International Journal of Educational Research and Technology
P-ISSN 0976-4089; E-ISSN 2277-1557
IJERT: Volume 7 [2] June 2016: 33-37

Website: www.soeagra.com/ijert.html
ICDS : 3.699[University of Barcelona, Barcelona]
Global Impact Factor: 0.765 Scientific Journal Impact Factor: 3.72
Journal Impact Factor (JIF) : 1.54

OPEN ACCESS

ORIGINAL ARTICLE

The Analysis of the Student Academic Performance Based on Data Mining

Sheng Zhang[1], Xiangqian He[2]
College of Medical Informatics
Chongqing Medical University, 1 Yixueyuan Road
Yuzhong District
Chongqing 400016, China
E-mail: vicky_zhang9206@hotmail.com: hexiangqian@cqmu.edu.cn

ARTICLE HISTORY
Received:
20.02.2016
Revised
26.03.2016
Accepted
12.05.2016

ABSTRACT
Educational management is a complex work. The student academic performance is of great concern to the university administrators. Academic decisions will result in academic performance changes. Subjective decision-making is possible without scientific approaches. Having been applied to the field of education, data mining is capable to extract information and develop significant relationships among variables stored in large data set. In this paper, association rule mining algorithms of data mining are used to provide an instructive suggestion to the academic planners in decision making by evaluating student data to study the factors that may affect the performance of students.
Keywords *Data mining, Academic performance, Association rules, Educational data mining, Association rule mining algorithms*

CITATION OF THIS ARTICLE
Sheng Zhang, Xiangqian He.The Analysis of the Student Academic Performance Based on Data Mining . Inter. J. Edu. Res. Technol. 7[2] 2016; 33-37
DOI: 10.15515/ijert.0976 4089.7.2.3337

INTRODUCTION

Nowadays, data mining has become the most potential area in computer science. The process of discovering valuable information from data set for further application is dealt with well using lots of data mining tools. Association rule mining (ARM) is one of the most popular technique of data mining. ARM is considered as an unsupervised learning method for discovering close relationships among patterns and rules in large databases. The patterns and rules are discovered through majority of frequent and repeated items in the data sets. The aim of ARM is to find out association rules that satisfy the predefined minimum support and confidence from a given database. ARM is widely applied for a variety of fields such as super market sales-prediction, telecommunication networks, financial forecast, medical diagnosis, risk management, etc [1]. Recently ARM has been used to help in enhancing the quality of educational systems.

Educational data mining is a potential research area which extract fertile unknown patterns from database for better understanding and improving student academic performance [2]. Cristóbal Romero et al. [3] compared Apriori and several rare association rule mining algorithms. It is shown that Apriori finds a huge number of frequent rules and other rare association rule mining algorithms Apriori-Inverse and Apriori-Rare discover rare association rules. These rules present infrequent students' behaviors/activities in Moodle, a free Learning Content Management Systems. Instructors can evaluate the relations between the on-line activities and the final mark obtained by the students. Behrouz Minaei-Bidgoli et al. [4] present an approach to classify students to predict their final year grade based on the features extracted from logged data in an education web-based system. Talavera L et al. [5] propose some preliminary experiments using clustering algorithms to characterize similar behavior groups in student data mining. Some areas such as student assessments in online courses and relevant evaluation of educational websites are used to improve educational standards and management [6,7,8,9]. Robertas et al. [10] analyze the informatics course examination results using association rules. Based on the strength of the association rules, the course topics are ranked according to their importance for final course marks and are improved to achieve higher student learning effectiveness and progress. Mukesh Sharma et al. [11] compared Apriori and PredictiveApriori using association rule mining tool weka.

In this paper, we presents an analysis of the performance score of students in final exam using association rule mining to discover the latent relationships among the basic information of students such as sex, nationality and mark. Student data is analyzed using Apriori, Predictive Apriori and Tertius to provide constructive recommendations to administrators for decision making in improving academic performance. This paper is organized as follows. Methods are presented in section 2. Results and discussion are given in section 3. Conclusions are provided in section 4.

METHODS

This study focuses on using three association rule mining algorithms of data mining technologies to dig into the factors affecting the student academic performance.

Association rule mining algorithms

Apriori

Apriori is one of the most popular algorithm of ARM. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database. Apriori generates candidate item sets of length k from item sets of length k-1. Then it prunes the candidates which have sub pattern. According to the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates. The algorithm discovers all item sets in the data that satisfy predefined *support* threshold (*minsup*). Then it generates association rules from those item sets constrained by predefined *confidence* threshold (*minconf*).

PredictiveApriori

PredictiveApriori is one of the variation of Apriori. Tobias Scheffer analyzed association rules from a view of Bayesian statistics [12]. PredictiveApriori based on expected predicted accuracy was proposed. Rules with largest expected predicted accuracy can be mined by setting up the number of the best rules. The algorithm is also a confidence-based ARM algorithm. Unlike using confidence in Apriori, the algorithm maximized the predictive accuracy of the rules during the mining process. It searches with an increasing support threshold for the best 'n' rules concerning a support-based corrected confidence value. Predictive accuracy of rules is among the 'n' best and it is not subsumed by a rule with at least the same expected predictive accuracy. The algorithm considered the influence of support and confidence on the expected predictive accuracy of the $X \Rightarrow Y$ association rules comprehensively.

Tertius

Tertius is a general-purpose first-order logic rule discovery algorithm [13]. It utilizes the first order logic that offers the ability to deal with structured, multi-relational knowledge. Tertius employs a top-down search over the space of possible rules. A attributes with on the average V values are given, and n literals are required when searching for rules. The number of rules mined using Tertius is of the order (AV)^n. The algorithm use the degree of confirmation as a measure for ranking given rules.

Experimental procedures

The data is extracted from the database of Chongqing Medical University, which is a collection of data associated with all information of faculties in school. Student data is a record keeping detailed information of students (e.g., student number, name, sex, nationality and mark). We exported data of 4666 students from the database. Cleaning data is the first step since the noise existing in data disturbs the subsequent steps. The data with missing values would be deleted. Then it would be imported into MySQL, which is used to store the student data. Then the preprocessing technique 'Unsupervised Discretization' was applied on the datasets. Unsupervised Discretization is applied to convert a range of numeric attributes into nominal attributes. For example, the numeric grades have been divided into categorical classes: ABSENT, FAIL, PASS, EXCELLENT, which are easier to understand for users. At last, association rules were obtained from Apriori, PredictiveApriori and Tertius algorithms using Weka, a data mining tool. The description of attributes of student is shown in Table 1.

Attributes	Description	Value
Sex	Male and Female	Male, Female
Nationality	Members of a particular ethnic group in China	Han , Tibetan , Yi , Manchu, Tujia
Type	One who upgrades from high school to university using HU to represent One who upgrades from junior college to university using JU to represent	HU, JU
Mark	Final mark obtained by the students	ABSENT, FAIL, PASS, EXCELLENT

Table 1. Attributes used for each student instance

• Sex: Among the 4666 students, 3567 students are female (76.45%), 1099 students are male (23.55%).
• Nationality: Among the 4666 students, Han students account for 91.13%, 118 Tibetan students for 2.53%, 59 Yi students for 1.26%, 59 Manchu students for 1.26%, 178 Tujia students for 3.82%.
• Type: 4620 of the 4666 students were upgraded from high school to university (HU) accounting for 99.01%. 46 students were upgraded from junior college to university (JU) accounting for 0.99%.
• Mark: 3574 of the 4666 students obtained good score in the final exam (76.60%), 997 students passed the exam with normal grade (21.37%), 58 students failed the exam (1.24%), 37 students were absent from the exam (0.79%).
The pie charts of four attributes are shown in Figure 1.

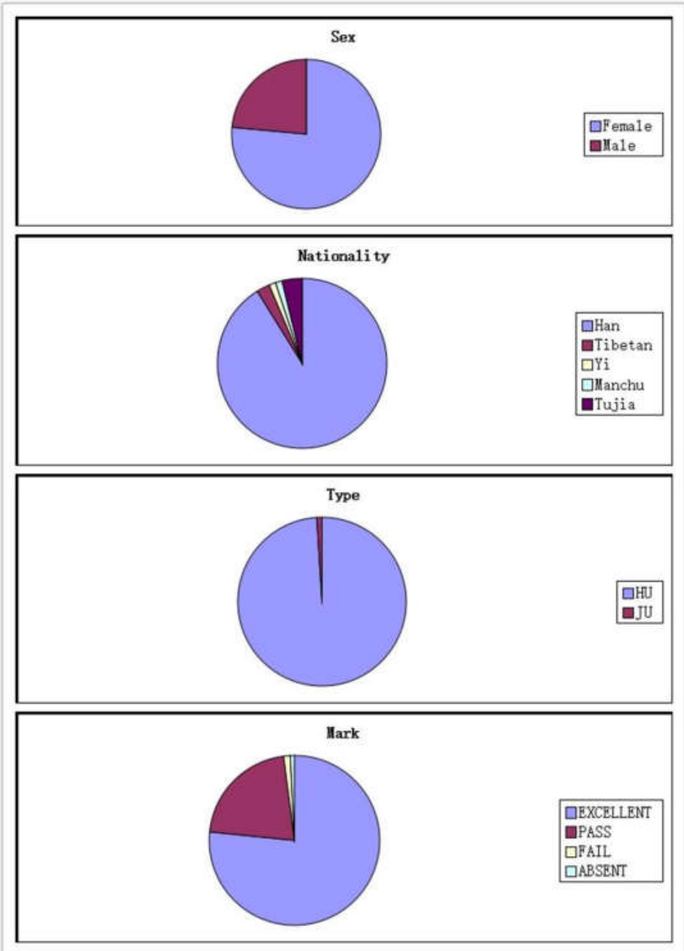

Figure 1. Value distribution for the attributes Sex, Nationality, Type, Mark

RESULTS AND DİSCUSSİON

Firstly, data are analyzed by Apriori, using different support threshold set at 0.05, 0.1, 0.2. We assigned the value 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 as the confidence threshold respectively. Rules are found using Apriori can be seen graphically in Figure 2.

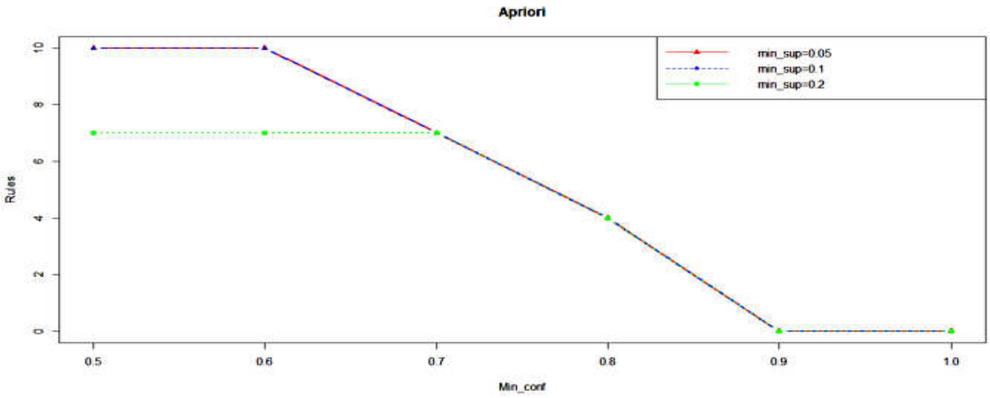

Figure 2. Graphical representation of effect of min_sup, min_conf on rules

Min_sup represents support threshold and Min_conf represents confidence threshold in Figure 2. We can find that the same number of rules are found when min_sup is set at 0.05 and 0.1. Rules are decreased when min_sup is increased. Thus, min_sup set at 0.1 is appropriate. The rules discovered by Apriori are shown in Table 2.

Rule	Antecedent	Consequent	Confidence
1	Sex=Female, Nationality=Han, Type=HU	Mark=EXCELLENT	0.82
2	Sex=Female, Nationality=Han	Mark=EXCELLENT	0.82
3	Sex=Female, Type=HU	Mark=EXCELLENT	0.81
4	Sex=Female	Mark=EXCELLENT	0.81
5	Nationality=Han, Type=HU	Mark=EXCELLENT	0.77
6	Nationality=Han	Mark=EXCELLENT	0.77
7	Type=HU	Mark=EXCELLENT	0.77

Table 2. Rules extracted using Apriori

As we can see, only the rule of Mark=EXCELLENT was mined. The meaning of rule 1: Sex=Female, Nationality=Han, Type=HU $\Rightarrow$ Mark=EXCELLENT can be described that The Han female students who were upgraded to university from high school got good scores in exam. The rest rules are parts of rule 1. Obviously, there are some limits in Apriori. Then, PredictiveApriori and Tertius were applied to the same datasets, The results of analysis are shown in Table 3 and Table 4 separately.

Rule	Antecedent	Consequent	Accuracy
1	Nationality=Yi	Mark=EXCELLENT	0.897
2	Nationality=Tujia	Mark=EXCELLENT	0.798
3	Type=JU	Mark=EXCELLENT	0.696
4	Sex=Female, Nationality=Han, Type=HU	Mark=EXCELLENT	0.554
5	Nationality=Manchu	Mark=FAIL	0.335
6	Nationality=Tibetan	Mark= FAIL	0.329
7	Type=JU	Mark= FAIL	0.212

Table 3. Rules extracted using PredictiveApriori

We can see that some rules with more interesting information were mined such as Nationality=Yi, Tujia, Manchu, Tibetan and Type=JU. We discovered that hidden patterns and relationships from the rules are helpful in decision making. It is shown in rule 1 and rule 2 that Yi and Tujia students got good scores in exam. Rule 3 and rule 7 shows that the test results of students who were upgraded to university from junior college are unstable. It is shown in rule 5 and rule 6 that Manchu and Tibetan students failed the

exam. Comparing the results with that of Apriori, more factors affecting the student academic performance were discovered in PredictiveApriori. From the knowledge discoveries above, student management department should focus on cultivating the outstanding student from minority groups Yi and Tujia. By comparison, the Manchu students and Tibetan students with poor grades might need help in their studies. Besides, students who were upgraded to university from junior college should be paid more attention by instructors.
The rules discovered by Tertius are shown in Table 4. It is shown that female students behave better in exam than male students. Instructors should provide male students with more guidance and care.

Rule	Antecedent	Consequent	Confirmation
1	Sex=Female, Nationality=Han, Type=HU	Mark=EXCELLENT	/*0.154411 0.118946 */
2	Sex=Female, Nationality=Han	Mark=EXCELLENT	/*0.151128 0.121946 */
3	Sex=Female, Type=HU	Mark=EXCELLENT	/*0.146929 0.140806 */
4	Sex=Female	Mark=EXCELLENT	/*0.143831 0.143806 */
5	Sex=Male	Mark=FAIL	/*0.111178 0.157951 */
6	Nationality=Yi	Mark=EXCELLENT	/*0.036697 0.001072 */
7	Nationality=Tibetan	Mark= FAIL	/*0.020854 0.017360 */

Table 4. Rules extracted using Tertius

CONCLUSİONS

In this paper, we have explored the applications of association rule mining algorithms over student data to analyze and find out student academic performance and to improve student management. The potential of the data mining algorithm for enhancing the effectiveness of academic planners in decision making is shown in this study. Academic administrators should pay more attention to male students, the students from minority groups and the students who were upgraded from junior college to university, which provides a certain theoretical basis for decision making and further education reform.

REFERENCE

1. Kotsiantis S, Kanellopoulos D. Association Rules Mining: A Recent Overview[J]. International Transactions on Computer Science & Engineering, 2010, 32(32):71-82.
2. Dogan B, Camurcu A Y. Association Rule Mining from an Intelligent Tutor[J]. Journal of Educational Technology Systems, 2008, 36(36):433-447.
3. Romero C, Romero J R, Luna J M, et al. Mining Rare Association Rules from e-Learning Data[C]// Educational Data Mining 2010, The 3rd International Conference on Educational Data Mining, Pittsburgh, PA, USA, June 11-13, 2010. Proceedings. 2010:171-180.
4. Minaei-Bidgoli B, Kashy D A, Kortemeyer G, et al. Predicting student performance: an application of data mining methods with an educational Web-based system[C]// fie. IEEE Computer Society, 2003:T2A-10.
5. Talavera, L., and Gaudioso, E. “Mining student data to characterize similar behavior groups in unstructured collaboration spaces”. In Proceedings of the Arti_cial Intelligence in Computer Supported Collaborative Learning Workshop at the ECAI ,Valencia, Spain, 2004.
6. Encheva S, Tumin S. Application of Association Rules for Efficient Learning Work-Flow[M]// Intelligent Information Processing III. Springer US, 2006:499-504.
7. Hsu H H, Chen C J, Tai W P. Towards Error-Free and Personalized Web-Based Courses[C]// Proceedings of the 17th International Conference on Advanced Information Networking and Applications. IEEE Computer Society, 2003:99-104.
8. Hsu P L, Lai R, Chiu C C. The hybrid of association rule algorithms and genetic algorithms for tree induction: an example of predicting the student course performance[J]. Expert Systems with Applications, 2003, 25(1):51-62.
9. Chan A Y K, Chow K O, Cheung K S. Online course refinement through association rule mining[J]. Acta Press, 2005.
10. Robertas Damaševičius. Analysis of Academic Results for Informatics Course Improvement Using Association Rule Mining[J]. Progress in Orthodontics, 2013, 14(1):1-15.
11. Mukesh Sharma, et al. Evaluating the performance of apriori and predictive apriori algorithm to find new association rules based on the statistical measures of datasets.[J]. International Journal of Engineering Research and Technology, 2012, 1:1-5.
12. Scheffer T. Finding Association Rules That Trade Support Optimally against Confidence.[J]. Lecture Notes in Computer Science, 2001, 2168(4):424-435.
13. Flach P A, Lachiche N. Confirmation-guided discovery of first-order rules with Tertius[J]. Machine Learning, 2000, volume 42(1-2):61-95(35).