

Improving Teaching quality of computer courses in Sri Lankan Universities using Data Mining Techniques

Ahamed Lebbbe Hanees

Department of Mathematical Sciences,
Faculty of Applied Sciences, South Eastern University of Sri Lanka,
Sammanthurai-320350, Sri Lanka
alhanees@seu.ac.lk

Abstract

With the rapid development of information technology, computer technology has been getting more widely used in daily life, thus, it is necessary for each university graduates grasp basic technical skills of computers. However, In Sri Lanka a teacher will be responsible for teaching many university students basic computer courses, so it is difficult to ensure the quality of the teaching with the enlarged students' enrollment in Sri Lankan Universities.

In this research we proposed combining data mining and online examination system to improve teaching of basic computer courses through Student Academic Performance Monitoring and Evaluation. The data mining methodology while extracting useful, valid patterns from higher education database environment contribute to proactively ensuring students maximize their academic output. This work develops a methodology by the derivation of performance prediction indicators to deploying a simple student performance assessment and monitoring system within a teaching and learning environment by mainly focusing on performance monitoring of students' continuous assessment (tests) and examination scores in order to predict their final achievement status upon graduation. Based on various data mining techniques (DMT) and the application of machine learning processes, rules are derived that enable the classification of students in their predicted classes.

KeyWords: Data Mining, Machine Learning, Association, Classification

Methodology

COMBINATION OF DATA MINING AND ON-LINE EXAM SYSTEM

We proposed details idea of how to combine data mining and on-line exam system to improve teaching effect.

Data Source

There are mainly two types of data, routine data and exam data. Routine data is generated from each class, including the contents of each class, the approaches adopted by teachers, the knowledge master situation. While the exam data is obtained after the examination, including the mark of each classroom occupied in test, the exam mark, usual mark and the final mark. Through test system, these data can be gathered together. Forexample, in my university, all the students of non computer major are required to do their course work and computer practice by examination system. Before using the test system, the teacher will input the title of each knowledge point entered into the exam system. When the test paper is needed, the random buildup of a test paper can be achieved by means of automatic buildup, handmade buildup, and handtyping, thus making the exam system more convenient. As the students finish their course work or exam, test system can generate detailed students' scores of each item of test paper. It is noteworthy that data format of data generated directly by the examination system is not supported by Weka, so teachers need to convert these data into the format which Weka can deal with.

Cluster Analysis and Outlier Analysis

Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense. A cluster of data objects can be treated collectively as one group. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, especially in data mining. Weka provides strong clustering and preprocessesfunction for data analysis,

There is no doubt that students of different individuals vary in their background, such as the individual's learning ability, interests and learning habits, the original foundation, effort, there is a huge difference. Therefore it needs to distinguish between student groups in order to make a personalized learning arrangement. Clustering method can be used for finding a similar students group, to facilitate the management and teaching. For each student has a lot of attributes, some of it apparently is not the learning ability of individuals affected, such as student ID, e-mail. For sake of conducting basic computer course teaching, first of all cluster analysis methods can be used to assign students into different cluster through their scores, and then the results of classification association rules for students to conduct the forecast.

If it exists $X, Y \Rightarrow Z$ rules. Well, according to student behavior X behavior and Y can be introduced Z -acts may occur. It can help us promptly

Develop strategies to encourage the production or containment of Z behavior. Moreover, for example, because there is limited teacher resource, when we concentrate on making individual instruction on student of poor foundation, we should exactly distinguish these students. However, there maybe some students whose original foundation is good, but they don't work hard, which

lead them to get low scores. Through outlier detection, we may identify these students, thus we can pay more time to help students of poor foundation.

Association Rule production

Frequent patterns are patterns (such as itemsets, subsequences, or substructures) that appear in a data set frequently. For example, a set of items, such as milk and bread, that appear frequently together in a transaction data set is a frequent itemset. It is very useful that association rules are employed for teaching evaluation. After mining association rule of different knowledge points the produced association rules can be used to guide the teaching. Application of association rules analysis papers, according to the scores of students analyze the degree of difficulty of each question to distinguish between degrees, correlation and other indicators.

Accordingly, the quality of examination questions can be made a comprehensive and accurate assessment of the topic and provide guidance for teachers. And thus it can be used to examine teaching situation and students to grasp the situation, to provide reference for future teaching. Furthermore, using association rule, we may build up a set of precaution mechanism. Through test system, the classification, association rules analysis, cluster analysis of historical data, we are able to find the potential relation between the failed courses and the key factors that caused the failure, thus gain the characteristics and commonness of the students needed to be warned. The analysis results, providing the basis for reducing the rate of new failure, increasing the pass rate, or even predicting the possibility of appearing similar precaution students, are helpful to the management of the students for universities.

Proposed DM techniques

The rule inductions and artificial neural network data mining techniques are proposed to use in the project fall under the category of machine learning that uses high end modeling techniques for uncovering hidden patterns and/or predicting outcomes. Supervised knowledge discovery explains relationships found. Hence for this project the data mining product will use software that parades in its arsenals C5.0, C&RT, ANN, CHAID, QUEST, Link Analysis, KMeans, and Kohonen etc.

For both classification and estimation in supervised and unsupervised tasks respectively. The supervised modeling techniques for the classification and estimation of students' data and DM theories and techniques of analyses to be fully applied.

The C5.0 algorithm served as a good representative of the decision trees in this project as its statistical property of information gain helped to determine which of the several attributes best represent the division of the training sets.

Conclusions

This research mainly discusses how to employ data mining to improve teaching of basic computer courses in Sri Lankan Universities, and discuss related key technologies. Especially, proposed to use an open resource data mining tools to mine useful information from daily exam to precaution against students who may fail in final examination or who cannot meet requirement of

basic computer courses. However, the proposal is still not so fulfill, which needs further improvements, that is this proposal only introduce using data mining tool to mine results from on-line exam system, but the main drawback of this method that the steps of data mining and data collection are separated. Data collecting work should be done alone and only after data collecting is finished data mining can be executed, and the data gathered from test system cannot be used immediately, which should be converted into the data format that Weka support, which may be a time consuming work, hence we have to develop an on-line examination which has function of data mining.

References

(1).Dong Caiyun, QuShouning, “Data mining and its application in university education system”, Journal of Jinan University (Sci. & Tech.), 2004, vol.18, no.1, pp.65-68

[2] Wang Mulin, “Data mining association rules on the management of technology in teaching”, Computer Knowledge and Technology.(Academic Exchange), 2007,.no.8, pp.525-527,

[3] Zhangming, Data Mining for Prewarning Students with Score Data.
[Master thesis] Wuhan: Wuhan University of Science and Technology, 2008 (in Chinese)

[4] Han Yaming, Xiao Yuwei, Zhang zhijun, “Application of association rules mining to arrangement of courses in universities”, Microcomputer Applications, 2007.vol.28, no.1, pp.54-57

[5] Ma xirong sun huazhi application data mining in teaching evaluation, computer engineering and application, 2003, No.19, pp.51-54

[6] Hong yuehua, Data Mining for Classroom Teaching Quality Based on Fuzzy Comprehensive Evaluation COMPUTER SCIENCE, 2008, No.2, 42-45

[7] WANG Tie-jun. The Application of Data Mining Technology in the Teaching Evaluation, modern computer, 2005, No.3, pp, 34