A Study on Educational Data Mining

Kapila Kundu\(^1\), Ramkala Punia\(^2\)

Research Scholar, Computer Science and Engineering Department, GJUS&T University, Hisar, India\(^1\)
Research Scholar, Computer Science and Engineering Department, GJUS&T University, Hisar, India\(^2\)

ABSTRACT: Data mining can be defined as a technique to find patterns or interesting information in large amount of data. Educational Data Mining (EDM) is an emerging approach which combines data mining and education system. EDM deals with developing and exploring methods for educational data, and using those methods to better understand students’ performance, and their learning environment. This paper includes the details of data mining methods and educational data mining application area or tasks. Under data mining methods and educational data mining tasks section, most common methods are included. Classification is most dominating method used in educational data mining. Among the various tasks, predictive modeling is a task that arouses the curiosity of the researchers.

I. INTRODUCTION

Data mining means extracting novel, interesting, useful and comprehensive knowledge from large amount of data \([1]\). It includes various methods which have been applied in different application area like clustering, classification, association rule mining and sequence mining. Educational Data Mining (EDM) is an emerging approach which combines data mining and education system. EDM deals with developing and exploring methods for educational data, and using those methods to better understand students’ performance, and their learning environment \([2, 3, 4, 5, 6]\). EDM includes a variety of methods that have been presented by the researchers in the field of educational data mining in recent years. Some methods are similar to basic data mining methods like classification, clustering, association rule mining. On the other hand, some methods are unique to educational data mining like regression, ensemble methods \([5, 6, 8, 9]\). EDM deals with many educational systems like traditional classroom, e-learning, learning management system, adaptive hypermedia educational systems, intelligent tutoring system, test/quizzes, and text/contents \([4, 7, 8]\).

There are numerous application areas or tasks that have been employed in educational environment. Romero (2010), Baker (2012) and Alejandro (2014) enlist many application areas and variety of functionalities related to students, teachers and about organization under the umbrella of EDM \([2, 3, 9]\). The most commonly used tasks by the researchers are as follows: student modeling, predicting students’ behavior, assessment, feedback and support, teacher, curriculum and domain knowledge, and predicting students’ performance \([9]\). These are the frequently used application areas in educational data mining.

II. DATA MINING METHODS

Data mining can be defined as a technique to find patterns or interesting information in large amount of data. There are many methods which are used in data mining like association rule mining, clustering, classification, text mining etc \([1]\).

In this proposal we will focus on classification and clustering.

II.1. ASSOCIATION RULE MINING

Association means relationship or togetherness or connection of objects. Association rule mining indicates associated relationship between the set of objects/items. An example of an association could be that, 90% of the people who buy cookies also buy milk (60% of all grocery shoppers buy both) \([1]\).
II.ii. CLUSTERING
Clustering is an unsupervised data mining technique used to classify data into different groups based on their inherent characteristics. Many researchers have applied clustering in student performance modelling, e.g., to divide the students of interest into different groups to look at the common characteristics of the target groups under consideration like the students who have failed, the student who have achieved average score and the students who have done very well [1].

II.iii. CLASSIFICATION
In classification, the predicted variable can either be a binary or categorical variable. Classification is a supervised learning technique which classifies the data on the basis of known class variable. Some popular classification methods include decision trees, random forests, and decision rules. Classifiers are typically validated using cross-validation, where part of the data set is repeatedly and systematically held out and used to test the goodness of the model. Cross-validation should be conducted at multiple levels. It is typically standard to cross-validate at the student level in order to ensure that the model will work for new students. Some common metrics are used for cross-validation in classification i.e. AUC, kappa, precision, recall, and accuracy [1, 12].

II.iv. REGRESSION
Regression deals with continuous data. It is used to find out the relationship between the predicted (dependent) variable and other independent variables. The most popular regression method is linear regression, and logistic regression. Another method, regression tree is also fairly popular. Regressors can be validated using Kappa, linear correlation or root mean squared error (RMSE) [1, 12].

III. EDUCATIONAL DATA MINING TASKS
The tasks are the application areas in educational field. There are various tasks described in the field of educational data mining like predicting student performance, predicting student behaviour etc. we have described six common tasks used frequently in educational environment [9]. These are as follows:

III.i. STUDENT MODELING
Student modeling is used to classify the learner according to his/her characteristics such as emotion, cognition, domain knowledge, learning strategies, achievements, features, learning preferences, skill, evaluation and affects[9, 11, 13, 14, 15, 16, 17].

III.ii. PREDICTING STUDENTS’ BEHAVIOR
The main aim of predicting student behavior is to mine particular pattern behavior in order to provide appropriate help to the user [9, 19-23].

III.iii. ASSESSMENTS
Main focus of assessment is to evaluate the learners’ domain knowledge, acquisition, skill development, and achieved outcomes, as well as reflection, inquiring, sentiments [25-29].

III.iv. STUDENT SUPPORT AND FEEDBACK
Student support and feedback is used to identify that the support given by the computerized educational system is relevant to enhance the learners’ performance and achievements, or to correct their misconceptions, bugs, and faults [9, 30- 33].

III.v. TEACHER SUPPORT, DOMAIN KNOWLEDGE AND CURRICULUM
Teacher support, domain knowledge and curriculum deals with curriculum development, teaching practices, domain knowledge and teacher support [9, 33-36].
III. vi. PREDICTING STUDENTS’ PERFORMANCE

Predicting students’ performance means to detect the significant attributes, relationship between the attribute, and other aspects that affect students’ performance. Many indicators help to detect the students’ performance [37–40]. These are as follows: efficiency, evaluation, achievement, competence, resource consuming, elapsed time, correctness, deficiencies, etc. The main target of this task is to assess the learner if he/she will be able to accomplish a given task or reach a specific learning goal [9].

IV. CONCLUSION

Educational Data Mining is very promising approach which extract novel and interesting information from educational data. Classification is most dominating task used under the umbrella of EDM. Predicting students’ performance is dominant task in educational data mining that has always aroused the curiosity of researchers. Understanding factors that affect the students’ performance is very important facet that helps in improving and understanding the educational landscape. By predicting student’s performance, students can be helped at the early stages and this can improve the academic success of students.

REFERENCES

...