



Review on Knowledge Discovery from Academic Database using Association Rule Mining

Mahendra Sahu¹, Narendra Dhawade¹, Smita Bagde¹, Rubina Sheikh¹, Poorvi Khare²

B.E, Dept. of CSE, Nagpur Institute of Technology, Mahurzari, Nagpur, Nagpur University, Maharashtra, India¹

Asst. Prof, Dept. of CSE, Nagpur Institute of Technology, Mahurzari, Nagpur, Nagpur University, Maharashtra, India²

ABSTRACT: Discovering the hidden knowledge from large volume of student database and applying it properly for decision making is essential for ensuring high quality education in any academic institution. This knowledge is extractable through data mining techniques. Association Rule Mining technique aims at discovering implicative tendencies that can provide valuable information for the decision maker. In this project, we are going to present an applied research on mining Association Rule using academic data of a university and use it for syllabus design. We are going to discover knowledge regarding the academic performance and personal statistics of students. And we will develop a technique to transform the existing relational database for student's academic performance into a universal database format using academic and personal data of a student. After that we will transform the universal format into a modified format for suitability of using Association Rule mining algorithm. We will use FP Growth algorithm for finding interested association rules from the transformed database which can be useful to extract knowledge of student's academic progress, decay in their potentiality, abandonment as well as retention of students. The impact of courses and curriculum and teaching methodologies are also found from the extracted knowledge which is beneficial for any institution of higher education.

KEYWORDS: Data Mining, Association Rule Mining, Fp-growth Algorithm, Knowledge Discovery, reasons for weak performance.

I. INTRODUCTION

Students are one of the fundamental elements of any academic institution. Indeed, the prime concern for an educational institution is to ensure qualified technical foundation, scholarly guidance and high standard education to all of its students. For a large educational institute like public university which generates large volume of data, it requires an efficient way to apply data mining techniques for obtaining knowledge on the development and performance improvement of academic activities. The knowledge acquired from the institutional database will be sufficient to look for answers to such questions as: Which factors determine better or worse academic performance of students? What are the causes behind the students' retention in the university? Why do students abandon from an educational institute? Concepts and techniques of data mining are essential to discover the hidden knowledge from large datasets.

Many universities of India enrolls the top most brilliant 1000 students selected by a competitive examination among many students competing higher secondary education. Among these 1000 students, top ranked students can get admission into the various departments like Computer Science and Engineering (CSE), Electronics Engineering (EC), Electrical Engineering (EE), Mechanical Engineering(M.E.). Although, the departments possesses most of the brightest students, statistics demonstrates that performance of some students degrades noticeably. Moreover the problem of retention as well as abandonment is also prevalent among the students. The main objective of this research study is to get knowledge of those students' academic progress, degradation of their merit, abandonment as well as retention of them. In this paper, we are going to developed a technique to discover knowledge using Association Rule Mining from institutional data of students who have completed their undergraduate in the give respective departments /



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Here, we are going to transform the existing relational database for students' academic performance into a universal database format using academic and personal data of a student. After that, we will further transform the universal format into a modified format for suitability of using Association Rule Mining algorithm. We are going to discover interested rules which interpret several important facts related to students' academic performance e.g., impact of personal information of students such that gender, residence etc. and the impact of course contents and pedagogy. We will also determine the impact of retention for particular courses. Addressing the abandonment issue, we have categorized the students based on their personal information who could not complete their graduation. Finally, we will present a guideline to apply the extracted knowledge to improve the academic performance and to make an optimization between abandonment and retention.

II. RELATED WORK

Techniques of Educational Data Mining (EDM) have been used to resolve educational research issues since 1993 [3]. Mining educational data through classification is an effective way to analyze students' performance from the extracted knowledge [4]. Automatic Clustering and decision rule data mining techniques are also applied for knowledge discovery based on academic data analysis [5]. It has been shown that how data mining algorithms can help discovering all possible relevant knowledge contained in databases obtained from Web-based educational systems [6]. Besides these, there is also an idea presented in the research [7] about the factors depended on students' performance and how Naïve Bayes classifier can be applied to calculate probabilities so that the final examination results can be predicted based on the findings

Proposed algorithm

A. Design Considerations:

- Select database
- Apply algorithm

```
(1)IF Tree contains a single path P
(2)THEN FOR EACH combination (denoted as) of the nodes in the path P
DO
(3)generate pattern[with support=minimum support of nodes in ;
(4)ELSE F
OR EACH ai
in the header of Tree
DO
(5)generate pattern=ai
[with support =ai
:support;
(6)Construct's conditional pattern base and then's conditional FP-tree Tree ;
(7)IF Tree 6 = ;
(8) THEN Call FP-gro
With(Tree;)
```

eq. (3)

II. PSEUDO CODE

```
(01) if Tree contains a single prefix path then // Mining single prefix-path FP-tree {
(02) let P be the single prefix-path part of Tree;
(03) let Q be the multipath part with the top branching node replaced by a null root;
(04) for each combination (denoted as  $\beta$ ) of the nodes in the path P do
(05) generate pattern  $\beta \cup a$  with support = minimum support of nodes in  $\beta$ ;
(06) let freq pattern set(P) be the set of patterns so generated;
```



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

```
}
(07) else let Q be Tree;
(08) for each item ai in Q do { // Mining multipath FP-tree
(09) generate pattern  $\beta = ai \cup a$  with support = ai .support;
(10) construct  $\beta$ 's conditional pattern-base and then  $\beta$ 's conditional FP-tree Tree  $\beta$ ;
(11) if Tree  $\beta \neq \emptyset$  then
(12) call FP-growth(Tree  $\beta$  ,  $\beta$ );
(13) let freq pattern set(Q) be the set of patterns so generated;
}
(14) return(freq pattern set(P)  $\cup$  freq pattern set(Q)  $\cup$  (freq pattern set(P)  $\times$  freq pattern set(Q)))
}
```

III. SIMULATION RESULTS

The analysis tool in which subject degradation declare. Our results shows that the metric total transmission energy performs better than the maximum number of hops in terms of network lifetime, energy consumption and total number of packets transmitted through the network.

IV. CONCLUSION AND FUTURE WORK

Knowledge Discovery from academic database is very useful to improve the academic performance of any education Institution. In the proposed work we have tried to overcome the challenges.

REFERENCES

- [1] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art", IEEE Transactions on Systems, 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
- [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 International Conference on Modelling and Simulation, 2009
- [5] Anurag Choubey, Dr. Ravindra Patel, Dr. J.L. Rana, "GRAPH BASED NEW APPROACH FOR FREQUENT PATTERN MINING". (IJCSIT) Vol 4, No 1, Feb 2012
- [6] Shaeela Aysha, Tasleem Mustafa, Ahsan Raza Sattar and M. Inayat Khan, "Data Mining Model for Higher Education System", European 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
- [8] Han Jiawei and Micheline Kamber, Data Mining: Concepts and Technique, Morgan Kaufmann Publishers, 2000
- [9] Hongjie Sun, "Research on Student Learning Result System based on Data Mining", IJCSNS International Journal of Computer Science and Network Security, Vol.10, No. 4, April 2010
- [10] Qasem A. Al-Radaideh, Emad M. Al-Shawakfa, Mustafa I. Al-Najjar, "Mining Student Data Using Decision Trees", ACIT' 2006: The International Arab Conference on Information Technology
- [11] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques", 3rd ed. San Francisco: Morgan Kaufmann Publishers, 2011.
- [12] (2014, Jun.) The Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology: General Information. [Online]. Available: <http://www.buet.ac.bd/cse/geninfo/index.php>
- [13] A.S.M. L. Hoque, R. Paul, S. Ahmed, "Preprocessing of Academic Data for Mining Association Rule," In Proceedings of the Workshop on Advances in Data Management: Applications and Algorithms (WADM), 2013.
- [14] B.K. Baradwaj, and S. Pal, "Mining Educational Data to Analyze Students' Performance." International Journal of Advanced Computer Science and Applications (IJACSA), 2(6), pp. 63-69, 2011.
- [15] V. Kumar, and V. Sharma, "Students-Examination-Result-Mining-A-Predictive-Approach." International Journal of Advanced Computer Science and Applications (IJASER), 3(11), Nov. 2012

BIOGRAPHY

Prof. Poorvi Khare is an assistant professor in Nagpur Institute of Technology .

Miss. Smita Bagde is student of final year studing in Nagpur Institute of Technology.

Miss. Rubina Shiekh is student of final year studing in Nagpur Institute of Technology.

Mr. Mahendra Sahu is student of final year studing in Nagpur Institute of Technology.

Mr. Narendra Dhawade is student of final year studing in Nagpur Institute of Technology.