

A Survey on Educational Data Mining using Association Rule Mining

Mehta Smruti Hemantkumar¹, Dr. Ashish Adholiya²

¹Research Scholar, Pacific Academy of higher Education and Research University, Udaipur, Rajasthan, India

²Assistant Professor of IT and Marketing, Faculty of Management, Pacific Academy of higher Education and Research University, Udaipur, Rajasthan, India

ABSTRACT

In data mining, the knowledge is extracted from the large data sets and the result is generated from various data mining techniques. Here, we used educational data mining for improving graduate students' performance. The most popular application of data mining in education is to predict student performance. The Educational Data Mining (EDM) spotlights on modeling and evaluate student's performance based on scores in an examination. Different techniques like neural networks, association rules, regression, Bayesian networks, and rule-based systems are applied to study the educational data. Association rules are an essential part of every research field. In this perception, we use association rule mining techniques for discovering relations between variables in large databases. Various algorithms are available for association rule mining. Apriori algorithm can be used with WEKA tool to extract the set of rules and scrutinize the given data to classify the student based on their academic performance.

Keywords : Data Mining, Frequent Pattern Mining, Apriori

I. INTRODUCTION

Education is an important part for the betterment and development of a nation. Today, higher institutions facing challenges in high competitive environment. Data mining is the process of extracting previously unknown and probably useful information from large amount of data. Now-a-days data mining have been carried out in education field.

Student's performance is the most important part of higher educational institutions. Owing to wide-ranging computerization and limited storage facilities, an enormous amount of information implanted in huge database belonging to different endeavor.

Data mining is a valuable utensil and extract useful patterns from data warehouses, which may be used to

find future trends. The most preferred aspects of data mining is to look for knowledge from enormous data. Data can be categorized in terms of size and length. With the advent of data mining tools and techniques, the data miner may extract knowledge from large data warehouses very easily which is better for the organizations [1].

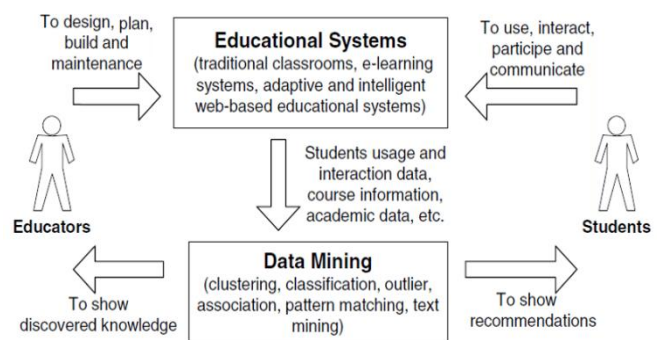


Figure 1 depicts the life cycle of data mining in education system [7].

Students: To provide activities to learners and also provides resources to improve their learning.

Educators: Find the students who require extra support, predict student's performance, make the learners group, find the learner's patterns to identify the most frequently made mistakes, analyze student's learning behavior.

Educational Systems: Compare data mining techniques and recommend the most useful one to develop data mining tools for educational purpose.

System Administrators: Utilize available resources more effectively to improve educational program.

This paper is divided into following sections: Second section represents the literature survey, Third section represents overview of frequent pattern mining, Fourth section represents Apriori Algorithm, Fifth section presents conclusion.

II. RELATED WORK

Paresh Tanna, Dr. Yogesh Ghodasara [2] described the knowledge is discovered from the large data sets. In data mining, Frequent Pattern Mining is important. Apriori algorithm is used for frequent pattern mining to generate candidate key and pruning techniques for the satisfaction of the desired objective. This paper exhibits the use of WEKA tool for association rule mining using Apriori algorithm.

Astha Soni, Vivek Kumar, Rajwant Kaur, D. Hemavathi [3] presented the data mining techniques to study the student's performance. Educational institutions have academic database with student details. The author prepared a formation to analyze the pupil's performance using Classification technique. Decision Tree, Support Vector Machine and Naïve Bayes are Classification algorithms which help in predicting student's performance. The author used three classification models to learn the predictive function. The methods are Decision Tree, Support Vector Machine and Naïve Bayes. From these three methods, SVM classification algorithm

gives effective result. The author concluded that academic information, family details and personal information have very strong impact on the students' performance.

Siddu P. Algur, Prashant Bhat, Nitin Kulkarni [4] defines data mining as a leading tool for academic and educational field. EDM is fruitful in developing new methods to discover knowledge from academic database and used for decision making. Here, author used Random Tree and J48 algorithms using Decision Tree. From these two classification models, the Random Tree is found good as compared to J48 classification model.

Kolo David Kolo, Solomon A. Adepoju, John Kolo Alhassan [5] described that this research is based on the use of a decision tree for predicting students' academic performance. The society improves the quality of its citizens on the basis of education. The author used Statistical Package for Social Studies (SPSS) to apply the Chi-Square Automatic Interaction Detection (CHAID) to produce the decision tree.

Dr. Varun Kumar, Anupama Chadha [6] performed a case study to improve the quality of education and discover the factors affecting to academic results to increase the success of students. Here, author used Association rule mining technique to enhance the quality of students' performances at Post Graduation Level.

Frequent Pattern Mining

Association rule mining discovers the relationships between large data sets. Association rules display the attributes that occur frequently together in a given dataset. Support, Confidence and Lift are the interestingness measures of association. Association uses if-then pattern to classify the relationships.

Support means how frequently the item is appearing in the data. Confidence means the number of times the if-then relationships are found true. Lift means the ratio of the observed joint probability of X and Y to the expected joint probability if they were statistically independent.

Apriori Algorithm

Apriori algorithm was the first algorithm that was proposed for frequent itemset mining. It was improved by R. Agarwal and R. Srikant. This algorithm uses two steps “join” and “prune” to reduce the search space. It is an iterative approach to discover the most frequent itemsets.

Join step: Let L_k is the itemsets available in a frequent set L_k . To find L_k , join operation is performed on L_{k-1} with itself.

Prune step: The size of C_k can be reduced by using the Apriori property.

According to the pruning rule, for any k -itemsets in C_k , if any $(k-1)$ subset of this is not belong to L_{k-1} , then the candidate can't be frequent and can be removed from C_k .

Apriori takes minimum support value and transaction database. For example, consider the following transaction database containing five transactions.

T1: Pen, Paper

T2: Pen, Eraser, Sharpener, Ink

T3: Pen, Eraser, Sharpener, Pencil

T4: Pen, Paper, Eraser, Sharpener

T5: Pen, Paper, Eraser, Pencil

The association rule in following transaction is:

{Paper, Eraser} \rightarrow {Sharpener, Pencil}

III. CONCLUSION

This paper is the review of Educational Data Mining. EDM has been introduced as up and coming research field. This field helps higher institutions to develop institutional effectiveness and learning process. This study will be helpful to teachers, learners and institutions to improve their performance. Association rule is an efficient technique for finding frequent itemset in transaction database. Apriori algorithm can be used to mine the frequent itemsets.

Cite this article as : Mehta Smruti Hemantkumar, Dr. Ashish Adholiya, "A Survey on Educational Data Mining using Association Rule Mining", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 3 Issue 8, pp. 451-453, November-December 2018. Journal URL : <http://ijsrcseit.com/CSEIT1838107>

IV. REFERENCES

- [1]. Keyuri S. Lakshmi Prabha, Dr.A.R.Mohamed Shanavas, “Application of Educational Data Mining techniques in e-Learning- A Case Study”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5), 2015, 4440-4443.
- [2]. Paresch Tanna, Dr. Yogesh Ghodasara, “Using Apriori with WEKA for Frequent Pattern Mining”, International Journal of Engineering Trends and Technology (IJETT), Volume 12, Number 3, Jun 2014.
- [3]. Astha Soni, Vivek Kumar, Rajwant Kaur, D. Hemavathi, “PREDICTING STUDENT PERFORMANCE USING DATA MINING TECHNIQUES”, International Journal of Pure and Applied Mathematics, Volume 119, No. 12, 2018.
- [4]. Siddu P. Algur, Prashant Bhat, Nitin Kulkarni, “Educational Data Mining: Classification Techniques for Recruitment Analysis”, I.J. Modern Education and Computer Science, 2016, 2, 59-65.
- [5]. Kolo David Kolo, Solomon A. Adepoju, John Kolo Alhassan, “A Decision Tree Approach for Predicting Students Academic Performance”, I.J. Education and Management Engineering, 2015, 5, 12-19.
- [6]. Dr. Varun Kumar, Anupama Chadha, “Mining Association Rules in Student’s Assessment Data”, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012.
- [7]. Sen, Umesh Kumar, “A Brief Review Status of Educational Data Mining”, International Journal of Advanced Research in Computer Science & Technology (IJARCST 2015), Vol. 3, Issue 1 (Jan. – Mar. 2015).