International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 5 3151 - 3154

Knowledge Discovery in Higher Education using Association Rule Mining

Prashant G. Tandale
Research Student, Bharati Vidyapeeth University
Institute of Management, Kolhapur
Maharashtra, India
prashant.g.tandale@gmail.com

Dr. K. M. Alaskar
Professor, Bharati Vidyapeeth University
Institute of Management, Kolhapur
Maharashtra, India
k.alaskar@rediffmail.com

Abstract— Since last few years, so many statistical tools have been used to analyze students' performance from different points of view. This paper presents data mining in education environment that identifies students' failure patterns using association rule mining technique. The identified patterns are analyzed to offer a helpful and productive recommendations to the academic planners in higher institutions of learning to improve their decision making process. This will also aid in the curriculum structure and modification in order to improve students' academic performance and reduce failure rate.

Keywords- *Higher Education, Knowledge Discovery, Frequent Pattern, Association Rule Mining, Apriori Algorithm*

*****

I. INTRODUCTION

Education is an essential element for the betterment and progress of a country. It enables the people of a country civilized and well mannered. [6] Today the important challenge that higher education faces, is reaching a stage to facilitate the universities in having more efficient, effective and accurate educational processes. To date, higher educational organizations are placed in a very high competitive environment and are aiming to get more competitive advantages over the other competitors. To remain competitiveness among educational field, these organizations need deep and enough knowledge for a better assessment, evaluation, planning, and decision-making. [5] The required knowledge cannot be gained from the tailor made software used now a day. Data mining incorporates a multitude of techniques from a variety of fields including databases, statistics, data visualization, machine learning and others.[7] The data mining technology can discover the hidden patterns, associations, and anomalies from educational data. This knowledge can improve the decision making processes in higher educational systems. Data mining is considered as the most suited technology appropriate in giving additional insight into the lecturer, student, alumni, manager, and other educational staff behavior and acting as an active automated assistant in helping them for making better decisions on their educational activities.[1] The data mining techniques can help the institutes in extracting patterns like students having similar characteristics, Association of students' attitude with performance, what factors will attract meritorious students and so on.[2]

II. REVIEW OF LETERATURE

C. Romero and S. Ventura (2010) survey the relevant studies carried out in the field of education. They have described the types of users, types of educational environments and the data they provide. Also they have explained in their work the common tasks in the educational environment that have been resolved through data mining techniques.[1]
Hua-long Zhao (2008) has done Multidimensional cube analysis by taking use of OLAP technology and has shown that the curriculum chosen by the students can depend upon many angles like teacher, semester and student. He has used Star model of data warehouse to the analysis of curriculum which can provide certain policy making support for different levels of education policy- maker in the school.
Fadzilah Siraj and Mansour Ali Abdoulha (2009) have used data mining techniques for understanding student enrolment data. They have done comparative study of three predictive data mining techniques namely Neural Network, Logistic regression and Decision tree. The results obtained can be used by the planners to formulate proper plan for the university.[4]
Shaeela Ayesha et al. (2010) discusses data mining technique named k-means clustering is applied to analyze student's learning behavior. Here K-means clustering method is used to discover knowledge that come from educational environment.[6]
W.M.R. Tissera et al. (2006) present a real-world experiment conducted in an ICT educational institute in Sri Lanka. A series of data mining tasks are applied to find relationships between subjects in the undergraduate syllabi. This knowledge provides many insights into the syllabi of different educational programmes and results in knowledge critical in decision making that directly affects the quality of the educational programmes. [7]
Hongjie Sun (2010) conducts a research on student learning result based on data mining. It is aimed at putting forward a rule-discovery approach suitable for the student learning result evaluation and applying it into practice so as to improve learning evaluation skills and finally better serve learning practicing.[9]
S. Anupama Kumar (2011) applied decision tree algorithm on student's internal assessment data to predict their performance in the final exam. The outcome of the decision tree predicted the number of students who are likely to fail or pass.

III. KNOWLEDGE DISCOVERY PROCESS

Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships which are helpful in decision making

A. Association Rule Mining

Association rule mining associates one or more attributes of a dataset with another attributes, to discover hidden and important relationship between the attributes, producing an if-then statement concerning attribute values in form of rules. The formal definition of association rule mining is

Let I = {I_1,I_2,…,I_m} be an itemset. Let D, the task-relevant data, be a set of a database transaction where each transaction T is a nonempty itemset such that T ⊆ I. Each transaction is associated with an identifier, called a TID. Let A be a set of items. A transaction T is said to contain A if A⊆ T .An association rule is an implication of the form A⇒B. where A ⊂ I, B ⊂ I, A ≠ Φ, B≠ Φ and A∩B= Φ. The rule A⇒B holds in the transaction set D with support s, where s is the percentage of transactions in D that contain A ∪ B (i.e., the union of sets A and B say, or, both A and B). This is taken to be the probability, P(A ∪ B). The rule A⇒B has confidence c in the transaction set D, where c is the percentage of transactions in D containing A that also contain B. This is taken to be conditional probability, P(B|A). That is

$$\text{Support } (A \Rightarrow B) = P(A \cup B)$$
$$\text{Confidence } (A \Rightarrow B) = P(B|A)$$

The association rule mining can be viewed as a two-step process:
1. Find all frequent itemsets: Each of these itemsets will occur at least as frequently as predetermined minimum support count.
2. Generate strong association rules from the frequent itemsets: The rules must satisfy minimum support and confidence. These rules are called strong rules[6]

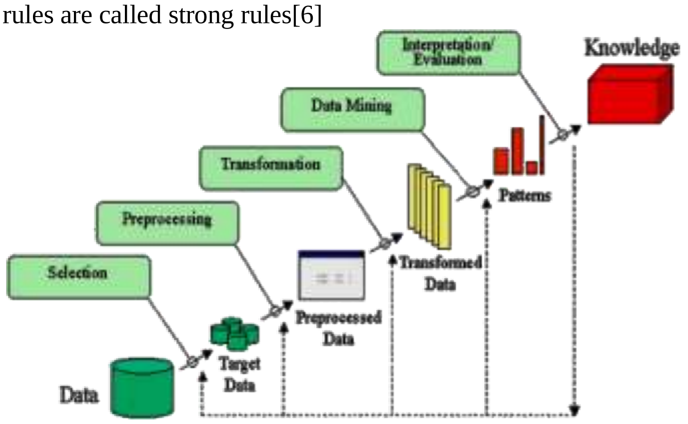

Fig. 1 The steps of Knowledge Discovery Process[8]

B. Apriori Algorithm

The algorithm starts by collecting all the frequent 1-itemsets in the first pass based on the minimum support. It uses this set (called L1) to generate the candidate sets to be frequent in the next pass (called C2) by joining L1 with itself. Any item that is in C1 and not in L1 is eliminated from C2. This is achieved by calling a function called 'apriori-gen'. This reduces the item size drastically. The algorithm continues in the same way to generate the Ck, of size k from the large itemsets of k-1, then reduces the candidate set by eliminating all those items in k-1 with support count less than minimum support. The algorithm terminates when there are no candidates to be counted in the next pass.[3]

C. Data Collection

The data required for this study is taken from MCA Programme of Institute of Management, Kolhapur: A constituent unit of Bharati Vidyapeeth University, Pune. This paper's purpose is to study reasons of high failure rate in MCAII class in Computer Networks (CN) subject. The data contains Marks obtained by the student in CN, his/her University from which Graduation is completed and entry type i.e. Normal (From MCA 1st year) and Lateral (Direct MCA 2nd year). In this research the association rule mining analysis is performed based on student's failed course. This identifies hidden relationship between the failed course and suggests relevant causes of the failure to improve the low capacity students' performance.

D. Data Transformation:

From academic year 2011-12 BVU has adopted Choice Based Credit System (CBCS). The grading system that is being followed by the University is The 10 point Grading System which is shown in Table I

Range Percentage Marks	Grade points	Grade
[75, 100]	10.0	O
[70, 74.9]	9.0	A+
[65, 69.9]	8.0	A
[60, 64.9]	7.0	B+
[55,59.9]	6.0	B
[50,54.9]	5.5	C+
[45, 49.9]	5.0	C
[40,44.9]	4.5	D
[00, 39.9]	0.0	F

TABLE I: The 10-Point Grading System

IV. GENERATING AND ANALYZING ASSOCIATION RULE

In this paper we have used WEKA, a data mining tool for association rule mining. The snapshot of actual data that is utilized to generate the rules is shown in Table II:

A. WEKA As A Data Miner Tool

WEKA stands for Waikato Environment for Knowledge Analysis, is a computer program that was developed at the University of Waikato in New Zealand. This application is completely written in JAVA and is compatible with almost every computing platform. Weka contains data mining tools for data pre-processing, classification, association, clustering, regression, association rules and visualization and provide outstanding results. Data set can be loaded in Weka in ARFF (Attribute relation File Format) format. We can also convert CSV (comma delimited) format into ARFF format.[13]

B. Data Analysis

Table III indicates frequency of grades obtained by the students. The instance size is 67.

Grade	Count
F	31
C	8
C+	2
B	2
D	24

TABLE III

C. Result Analysis using WEKA

Fig. 2 is the snapshots showing different Association Rules with various Support and Confidence factors

Lateral	CN	University
N	F	BVU
N	C	OTH
N	F	OTH
N	C	BVU
N	C+	OTH
N	B	OTH
N	B+	OTH
N	C	BVU
N	C+	OTH
N	B	OTH
N	B+	BVU
N	B+	OTH
N	B+	OTH
N	F	BVU
N	A	BVU
N	D	OTH
N	D	OTH
N	D	OTH
N	F	OTH
N	D	OTH
Y	D	OTH
Y	C	OTH
Y	C	OTH
Y	D	OTH
Y	D	OTH
Y	F	BVU
Y	D	OTH
Y	D	OTH
Y	B	BVU

TABLE II: Data Preprocessed (transformed) to contain Grades

In the above test we come to know the following associations. For all rule we applied the min. support 0.2 and min. confidence 0.4. Size of large itemsets L_1= 6 and L_2=9. By observing rule 2, we can say 24 students have obtained D grade and they belongs to other university. Means 100% D grade students belongs to other university with confidence 0.71. Rule 5 tells us that 14 students are in fail grade out of 24 students who have completed their graduation form BVU with confidence level 0.56.

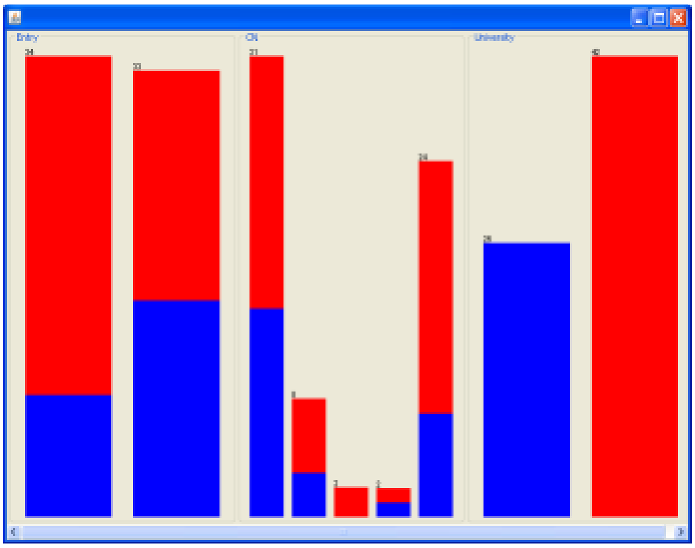

Fig 2 Snapshot of preprocessing in Weka

From all above rules we can say that the lateral entry students and other university students have scored very less as compared to BVU students and Normal students attending classes from MCA 1. So we suggest the authorities to have satisfactory performance of these students and to increase their understanding level, bridge courses and remedial classes should be assigned so that the result will improve. The generated association rules are of great help for the curriculum planners and academic mangers. The knowledge discovery from the above study would not only benefit students but the concerned academic institute in improving the quality of their students so that they can be better placed in their jobs which indirectly will help the institute in better intake of the students

V. CONCLUSION

In this paper we have analyzed the use of a data mining technique called association rule mining to study the quality of students' performances at Post Graduation level. The mined association rules exposed us various factors like student's Home University, Admission type, Grades obtained by students who have failed to attain a satisfactory level of performance in the Post Graduation level. This study has bridge the gap in educational data analysis and shows the use of the association rule mining algorithm for enhancing the effectiveness of academic planners in higher institutions.

ACKNOWLEDGMENT

We are thankful to Prof. A A Basade, Professor of Statistics, BVU, Pune for taking efforts for handling the statistical issues. Also we thankful to Dr. Nitin Nayak for supporting and motivating us to do research.

REFERENCES

[1] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art", *IEEE Transactions onSystems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol 40., No. 6, November 2010.

[2] M. A. Anwar and Naseer Ahmed, "Knowledge Mining in Supervised and Unsupervised Assessment Data of Students' Performance", *2011 2nd International Conference on Networking and Information Technology IPCSIT vol.17 (2011)*.

[3] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth. "The KDD process for extracting useful knowledge from volumes of data", *CACM* 39 (11), pp. 27-34, 1996.

[4] Fadzilah Siraj and Mansour Ali Abdoulha, " Uncovering hidden Information within University's Student Enrollment Data using Data Mining", *Third Asia InternationalConference on Modelling and Simulation*, 2009.

[5] Fangjun Wu,"Apply Data Mining to student's choosing Teachers under complete Credit Hour", *Second International Workshop on Education Technology and Computer Science*, 2010.

[6] (Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar and M. Inayat Khan, "Data Mining Model for Higher Education System", *Europen Journal of Scientific Research*, Vol.43, No.1, pp.24-29, 2010

[7] W.M.R. Tissera, R.I. Athauda and H. C. Fernando "Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining", *IEEE International Conference on Information Acquisition*, 2006*FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.

[8] Han Jiawei and Micheline Kamber, *Data Mining: Concepts and Technique*, Morgan Kaufmann Publishers, 2000

[9] Hongjie Sun, "Research on Student Learning ResultSystem based on Data Mining", *IJCSNS International Journal of Computer Science and Network Security*, Vol.10, No. 4, April 2010.

[10] Ceglar, J.F Roddick. "Association mining". *ACM Computing Surveys*, 38:2, pp. 1-42, 2006.

[11] S. Kotsiantis, , D. Kanellopoulos. "Association Rules Mining" *A Recent Overview.GESTS Int. Transactions on Computer Science and Engineering*, Vol. 32 (1), pp. 71-82, 2006

[12] Dr. Varun Kumar and Anupama Chadha, "Mining Association Rules in Students' Assessment Data", *International Journal of Computer Science Issues*, Vol 9, Issue 5, September 2012.

[13] www.wikipedia.org

[14] http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html

[15] www.google.com

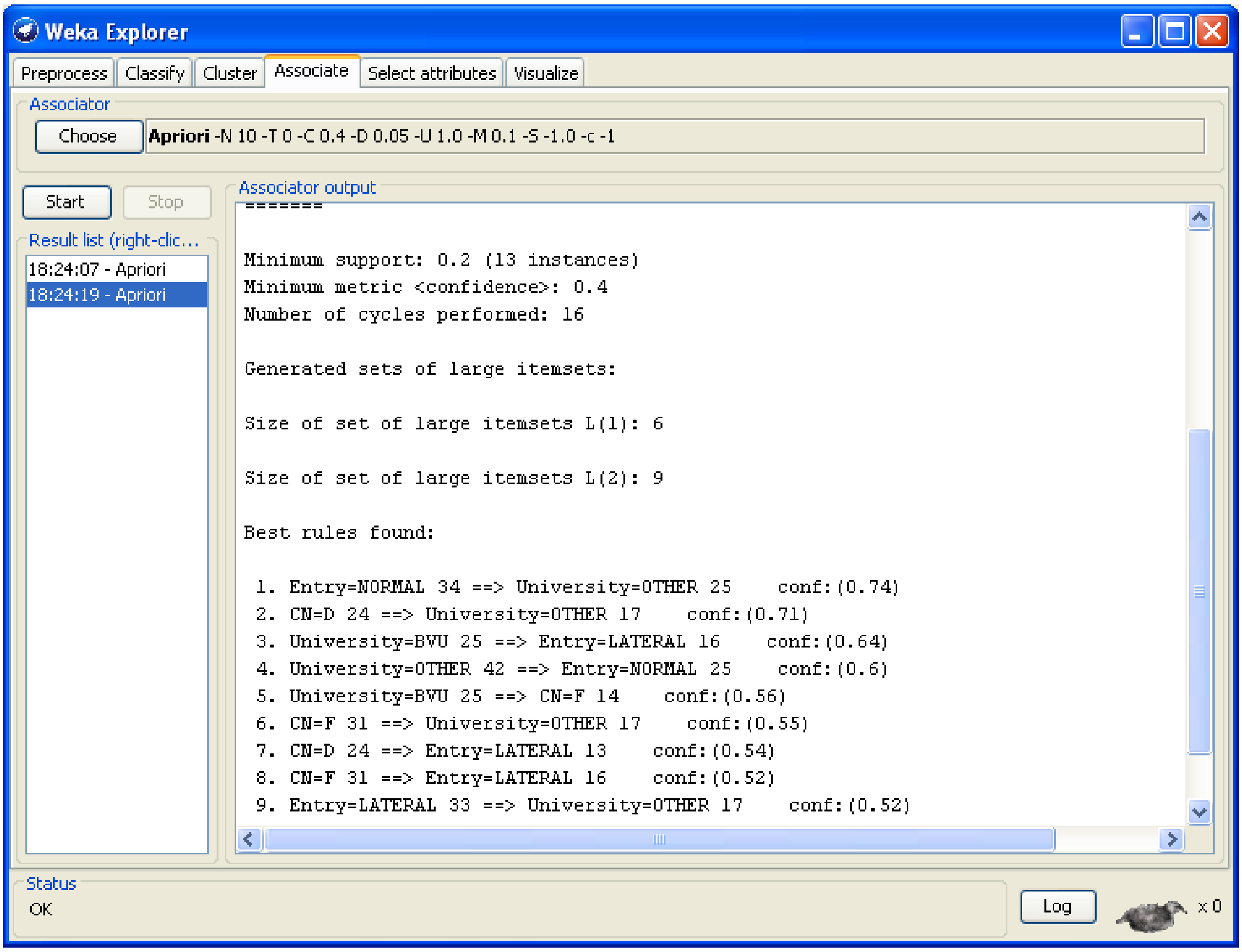

Fig. 3 Snapshot of Association Rule Mined