The Recommender System of Teaching Development Guidelines Using Text Mining Techniques

Wanvipa Wongvilaisakul¹
and Uraiwan Inyaem²

Department of Information Technology, Faculty of Engineering and Technology,¹
Panyapiwat Institute of Management, Nonthaburi, Thailand, 11120
Departments of Mathematics and Computer Science, Faculty of Science and Technology,²
Rajamangala University of Technology Thanyaburi, Pathumthani, Thailand, 12110
¹wanvipawon@pim.ac.th
²uraiwan.inyaem@rmutt.ac.th

Abstract - In the present days, it is hard to avoid an overcrowded classroom condition; moreover, at the same time, we all want the most effective and efficient learning styles. The question is to find the way to design an effective learning style for classroom with large numbers of students. This is to fully correspond the need of the students. There are also many limitations; for example, number of students, time to make the analysis, and the development of the learning style to suit the students without any bias towards it. Therefore, in order to reduce the limitations, this article will present a new system to guide the development of the learning style by the use of text mining techniques. These techniques will display the concept to organize feedback from the students after each learning session. It will also help with developing the learning style to make it suitable for the students. The researchers choose to collect information of the students via e-learning methods. This helps collecting information of the students at all time. The text mining techniques is also being used as an analysis process. This will represent the analysis of the correlation between new information presented. The dataset used in this research is undergraduate students at Panyapiwat Institute of Management during the academic year of 2012-2013. The total number of dataset in this research is 7,813 opinion messages studying in Management Information System subject. The research found out that the top 5 activities of the students are as follows; 1) interesting contents lecture, 2) presentation with video, 3) providing examples/exercises, 4) quiz after each class, and 5) game with reward to help with revision. They also help to suggest activities use to further improve the learning styles for more effective in the near future.

Keywords - Recommender System, Text Mining, Feedback, E-Learning

I. INTRODUCTION

The learning style in the present days focuses on a principle of student-centered learning. Therefore, teachers need to choose variation of teaching methodology as well as giving an opportunity for students to participate in the classroom. This means that students will be able to fully develop their own abilities. Teachers will be able to evaluate each learning activity and realize if this has attain the goal previously set. This will be used as an approach to solve any problem which may occur during the activities. They also will be able to control the outcome of the activities and used it in the preferable way. This will be based on the feedback from the students. From the research on “The Development of Instructional Activities using Learning Feedback for Participation of Learners: In
Case Studies of Information Technology Organization Behavior and Data Structure and Algorithm.”, the objective is to find an approach to the development of learning styles activities based on the feedback of the students by using dialogue journal which works as students’ short journal. The students will be able to continuously write and respond to the journal as if it was a dialogue between teachers and students. Moreover, each dialogue will be based on the students; whether it’s the suggestion or attitudes related to the learning. This will be used as a tool for this research to help the students record their own opinions and feedback from each classroom. The teachers will be able to interact with the students about the topic they have previously studied in the class. It will be used to analyze and further developed the learning activities based on opinions from students and teachers. According to the components of how to develop learning and teaching activities by using feedback from students, each component is concerned about 1) students’ needs, 2) students’ curiosity, 3) students’ feedback, 4) students’ schemata, 5) students’ objective, 6) learning environment, 7) system of learning, 8) students’ favorite topics, and 9) students’ learning problems [1]. As the research mentioned, it found out that there is also another limitation regarding to the tool for collecting information. This is because this type of dialogue journal is suitable for classroom with small number of students. This is because it takes a period of time in order to read and analyze the students’ opinion messages. Additionally, the teachers need to use the information and previous feedback to analyze. Therefore, it will not be able to apply to the class with large number of students. In conclusion, to reduce the limitations as resulted in this research, teachers have to collect information based on the students’ feedback. This is done in the classroom to further guide and reaches out the ways to develop learning styles in order to suit both students and teachers. The rest of this paper is organized as follows. Section 2 described related research of the recommender system of teaching development guideline. Section 3 presented components of the proposed system in detail. Section 4 discussed the result of proposed system. The final section concluded the paper.

II. RELATED WORKS
The Recommender System is a system that helps screen any information that is supposedly matching the need from most of the users. It also advises information to the users. The way that the recommender system works is that it provides information in advance. It is done by presenting the information needed for the users automatically. In general, the development of the recommender system uses 3 techniques to screen the information which are: Content-based filtering, Collaborative filtering, and Hybrid filtering [2]. The recommender system has been used in various types of teaching that is related to the needs and proficiency of student that may be difficult and complex. Therefore, it has been applied to use via data mining techniques. This is a way to search for the hidden format in the large quantity of data. This is to create new knowledge that will benefit through registration of the data and past learning results. This will be analyzing through the data form and rule of relation. Therefore, it will support new student in choosing to register in each subject in other semesters during the entire learning experience [3]. Text Mining is one of the Data Mining forms. It is also a part of the component in the Natural Language Processing. The study of Information Retrieval, Machine Learning, Statistic, and Language Evaluation focuses on Text Categorization, Text Clustering, Extraction, Sentiment Analysis and Document Summarization as well as the finding of entity relation modeling. Moreover, the Language Evaluation is considering one branch of Artificial Intelligence: AI which focuses mainly on the problems and understanding of human languages. The main components are the knowledge of language, grammar, and Statistic components [4]. Hany Mahgoub has explained about text mining by using the techniques of Association rules as follow;
A. Text Preprocessing Phase

1) Transformation: it is the way to change the format of the document with different format from the standard format; for example, XML format. This is so the file document will be in the same format.

2) Filtration: it is also a text extraction. This is done by choosing only the part that is needed; by word, syllable, clause or sentence to represent the document. It will also help removing the Stopword by removing an unimportant word for the document. This will not change the understanding of the entire document; for example, the clause or pronoun.

3) Stemming: it is the grouping of words with the similar meaning in order to be in the same group

4) Indexing: it is a way to use any document with the standard format; such as, XML that has been carefully chosen to create volume index. This can be created by both the users or automatically by the system.

B. The Use of Algorithm

For information finding that need an indexing from the entire document with different techniques from pictures is part of an Association Rule Mining phase. This is a method to find the correlation of the text to correspond with the lowest secure feeling by the users.

C. Visualization Phase

Visualization Phase is the creation of computer graphic picture that could present

III. METHODOLOGY
OF THE PROPOSED SYSTEM

The proposed framework consists of 5 parts as shown in Fig 1. The process can be explained in detail as follow: Firstly, students studied each chapter content and activities in e-Learning system. Secondly, students finished each chapter contents so they assent each chapter and eventually, all opinion messages were sent to online storage in e-Learning. Thirdly, students’ opinion messages were sent to teachers and text analysis. The natural opinion messages can be the information needed precisely and accurately instead of using text. The users will be able to interpret the information more effectively [5]. The research is from A Recommendation System to Choose Study Program based on Repertory Grid. The main objective of this research was to propose an algorithm to develop a recommendation system to choose study program based on repertory grid technique. This is because it can do a better reasoning advice than the system that operates without knowing the number of the previous users in the system. It also can work fast because it does not need to calculate the similarity of the main users along with different other users [6]. The text mining technique has been used to develop different type of researches; for example, using text mining technique for agricultural research. This benefits the use of references and many other research of agricultural information in Thailand. It can also apply with the research by the experts regarding agriculture in the near future [7]. Moreover, there is still another research that uses text mining to analyze the news from the 3 Southern provinces in Thailand. This will help predicting the behaviors of the terrorists and also to support the decision making of natural disaster protection that may occur in the near future. There will be a collection of related news from different sources and store it within the database. After this has been done, the model will be built to find the correlation and the format for structure correlation of the events will be created [8]. not indicate any meaning so we proposed a new methodology for text analysis for the recommender system of teaching development guideline by using text mining techniques. In addition, the result from previous process is teaching development guideline as suggestion activities. Finally, teachers used suggestion activities to develop teaching by themselves for the future usage. From above described framework, in this paper, we focus in third process. We started with preparation data for analysis process. The steps were described in detail. Dataset in the research used 7,813 opinion messages of bachelor degree students at Panyapiwat Institute of Management who registered in subject of Management Information System between academic years of 2012-2013. The data was fetched via e-Learning and it was
shown in text file form. Dataset was stored in local storage that was taken preparation data step by step as follow. First step is removal some stop words and edit misspelling or over-writing. Second step is to remove HTML tags from opinion messages. After that is tokenization for opinion message lexeme tokenized into term. Final step is filler tokens by length. The data was made after preparing process and then the next paragraph we described methodology of practical data analysis which is to prepare dataset was processed for data analysis.

**A. Feature Extraction**

For this research, we will use statistic for choosing feature selection. We selected the related properties students such as student id, student name, student surname, and gender as well as opinion messages.

**B. Text Data will be Analyzed into Structured Data Form.**

Bag of word process is selection of related recommender activities as keywords such as need, do, understand, and confuse. The process is finding frequency of word in sentence.

**C. TF**

TF comes from Term Frequency which is the amount of word in each opinion messages of students or all opinion messages of students. TF-IDF means the result of multiplicity between the amount of occurrence word and the amount of specific occurrence word in class. The result is shown that the class of each record of opinion sentence.

**D. Finding of Association Rule**

Finding of Association Rule is used between the students’ opinion message and the recommender activities and relationship between the recommender activities and view of plan learning. In this case, we used FP-Growth algorithm for creating association rule. There are two steps of FP-Growth algorithm process as follows; first is to read data in database and create FP-tree and last process is calculation support value from FP-tree.

**E. K-Means Algorithm**

K-Means Algorithm [9, 10] is a classification of partitioned clustering and popular clustering. K is amount of group or cluster that is diving data depending on k variable. The k-means concept is steps as follow; first is to define the centroid of each clustering. Second is to find the distance between each data and mean of each cluster, Euclidean distance formula is used to calculate in this method. Third is to define the nearest cluster for each data. Forth is to calculate the centroid of each cluster. Fifth is to repeat the ordering process until data staying in stable cluster. In the paper, classifier used k-means algorithm to classify the word from opinion message extraction process into four classes such as positive polar, request polar, general polar, and negative polar.

**Fig. 1. Proposed System Framework**

**TABLE I**

**DISPLAYS THE ASSOCIATION RULE ACCORDING TO STUDENTS’ OPINION**

<table>
<thead>
<tr>
<th>Message</th>
<th>Keyword</th>
<th>Activity</th>
<th>View</th>
</tr>
</thead>
<tbody>
<tr>
<td>I want variety instructional media, such as, video clips.</td>
<td>want (want)</td>
<td>open the video clips</td>
<td>learners' needs</td>
</tr>
<tr>
<td>I really like the games which compete within the classroom. It is very enjoyable.</td>
<td>like (enjoy)</td>
<td>game with reward</td>
<td>learners' favorite topics</td>
</tr>
<tr>
<td>I am still confused with the contents of this chapter. There should be revision before starting new chapter.</td>
<td>confused (confused)</td>
<td>contents revision (contents)</td>
<td>learners' learning problems</td>
</tr>
</tbody>
</table>
IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed system prevented the students study with content and activity in e-Learning system and they may send feedback into e-Learning. Students’ opinion messages were analyzed with text analytical system. The research showed the suitable activities recommender for students which come from the online students’ opinion message. The recommender activity is recognized with the bag of words as a keyword. The keyword collection considered the frequency word in database. The results of this system are suggestion activity that can help recommender activity use to improve in future class. The summary percentage of views between academic year 2012-2013 as shown in Fig 2. such as 1) students’ needs, 2) students’ curiosity, 3) students’ feedback, 4) students’ schemata, 5) students’ objective, 6) surrounding of learning, 7) system of learning, 8) students’ favorite topic, and 9) students’ learning problems. Nine views indicated students’ different academic year and teachers with the same activity alike. The developed system classified keyword in each opinion message with classification algorithms. K-means method is used with the classification dataset. Each view prohibited suggestion activity for students while recommender message showed students’ name, surname, teacher name, and keyword. Each suggestion activity was promoted the percentage number of students while students may know their own suitable activities. The results of top 5 suggestion activities of the students are as following; 1) interesting contents lecture, 2) presentation with video, 3) providing examples/exercises, 4) quiz after each class, and 5) game with reward to help with revision. In this research, the students’ opinion messages were considered the polar opinion which is divided into four types; such as, positive polar, request polar, general polar, and negative polar. The beginnings of positive polar words (+) were used to positive thinking for teaching system or teacher. Examples are like (เข้าใจ) and understand (เข้าใจ). Secondly, the negative polar words (-) were used to invert the opinion message such as unlikely (ไม่ชอบ) and confuse (ไม่ผ่อน). Next, general polar words (0) were used to a non-specific domain. Examples are do (ทำ) and tell (บอก). Last, request polar words (*) were used for a want or request from teacher. Examples are want (ต้องการ) and need (ต้องการ). According to the students’ opinion Message, the researcher has used to find the result of the association rule. This is to present the learning activities which is correspond to all aspects of students’ needs. This can be seen in the example in table I.

V. CONCLUSION

This paper presents a new system to guide the development of the learning style by the use of text mining techniques. This will display the concept to organize feedback from the students after each learning sessions. It will also help with developing the learning style to make it suitable for the students. The dataset used in the research is bachelor degree students at Panyapiwat Institute of Management in the academic year of 2012-2013. The total number of dataset in this research is 7,813 opinion messages studying in Management Information System subject. There are nine views using in the research such as 1) students’ needs, 2) students’ curiosity, 3) students’ feedback, 4) students’ schemata, 5) students’ objective, 6) learning environments, 7) system of learning, 8) students’ favorite topics, and 9) students’ learning problems. The research found out that the top 5 activities of the students are as follow;
1) Interesting contents lecture, 2) presentation with video, 3) providing examples/exercises, 4) quiz after each class, and 5) game with reward to help with revision. The student’s opinion messages were classified with k-means algorithm as the polar opinion which is divided into four types; positive polar, request polar, general polar, and negative polar. They also help to suggest activities use to further improve the learning styles to be better effective in the near future.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)


International Journal of the Computer, the Internet and Management Vol.22 No.3 (September-December, 2014) pp. 65-70