Improving Students' Performance using Educational Data Mining

Falguni Ranadive[1]
Principal, Parul Institute of Computer Application, Waghodia

Dhaval Mehta[2]
Asst. Professor, Navrachana University, Vadodara

Abstract

Due to successful applications in the business domain, more and more researchers now interested on applying data mining techniques in the education sector. In this paper, we discuss how data mining can help to improve an education system by enabling better understanding of the students. The extra information can help the teachers to manage their classes better and to provide proactive feedback to the students. This paper proposes the methodology to predict students' performance based on their psychological characteristics, internal and final examination results. The educational systems currently face number of issues such as high dropout rates, identifying students in need, personalization of training and predicting the quality of student interactions. Data mining provides a set of techniques, which can help the educational system to overcome these issues.

Keywords: Educational Data Mining, Predicting student performance, Classification, Prediction

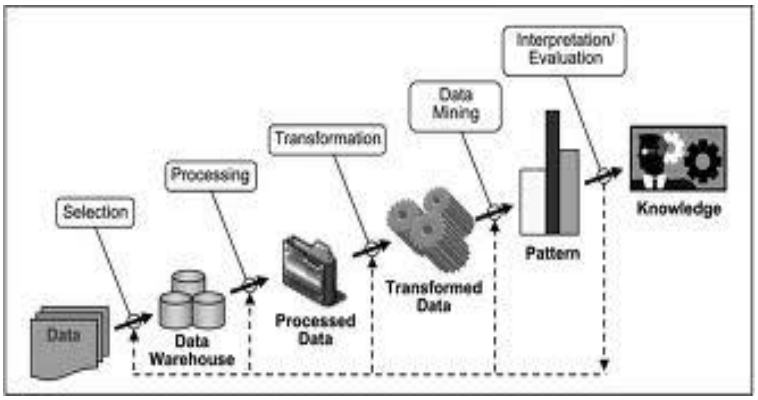
Introduction

Educational data mining (EDM) is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context. EDM uses computational approaches to analyze educational data in order to study educational questions. Educational opportunities, competitive and dynamic environment demands for quality education in respective domains. The education globalization leads to more and better opportunities for students to receive high quality education at institutions all over the world. The educational institutions are confronted with a severe competition among each other, try to make students capable enough to successfully complete the educational process and the retention ratio is reduced. Although educational data mining has not achieved greater success as compared to other domains such as e-commerce, some institutes are able to make use of data mining to discover knowledge and improve the learning experience of students as well as increase their profits. With the rising of learning management systems, one can expect if the data mining could integrate with learning management systems, then it would become possible to discover real-time knowledge and provide personalized learning experience to the students.

Data Mining Techniques

Data mining, also popularly known as Knowledge Discovery in Database, refers to extracting or "mining" knowledge from large amounts of data. Data mining techniques are used to

operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. While data mining and knowledge discovery in database are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. The sequences of steps identified in extracting knowledge from data are shown in Figure 1.



[Fig. 1 Knowledge Extraction from Raw Data]

Various algorithms and techniques such as Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases.

1. Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

2. Clustering

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification.

3. Predication

Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict. Unfortunately, many real-world problems are not simply prediction. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets)

may be necessary to forecast future values. The same model types can often be used for both regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models.

The Big Five personality test

In psychology, the Big Five personality traits are five broad domains or dimensions of personality that are used to describe human personality. The theory based on the Big Five factors is called the Five Factor Model (FFM). The Big Five factors are:

- Openness
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism

The big five personality traits are the best accepted and most commonly used scientific measure of personality and have been extensively researched.

Decision tree induction

Decision tree learning, used in data mining, uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. More descriptive names for such tree models are classification trees or regression trees. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

In data mining, a decision tree describes data but not decisions; rather the resulting classification tree can be an input for decision making.

K-means Clustering

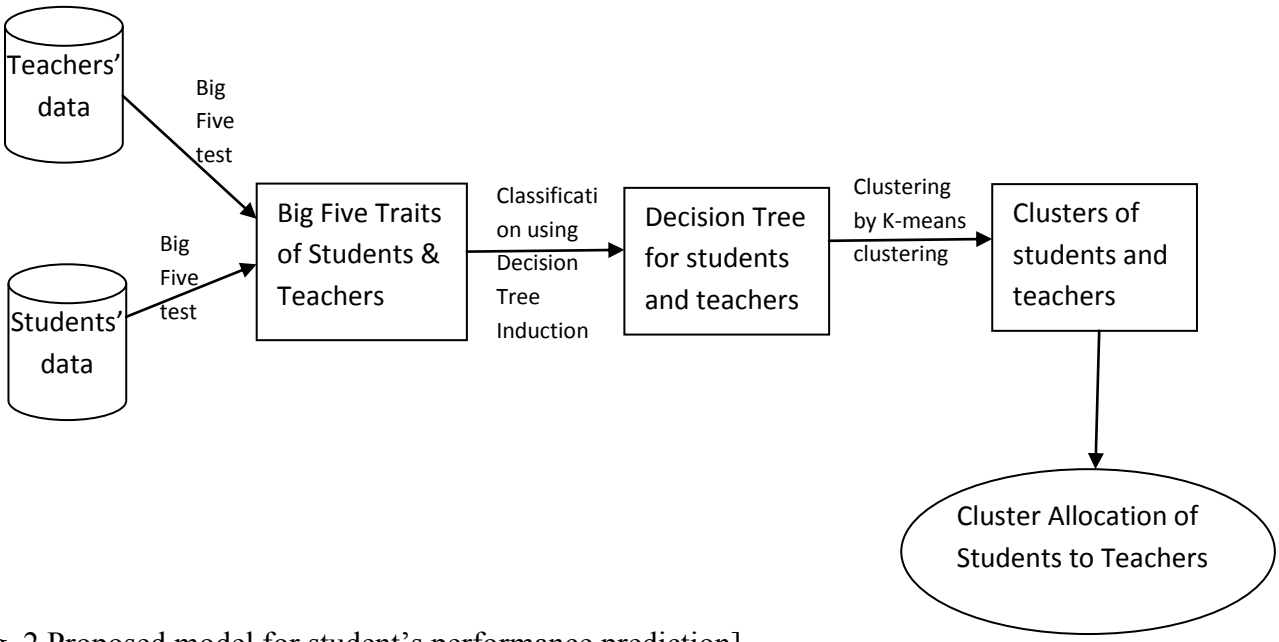
In data mining, k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. This results in a partitioning of the data space into groups of similar data.

Methodology

Data Mining has been used to study the factors leading students to choose to engage in behaviors which reduce their learning and to understand the factors influencing university student retention. Student profiling can be done by any one of the psychological methods such as Big 5 methods or Psychometric profiling. These methods can identify personality traits of each student.

A similar example of the social application of data mining is its use in expertise finding systems, whereby descriptors of human expertise are extracted, normalized and classified so as to facilitate the finding of experts, particularly in scientific and technical fields. In this way, data mining can facilitate Institutional Memory.

Expertise extraction for teacher as well as students will be done using one of the various tried and tested psychological methods mentioned earlier. Students will be classified using decision tree induction method in three categories 'Slow learner', 'Medium learner' and 'Rapid learner' based on their performance in various parameters like, continuous evaluation in class room, class tests, mid-semester exams etc. The expertise attributes will be analyzed using k-means method for clustering and the teacher will be allocated for the students lacking in particular expertise for academic as well as personal attribute improvement. Various visualization methods like, bar graph, pie graph and line chart will be used for output of analysis.



[Fig. 2 Proposed model for student's performance prediction]

Future Scope

The proposed methodology can be extended to accommodate the evaluation of the method after the group/groups of students who are weak in the expertise area are allocated to teacher having respective expertise.

References

[1] Alaa el-Halees (2009) Mining Students Data to Analyze e-Learning Behavior: A Case Study.

[2] Behrouz.et.al., (2003) Predicting Student Performance: An Application of Data Mining Methods With The Educational Web-Based System Lon-CAPA © 2003 IEEE, Boulder, CO.

[3] Connolly T., C. Begg and A. Strachan (1999) Database Systems: A Practical Approach to Design, Implementation, and Management (3rd Ed.). Harlow: Addison-Wesley.687

[4] Erdogan and Timor (2005) A data mining application in a student database. Journal o

Aeronautic and Space Technologies July 2005 Volume 2 Number 2 (53-57)

[5] Galit.et.al (2007)Examining online learning processes based on log files analysis: a case study Research, Refelection and Innovations in Integrating ICT in Education.

[6] Henrik (2001) Clustering as a Data Mining Method in a Web-based System for Thoracic Surgery: © 2001

[7] Han,J. and Kamber, M., (2006) "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor.

[8] Kifaya(2009) Mining student evaluation using associative classification and clustering Communications of the IBIMA vol. 11 IISN 1943-7765.

[9] ZhaoHui. Maclennan.J, (2005). Data Mining with SQL Server 2005 Wihely Publishing, Inc.

[10] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M.Inayat Khan (2010) European Journal of Scientific Research, ISSN 1450-216X Vol.43 No.1 (2010), pp.24-29