Mining Educational Data to Analyze the Student Motivation Behavior

Kunyanuth Kularbphettong, Cholticha Tongsiri

Abstract—The purpose of this research aims to discover the knowledge for analysis student motivation behavior on e-Learning based on Data Mining Techniques, in case of the Information Technology for Communication and Learning Course at Suan Sunandha Rajabhat University. The data mining techniques was applied in this research including association rules, classification techniques. The results showed that using data mining technique can indicate the important variables that influence the student motivation behavior on e-Learning.

Keywords—association rule mining, classification techniques, e-Learning, Moodle log Motivation Behavior

I. INTRODUCTION

WITH the rapidly growth of the Internet, the web-based educational system has been increasingly used as an important tool to support learners and teachers. There are many benefits for information sharing and collaboration between learners and teachers in a course. Learners can take a web-based class to enhance their knowledge at any time and any place and teachers can easily create their online classes and monitor student’s performance as well. Moodle is a popular Learning Management System (LMS) that supports educators to create the effective online courses. However, it did not provide the function to access and estimate learner’s motivation behavior. Also Student’s logs in a Moodle can show students’ interactions like reading, writing, taking exam, and doing various tasks [1]. With a huge of accumulated information daily from e-Learning course, it is very difficult to analyze this data manually. Although there are some tools that help to report useful information which in turn is very valuable for analyzing student’s pattern behavior [2], they do not offer specific features teacher need to track and evaluate all the students’ activities in class [3].

Data Mining Techniques is the promising methodology to extract valuable information in this objective. Data Mining can analyze relevant information results and produce different perspectives to understand more about the students’ activities so as to customize the course for student learning. For this paper, we used the data set of the Information Technology for Communication and Learning Course at Suan Sunandha Rajabhat University.

The data is composed of demographic information of each student, information in this class and grade student obtained. After data preparation process, data mining techniques are applied to discover the valuable knowledge for analysis student motivation behavior. And the rest of this paper is organized as follows. Section 2 reviews about the related literatures and the related methodologies used in this work. Section 3 presents the implementation based on the purposed data mining techniques. In section 4 the result and discussion is presented. Finally, we conclude the paper with future research issues in section 5.

II. RELATIVES WORKS AND THE METHODOLOGIES

In this section, we illustrate the literature search and the specified methodologies used in this project.

A. Relates Works

A literature search shows that most of the related researches have deployed data mining techniques to analyze student’s learning behaviors by following this: According to C. Romero at el [4], the research was shown the usefulness of the data mining techniques in course management system and the rules can help to classify students and to detect the sources of any incongruous values received from student activities. K. Waiyamai applied data mining approach to develop new curricula and to assist engineering students for choosing an appropriate major [5] and B.Minaei-Bidgoli et al [6] presented an approach to classify students in order to predict their final grade based on features extracted from logged data in an education web-based system Data mining techniques like association rule mining were applied in [7],[8] to extract the patterns and to evaluate the activities of on line course and classification and association rule mining algorithms are discussed and demonstrated in [9]. Also there are many researches that have been investigated in the on-line learning environment. For example, West et al investigated impact of learning style on e-learning by using Statistics [10] and Kerdprasop et al used Rule induction rough set to Classify student knowledge level [11].

B. The Methodologies

Data Mining is the data analyzing approach from different perspectives to summarize the useful information results. The data mining process uses many principles as machine learning, statistics and visualization techniques to discover and present knowledge in an easily comprehensible form. There is another definition as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data” [12], [13].
Association rule mining is used to discover interesting relationships between variables in databases. According to Agrawal et al [14], an association rule explains a close correlation between items in a database in the form of \( x \Rightarrow y \) where \( x \) and \( y \) are sets of Item set \( I = I_1, I_2, \ldots, I_m \) is a Item set of \( m \) distinct attributes. The rule is indicated \( x \) implies \( y \) whereby \( x \) is called antecedent and \( y \) is called consequent. There are two importance thresholds for measurement association rule mining, minimum support and minimum confidence. The support of a rule \( x \Rightarrow y \) is the probability of the Item set \( \{ x, y \} \) that means the relevance of the rule and the confidence of a rule \( x \Rightarrow y \) is the conditional probability of \( y \) given \( x \) that indicate the accuracy of the rule.

\[
\text{Confidence (} x \Rightarrow y \text{)} = \frac{\text{support (} x \cup y \text{)}}{\text{support (} x \text{)}} \quad (1)
\]

\[
\text{Support (} x \Rightarrow y \text{)} = \frac{\text{support (} x \cup y \text{)}}{\text{Total number of transaction in D}} \quad (2)
\]

Set of transactions: \( D = \{ d_1, d_2, \ldots, d_n \} \) each \( d_i \subseteq I \)

Confidence is a significant measure of the association rules to indicate how to strength of the mined rules. If the confidence of the association rule \( x \Rightarrow y \) is 80%, it means that 80% of the transactions that contain \( x \) also contain \( y \), based on users to indicate the specified minimum confidence [15].

Apriori Algorithm is an influential algorithm for association rule mining, purposed by proposed by Agrawal [16]. The Apriori Algorithm is used level-wise search for frequent item sets, the sets of items that have minimum support.

A decision tree is one of the most well known classification algorithms that are commonly used to examine data and induce the tree in order to make predictions [17]. The purpose of the decision tree is to classify the data into distinct groups or branches that generate the strongest separation in the values of the dependent variable [18]. Mitra S and Acharya T [19] classified the important decision tree methods from the machine learning community: IDS, C4.5, CART; and large databases: SLIQ, SPRINT, SONAR, RainForest.

The ID3 algorithm, introduced in 1986 by Quinlan Ross [20], uses the top-down induction of decision trees. ID3 uses information gain to build the leaves of the tree. A greedy search is used and the best attribute is selected by it and never reconsiders earlier choices by looking back [21].

J48 is an open source Java implementation of the C4.5 algorithm under WEKA data mining platform. C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is a software extension and thus improvement of the basic ID3 algorithm designed by Quinlan [22]. J48 uses gain ratio to classify the decision tree.

III. EXPERIMENTAL SETUP

The data set used in this research is collected the student’s data from the Information Technology for Communication and Learning Course at Suan Sunandha Rajabhat University, in the period of 2010-2011.

As shown in Fig.1, the student data set was 3,140 records and it is composed of personal records, course (face-to-face) records and students’ log file from e-Learning system.

Moodle has been used for this course. Moodle is an well known open source software for learning management system that offers teacher create and manage online classes effectively. From this Course, in the class (face-to –face) room, student must be required to attend in course room, to do exercises and project, and to take exams in class. Also, in e-Learning class, student needs to take pre and post quizzes online, to review and use materials on e-Learning system and to participate in exercises.

Also, the results of students’ grade are collected in the last. From Fig.2, the data is preprocessed, and transformed to be appropriated format in order to apply data mining techniques to discover valuable information. We removed irrelevant attributes and reduced data sets contained with missing value. And all continuous attributes have been transformed to nominal attributes. There are various methodologies to transform numerical attributes to discrete attributes like equal width, equal frequency, clustering principles and etc. In this research, the equal width method was used to partition the value of continuous attributes into five nominal values: VERY LOW, LOW, MEDIUM, HIGH and VERY HIGH.

The table I shows the significant information of this research after preparation data. Data was analyzed by WEKA. WEKA, the Waikato Environment for Knowledge Analysis, is a collection of machine learning algorithm to analyze data set for data mining tasks [23].
TABLE I
IMPORTANT ATTRIBUTES

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PreTest_number</td>
<td>Identification number of pretest</td>
</tr>
<tr>
<td>PostTest_number</td>
<td>Identification number of posttest</td>
</tr>
<tr>
<td>Result-PreTest_number</td>
<td>Mark obtained from pretest</td>
</tr>
<tr>
<td>Result-PostTest_number</td>
<td>Mark obtained from posttest</td>
</tr>
<tr>
<td>Time-of-PreTest_number</td>
<td>Time spent on pretest</td>
</tr>
<tr>
<td>Time-of-PostTest_number</td>
<td>Time spent on posttest</td>
</tr>
<tr>
<td>SumofPreTest</td>
<td>Total mark obtained from all pretest</td>
</tr>
<tr>
<td>SumofPostTest</td>
<td>Total mark obtained from all posttest</td>
</tr>
<tr>
<td>Time-of-viewing-material</td>
<td>Total time spent on viewing material</td>
</tr>
<tr>
<td>number-of-Attendance</td>
<td>Number of Attendance in class</td>
</tr>
<tr>
<td>Assignment_Number</td>
<td>Identification number of assignment</td>
</tr>
<tr>
<td>Result_Assignment_Number</td>
<td>Mark obtained from assignment</td>
</tr>
<tr>
<td>Result_InClassTest_number</td>
<td>Mark obtained from in class test</td>
</tr>
<tr>
<td>Project_score</td>
<td>Mark obtained from project</td>
</tr>
<tr>
<td>Midterm-score</td>
<td>Mark obtained from Midterm</td>
</tr>
<tr>
<td>Final-score</td>
<td>Mark obtained from Final</td>
</tr>
<tr>
<td>Grade</td>
<td>Final mark obtain from this class</td>
</tr>
</tbody>
</table>

IV. RESULTS OF EXPERIMENTAL

A. Apriori Algorithm

Apriori algorithm has been used for this research and evaluated with a minimum support of 0.2 and a minimum confidence of 0.9 to presents significant relationships between the activities of students on this class and their final scores. Fig 3 shows examples from the results of Apriori algorithm.

From the results of Apriori approach, there are many unnecessary rules teachers are not interested in. Therefore, for this research, the selected rules are explained in the Fig.4, based on decision of teacher.

B. Decision Tree Algorithms

ID3 and J48 are utilized in this research and Fig 6(a) and (b) present the results of these algorithms.

“Grade”, is significantly attributed to student motivation behavior. Accordingly, it was being discretized in to five levels: VERY LOW, LOW, MEDIUM, HIGH and VERY HIGH. Based on the derived testing result, 30% of the data set used for training, the result is depicted in Figure 5.
V. CONCLUSION AND FUTURE WORK

In this paper, we show the data mining techniques can be useful to model the student behavior outcomes and the preliminary result presents a promising progress in this prototypes model for the ongoing improvement of e-Learning course.

This model can be beneficial to similar courses to share and discover students’ motivation behavior. However, in term of the future experiments, we are looking forward to research about other data mining techniques to enhance this project and also apply the tool to help teachers in their class.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial subsidy provided by Suan Sunandha Rajabhat University.

REFERENCES


Kunyanuth Kularbphettong received the B.S. degree in Computer Business, M.S. degree in Computer Science, and Ph.D degree in Information Technology. Her current research interests are in Multi-agent System, Web Services, Semantic Web Services, Ontology and Data mining techniques.

Cholticha Tongsiri is interested in Web-Based Educational Learning, and Data mining techniques. She is studying MS. In Management Information System.