Learner's Prognostic Analysis Using Class Association Rule

[1]D. Magdalene Delighta Angeline[*], [2]Dr. P. Ramasubramanian, [3]I. Samuel Peter James
[1]Computer Science & Engineering, Dr.G.U.Pope Col. of Engineering, India
[2]Principal, Annai Vailankani College of Engineering, India
[3]Research School Bharathiyar University, India

Abstract— Data Mining is a tool that helps in understanding the samples extracted with the set of rules and to develop relationships among variables stored in large data set. Mining the knowledge from the database and providing a better solution for a specific problem yields good outcome to the developers. The educational institution focuses on obtaining good results that can be achieved by using the data mining techniques to predict the performance of the students and to impart the quality of education. In this paper, rules are extracted with the help of class association rule mining algorithm. This would help to identify the average and below average learners and to improve their performance to provide superior results.

Keywords— Class Association Rules, Classification, Data mining, Knowledge Discovery, Predictive Models

I.　INTRODUCTION

　　Education is an essential element for the betterment and growth of a country. Mining in educational environment is called Educational Data Mining, concern with developing new methods to discover knowledge from educational databases[2],[3], in order to analyze students trends and behaviors toward education[1]. Lack of deep and enough knowledge in higher educational system may prevent system management to achieve quality objectives, data mining methodology can help bridging this knowledge gaps in higher education system. A better student model yields better instruction, which leads to improved learning. More accurate skill diagnosis leads to better prediction of what a learner knows which provides better assessment. Better assessment leads to more efficient learning overall. The primary goals of Data Mining in practice tend to be Prediction and Description [1]. Predicting performance involves variables like Attendance, IAT Marks, Assignment Grades etc. in the student database to predict the unknown or future values of interest.

　　Data mining is the core process of knowledge discovery in databases. It is the process of extracting of useful patterns from the large database. To analyze the large amounts of collected information, the area of Knowledge Discovery in Databases (KDD) provides techniques which extract interesting patterns in a reasonable amount of time. Therefore, KDD employs methods at the cross point of machine learning, statistics and database systems. Data mining is the application of efficient algorithms to detect the desired patterns contained within the given data. Association rules mining are one of the major techniques of data mining and it is perhaps the most common form of local-pattern discovery in unsupervised learning systems. The technique is likely to be very practical in applications. Association Rules will permit you to discover rules of the kind If X then (likely) Y where X and Y can be particular items ,values, words, etc. , or conjunctions of values, items, words, etc. An association rule is composed of two item sets:

　　1.　Antecedent or Left-Hand Side (LHS)
　　2.　Consequent or Right-Hand Side (RHS)

　　It describes the relationship between Support, Confidence and interestingness. The support and confidence are usually referred as interestingness measures of an association rule. Association rule mining is the process of finding all the association rules that pass the condition of min-support and min-confidence. In order to mine these rules, first the support and confidence values have to be computed for all of the rules and then compare them with the threshold values to prune the rules with low values of either support or confidence.

　　Classification is a predictive data mining technique which makes prediction about values of data using known results found from different data [12]. Predictive models have the specific aim of allowing us to predict the unknown values of variables of interest given known values of other variables. Predictive modeling can be thought of as learning a mapping from an input set of vector measurements to a scalar output [13]. There are two types of predictive modeling, namely classification, for categorical dependent variables, and value prediction, for continuous dependent variables. Classification is appropriate if the goal is to predict group membership of new records based on their characteristics. Classification maps data into predefined groups of classes which is often referred to as supervised learning because the classes are determined before examining the data. Prediction models that include all personal, social, psychological and other environmental variables are necessitated for the effective prediction. There are several algorithms for data

classification such as decision tree and Naive Bayes classifiers. With classification, the generated model will be able to predict a class for given data depending on previously learned information from historical data.

The main objective of this paper is to use data mining methodologies to study learners' performance in the courses. Data mining provides many tasks that could be used to study the learner performance by performing classification of data.

II. LITERATURE REVIEW

In [2], gave a case study that use students data to analyze their learning behavior to predict the results and to warn students at risk before their final exams. [3] used educational data mining to identify and enhance educational process which can improve their decision making process. Predicting student performance, one of the tasks in educational data mining, has been taken into account recently [11],[10]. It was selected as a challenge task for the KDD Cup 2010. To address the problem of predicting student performance, many papers have been published but most of them are based on traditional classification/regression techniques. Recently, [9], [15], [10] have proposed using recommendation techniques, e.g. matrix factorization, for predicting student performance. The authors have shown that predicting student performance can be considered as rating prediction since the student, task, and performance would become user, item, and rating in recommender systems, respectively. Modeling of student performance at various levels is discussed in [4], [5], and [6]. [4] applied a data mining technique based on association rules to find weak tertiary school students (n= 264) of Singapore for remedial classes.

Han and Kamber [14] describes data mining software that allow the users to analyze data from different dimensions, categorize it and summarize the relationships which are identified during the mining process. Galit [2] gave a case study which is to analyze the learners learning behavior to predict the results and to warn students at risk before their final exams. Al-Radaideh et al. [12], applied the data mining techniques, particularly classification to help in improving the quality of the educational system by evaluating learners data. Al-Radaideh, et al [12] applied a decision tree model to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. Three different classification methods namely ID3, C4.5, and the Naive Bayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models

A. Association rule

Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time. Association rule generation is usually split up into two separate steps:

1. First, minimum support is applied to find all frequent itemsets in a database.
2. Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

While the second step is straightforward, the first step needs more attention.

B. Mining class association rules (CAR)

Normal association rule mining does not have any target where CAR finds all possible rules that exist in data, i.e., any item can appear as a consequent or a condition of a rule. However, in some applications, the user is interested in some targets. E.g, the user has a set of text documents from some known topics. He/she wants to find out what words are associated or correlated with each topic. So, mining class association rules gives solution to the above mentioned. A Class Association Rule (CAR) is a special type of Association Rule (AR) that describes an implicative co-occurring relationship between a set of items and a pre-defined class, expressed in the form of an

"IF antecedent (input-attributes) THEN

consequent (class)"

rule. The set of CARs thus consists of all the possible rules that are both frequent and accurate. The association rule for finding the grade of a learner is as follows:

- If $x >= 90$ then grade $=A$.
- If $80<=x<90$ then grade $=B$.
- If $70<=x<80$ then grade $=C$.
- If $60<=x<70$ then grade $=D$.
- If $x<50$ then grade $=E$.

C. Problem definition

Let T be a transaction data set consisting of n transactions. Each transaction is also labeled with a class y. Let I be the set of all items in T, Y be the set of all class labels and $I \cap Y = \varnothing$. A class association rule (CAR) is an implication of the form

$$X \rightarrow y, \text{ where } X \subseteq I, \text{ and } y \in Y.$$

D. Mining algorithm

Unlike normal association rules, CARs can be mined directly in one step. The key operation is to find all rule items that have support above minsup. A rule item is of the form:

(C, y)

where C is a set of items from I (i.e., $C \subseteq I$), and $y \in Y$ is a class label.

Each rule item basically represents a rule:

$$C \rightarrow y,$$

E. Multiple minimum class supports

The multiple minimum support idea can also be applied here. The user can specify different minimum supports to different classes, which effectively assign a different minimum support to rules of each class. For example, a data set with two classes, Yes and No is assumed.

- Rules of class "Yes" to have the minimum support of 5%
- Rules of class "No" to have the minimum support of 10%.

By setting minimum class supports to 100%, it is being told the algorithm to not generate rules of those classes. This is a very useful trick in applications.

III. EXPERIMENT

In an educational institution the overall performance of a learner is determined by internal assessment as well as external assessment. Internal assessment is made on the bases of a learner's assignment marks, class tests, lab work, attendance, previous semester grade and his/ her involvement in extra curriculum activities. While at the same time external assessment of a learner based on marks scored in final exam. The proposed model makes prediction about poor, average and good ratio of students based on class performance as well as system inform the learners about the ratio of class attendance.

The proposed model also deals with entrance ratio of learners in a particular department and exit ratio after successful completion of a degree. The proposed model identifies the weak students before final exam in order to improve his/her performance. Teachers can take appropriate steps at right time to improve the performance of learner in final exam. It deals with both kind of assessments especially internal assessment in order to predict learners whose performance is low. This model check the performance of learner at different levels before final exam in order to predict poor learners and take appropriate steps to save them from failure.

A. Data Set

The data set used in this study was obtained from department of Computer Science, Dr.G.U.POPE College of Engineering.

B. Data mining Process

Data mining process consist of following steps:

B1. Data Preparation

Initially, the data was stored in five different tables. The tables are personal data, class attendance, internal performance, class performance and behavior. Then data stored in different tables was joined in a single table and errors were removed.

B2. Data Transformation

In this step we determined the fields of study used for analysis. Data is inform of yes/no is transformed in form of 1/0.

C. Rule Generation

The rules were generated by using CAR technique. The implementation was done in data mining software.

TABLE I SAMPLE RULE SET GENERATED BY CAR

If assignment_grade= 'A' and attendance =86.98 and IAT mark = 86 then final_result='good'
If assignment_grade= 'A' and attendance =85.86 and Academic_Trouble='no' then final_result='good'
If assignment_grade= 'B' and attendance =86.98 and IAT mark = 60 then final_result='average'
If assignment_grade= 'C' and attendance =60 and Academic_Trouble='yes' then final_result='good'
If assignment_grade= 'D' and attendance =65 and IAT mark = 11 then final_result='poor'

TABLE II THE NUMBER OF GENERATED IF-THEN RULES OF EACH LENGTH FOR THE EDUCATIONAL DATA.

Length of rules	1	2	3	4	5
Number of rules	47	35	27	23	19

D. Discussion on the result

A set of rules were generated specific to each class and analysis was done for the given data to classify the learner based on their performance in academics. The best rules were selected where the performance identification was set as the target of the learners. The rules were evaluated based on the confidence and support.

Fig. 1 The analysis chart which shows overall Percentage of learner's Final Performance
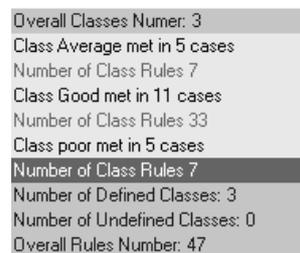

Fig. 2 Class Rules

The number of class rules generated was 47 with probability of 99.99%.

IV. CONCLUSION

The data classification is performed using class association rule mining in this study. The information generated after the implementation of data mining technique may be helpful for instructor as well as for learners. This work classifies the categories of learner's performance in their academic qualification by reducing failure ratio by taking appropriate steps at right time to improve the quality of education.

REFERENCES

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

[2]   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.

[3]   Erdogan and Timor (2005) A data mining application in a student database. Journal of Aeronautic and Space Technologies July 2005 Volume 2 Number 2 (53-57)

[4]   Y. Ma, B. Liu, C.K. Wong, P.S. Yu, and S.M. Lee, ―Targeting the Right Students Using Data Mining‖, Proceedings of KDD, International Conference on Knowledge discovery and Data Mining, Boston, USA, 2000, pp. 457-464.

[5]   S. Kotsiantis, C. Pierrakeas, and P. Pintelas, ―Prediction of Student's Performance in Distance Learning Using Machine Learning Techniques‖, Applied Artificial Intelligence, Vol. 18, No. 5, 2004, pp. 411-426.

[6]   P. Cortez, and A. Silva, ―Using Data Mining To Predict Secondary School Student Performance‖, In EUROSIS, A. Brito and J. Teixeira (Eds.), 2008, pp.5-12.

[7]   Thai-Nghe, N., Drumond, L., Horvath, T., Krohn-Grimberghe, A., Nanopoulos, A., and Schmidt-Thieme, L. 2011. Factorization techniques for predicting student performance. In Educational Recommender Systems and Technologies: Practices and Challenges (In press), O. C. Santos and J. G. Boticario, Eds. IGI Global.

[8]   Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., and Schmidt-Thieme, L. 2010. Recommender system for predicting student performance. In Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010). Vol. 1. Elsevier's Procedia CS, 2811 { 2819.

[9]   Thai-Nghe, N., Gantner, Z., and Schmidt-Thieme, L. 2010. Cost-sensitive learning methods for imbalanced data. Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN 2010).

[10] Thai-Nghe, N., Horvath, T., and Schmidt-Thieme, L. 2011. Personalized forecasting student performance. In Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies (ICALT 2011) (to appear). IEEE CS.

[11] Toscher, A. and Jahrer, M. 2010. Collaborative _ltering applied to educational data mining. KDD Cup 2010: Improving Cognitive Models with Educational Data Mining.

[12] Q.A.AI-Radaideh, E.M. AI-Shawakfa, and M.I. AI-Najjar, "Mining Student Data using Decision Trees", International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.

[13] David Hand, Heikki, Mannil Padraic smyth, "Principles of Data Mining" PHI

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

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