to count the number, and indeed is street smart, which makes researchers and practitioners have long concentrated on it. Voice knowledge has been regarded as something that could not be articulated and acquired, however the functionality of STT App-based dictation provides a way convenient to manage voice knowledge. Also, the feature of 'SVM-based topic identification' makes the prototype system possible to acquire knowledge with understanding the subject included in the knowledge. Dictating knowledge without knowing its meaning cannot be regarded as 'acquisition of knowledge.' Dictating voice knowledge and identifying its topic must be paired together to attain the true acquisition of voice knowledge, which is one of primary contributions of this research.

4.2 Statistical Analysis to Measure the Accuracy of Acquisition

The accuracy of acquisition can be influenced by two components : STT application and SVM module. The accuracy of STT App-based dictation is very crucial, because it initiates the execution of knowledge acquisition process.

Unless the accuracy of dictation is guaranteed, the result of acquisition cannot be satisfactory. The accuracy of STT App-based dictation, however, depends on the speaking habit of speakers, not on the recognition ability of applications. Of course the function of recognition must be enhanced so that the application can recognize every pattern of speaking; however this work must be done by researchers in the area of voice recognition technologies. Considering the issue of recognition accuracy is out of this research's concern. Although the accuracy of STT App-based dictation was observed to be very high, even a little error could have bad effects on the final results. Therefore, in this research, assuming that the performance of STT application is satisfactory enough not to have any effects on the final results is more reasonable. On the contrary, the accuracy of the SVM module can and must be considered in this research. The accuracy of the SVMs has been verified to be very high. If the prediction model has been trained sufficiently, then the SVMs output very accurate and correct results. The accuracy can refer two aspects: one is for comparing manual classification and the other is for measuring the correctness of classification. The LibSVM de-

ployed in this research has been proved to outperform the manual classification (Hsu et al., 2001). Measuring the correctness of classification means how much the outputs are classified into the correct categories, and usually it can be measured in statistical considerations. To measure the correctness of classification, three to five documents per each category are collected, and totally 40 test documents are inputted to check the accuracy.



<Figure 11> Prediction Accuracy of the SVM Module

<Figure 11> shows the prediction accuracy which is automatically calculated by the LibSVM. The accuracy of classification is indicated to be 82.5% with 1.025 MSE. This means 33 documents out of 40 documents are correctly classified into proper categories. Comparing to the results of the previous studies (Hsu et al., 2001; Kang et al., 2008), this result shows that the SVM module of this research classifies accurately. Of course, the accuracy needs to be improved around 90% to gain the credibility of classification. Small number of sample documents for training is the main reason for the relatively low level of accuracy. If sufficient number of training documents is provided in training the model, more accurate result of classification can be expected.

5. Conclusions

The reason why knowledge cannot be effectively accumulated is not because we do not know what we know, but because we do not know how we articulate it. If there is way to document we know, then a tremendous amount of knowledge can be identified and accumulated in knowledge repositories. The limitations in the methodology of knowledge acquisition have restricted the practices of knowledge management. After something to be managed have been prepared, specific technologies to manage them can finally be valuable. This paper tries to provide an effective solution to this problem by truly automating the process of knowledge acquisition. The amended knowledge acquisition methodology, proposed by this research, enhances the conventional approaches to automatically acquire knowledge by applying the Smartphone operated on the cloud computing environment and the SVM-based topic identification. The context of knowledge use is also considered using the Smartphone as a sensor to identify context data. To promote the use of acquired knowledge, the context data which play the role of the meta-knowledge must be considered.

Among various types of knowledge, this research just focuses on the voice knowledge which verbally communicated in conversations. Knowledge in conversations is to be acquired

by identifying the topic of an electronic document containing the dictated texts which have been converted from the recorded conversations. Therefore, other types of knowledge, such as hard copy- and activity-based knowledge, can be also acquired by applying the same procedure if, and only if, the electronic document is provided. The optical character recognition technology can be applied to convert the hard copy document into the electronic document, and the sensor- based activity recognition technology can be applied to convert the monitored actions into the textual data. Since knowledge can make its appearance to the real world via convertsations, documents, and activities, observing such human outputs and applying the proposed methodology to them can deliver a full-scale management of knowledge.

The prototype system includes two main sub-modules that are synchronized: the STT module and the SVM module. Since the main functions of the prototype system are characterized by the main modules, the proposed prototype does not consider how the business rule can be automatically formulated. In this research, to deliver main ideas concerning capturing and saving knowledge according to corresponding topic and context data as the meta-knowledge, the business rule is supposed to be well-defined and provided. To fully automate the process of knowledge acquisition, the business rule must be also automatically formulated by associating required information according to the identified topic and captured context data. However defining the category of context data which associate the required information to form a rule requires not a little time and expense. One possible solution to this problem can be the deployment of ontology technology. Of course this research also partially applies ontology-based category to train the SVM module, however it does not cover every corner of situations users can confront. Defining context category, and therefore relating the context data with the candidate business rule based on ontology technology is expected to automate the process to formulate business rules.

A discussion can include several pieces of knowledge. Therefore, a sentence-by-sentence analysis is necessary to clearly differentiate one piece of knowledge from another. However, the prototype is designed to simply store the whole discussion in terms of extracted topics. Although a discussion can be considered as a piece of knowledge (Anerousis and Panagos, 2002), each piece must be distinctively identified for its realistic use. For this, sentences within a discussion must be segmented, and each sentence must be analyzed in terms of its accidence. In examining each part of a sentence, knowledge-resident parts must be analyzed by expert linguists. Although this paper also assumes that a discussion is a piece of knowledge, there should be a linguistic analysis of every sentence to ensure the reliability of acquired knowledge.

This paper contributes to the knowledge management literature by suggesting a new methodology to automatically acquire human knowl-

edge based on the capabilities of text mining and cloud computing technologies. This study is different from conventional studies in that it provides and implements an automated knowledge acquisition methodology. Organizations seeking to manage organizational knowledge for strategic purposes can apply the proposed approach to their business environment. As long as knowledge possessors are agreeable to sharing and monitoring their knowledge, the proposed methodology and the prototype system should facilitate organizations' knowledge acquisition efforts.

References

Anerousis, N. and Panagos, E., "Making Voice Knowledge Pervasive", *IEEE Pervasive Computing*, Vol.1, No.2(2002), 42~48.

Bae, K., "Self-Tour Service Technology based on a Smartphone", *Journal of Intelligence and Information Systems*, in Korean, Vol.16, No.4(2010), 147~157.

Basu, A., Watters, C., and Shepherd, M., "Support Vector Machines for Text Categorization", *Proceedings of the 36th Hawaii International Conference on System Sciences*, 2002.

Chang, C. C. and Lin, C. J., "LIBSVM : A library for support vector machines", available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001.

Dumais, S., Platt, J., Heckman, D., and Sahami, M., "Inductive learning algorithms and representations for text categorization", *Proceedings of the 7th International Conference on Information and Knowledge Management*, 1998.

Hsu, C. W., Chang, C. C., and Lin, C. J., "A Practical Guide to Support Vector Classification : LibSVM Tutorial", available at http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf., 2001.

Joachims, T., "Text categorization with support vector machines : Learning with many relevant features", *Proceedings of the European Conference on Machine Learning*, 1998.

Kang, H., Suh, E., and Yoo, K., "Packet-based context aware system to determine information system user's context", *Expert Systems, with Applications*, Vol.35, No.1-2(2008), 286~300.

Kim, S. and Ahn, H., "Development of an Intelligent Trading System Using Support Vector Machines and Genetic Algorithms", *Journal of Intelligence and Information Systems* (in Korean), Vol.16, No.1(2010), 71~92.

Kwon, O. and Lee, J., "Text categorization based on k-nn approach for web site classification", *Information processing and management*, Vol.39(2003), 25~44.

Liu, Q., Wang, G., and Wu, J., "Secure and privacy preserving keyword searching for cloud storage services", *Journal of Network and Computer Applications*, 2011, article in press.

Meyer, D., Leisch, F., and Hornik, K., "The support vector machine under test", *Neurocomputing*, Vol.55(2003), 169~186.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., and Euler, T., "YALE : Rapid Prototyping for Complex Data Mining Tasks", *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD-06)*, 2006.

Pallis, G., "Cloud computing: the new frontier of

Internet Computing", *IEEE Internet Computing*, Vol.14, No.5(2010), 70~73.

Pamies-Juarez, L., García-López, P., Sánchez-Artigas, M., and Herrera, B., "Towards the design of optimal data redundancy schemes for heterogeneous cloud storage infrastructures", *Computer Networks*, Vol.55(2011), 1100~1113.

Park, S., Lee, J., and Kang, J., "The Effect of Knowledge Acquisition through OntoRule : XRML Approach", *Journal of Intelligence and Information Systems*, in Korean, Vol.11 No.2(2005), 151~173.

Rennie, J. D. M. and Rifkin, R., "Improving multiclass text classification with the support vector machine", *CBCL Paper* #210/AI Memo #2001-026, Massachusetts Institute of Technology, Cambridge MA, October, 2001.

Schmidt, A., Beigl, M., and Gellersen, H., "There is more to context than location", *Computers and Graphics*, Vol.23(1998), 893~901, In Zhou, J., Gilman, E., Palola, J., Riekki, J., Ylianttila, M., and Sun, J., "Context-aware pervasive service composition and its implementation", *Personal and Ubiquitous Computing*, Vol.15(2011), 291~303.

Yoo, K., "SVM-based knowledge topic identification toward the autonomous knowledge acquisition", *Proceedings of the IEEE 9th International Symposium on Applied Machine Intelligence and Informatics (SAMI),* (2011), 149~154.

Zhou, J., Gilman, E., Palola, J., Riekki, J., Ylianttila, M., and Sun, J., "Context-aware pervasive service composition and its implementation", *Personal and Ubiquitous Computing*, Vol.15(2011), 291~303.

Abstract

모바일 기기와 가상 스토리지 기술을 적용한 자동적 및 편재적 음성형 지식 획득

유기동*

최근에 들어 많은 관심과 인기 속에 사용되고 있는 스마트폰은 클라우드 컴퓨팅의 편재적 기능성을 접목하여 즉각적인 지식의 획득에 효과적으로 활용될 수 있다. 또한 지식의 주제어 또는 명칭을 자동으로 파악하여 해당 지식을 저장할 수 있다면 전반적인 지식 획득 과정이 자동화될 수 있다. 본 논문은 텍스트마이닝 기반 주제어 추출 기술과 클라우드 스토리지 기반 스마트폰을 접목하여 지식이 발생되는 지점 및 시점에 즉각적으로 해당 지식을 획득할 수 있는 학제적 방안을 제시한다. 이를 위해 스마트폰은 지식이 포함된, 지식소유자의 대화를 녹음하는 역할을 함과 동시에 지식소유자의 대화의 내용을 부가적으로 특성화할 수 있는 상황정보를 채취할 수 있는 센서의 역할을 수행한다. 또한 기계학습 알고리듬 중 텍스트마이닝 분야에서 우수한 성능을 나타내는 것으로 알려진 Support Vector Machine 알고리듬을 사용하여 해당 대화의 주제어를 추출한다. 파악된 주제어와 상황정보를 연관시켜 일종의 비즈니스 규칙을 생성할 수 있으며, 최종적으로 규칙, 주제어, 상황정보, 그리고 문서화된 대화를 종합하여 하나의 지식을 자동으로 획득할 수 있다.

Keywords : 자동 지식 획득, 주제어 파악, 클라우드 스토리지, 스마트폰, 지식경영

* 단국대학교 경상대학 경영학부

저 자 소 개

유기동

포항공과대학교(POSTECH) 산업경영공학과를 졸업하고, 동대학원에서 경영정보 시스템 석사 및 박사 학위를 취득하였다. 현재 단국대학교 경상대학 경영학부 조교수로 재직 중이다. Expert Systems with Applications, Electronic Commerce Research and Applications, Journal of Knowledge Management, 경영정보학연구, 지식경영연구 등에 주로 논문을 발표하였다. 주요 관심분야는 지식경영 및 지식관리시스템, 유비쿼터스 컴퓨팅, 차세대형 경영정보시스템, 컨텍스트 기반 자율적 컴퓨팅, 지능적 지식서비스, 정보전략 기획 및 성과평가 등이다.