RESEARCH ARTICLE OPEN ACCESS

Comparative Study of various association rule-mining algorithms applied with special reference to Educational Data Mining

Ms. Juhi Arora, Dr. ManaliKshirsagar
(Computer Science & engineering, Rajiv Gandhi College of Engineering, Nagpur.
Email: ajuhi18@gmail.com)
(Computer Science & engineering, Rajiv Gandhi College of Engineering, Nagpur
Email : manali_kshirsagar@yahoo.com)

Abstract:

Data is important property for everyone. Large amount of data is available in the world. Educational Data Mining helps in understanding the trends, inclination of students as well as their performance. This analysis can help in adapting and designing student cerbic teaching-learning methods. In this paper, we have studied what is Educational data mining, various techniques of EDM and methods used for association-rule mining algorithm. Algorithms are compared based on some performance factors like accuracy, data support, execution speed, etc.

Keywords — Data Mining (DM), Educational data mining (EDM), Educational systems, Association rule-based algorithm.

I. INTRODUCTION

Data mining is the computing process of discovering patterns in large data set involving methods at the intersection of machine learning, statistics, and database system. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.

Educational data mining (EDM) is the latest advancement in education field, dealing with the development of techniques for exploring and analysing the huge data that come from the educational database. This involves analysing the results in depth in order to monitor the student's academic activities closely. Competition between the institutions is an all-time high forcing management to focus on increasing enrolment and registration whilst controlling costs. Data mining, which is defined as the process of extracting previously unknown knowledge, and detecting the interesting patterns from a massive set of data, offers a way of dealing with this problem in educational institutions.

In most of the colleges several assessments are prepared to evaluate student performance. Types of assessments that are common to all the colleges and institutions are assignments, sessional, and practical. With the help of sessional marks, assignments and practical the professor assesses the performance of students but this become the very tedious task and need to maintain the records. However, it also takes time to do these tasks because they need to formulate their own calculation for evaluation of result. These tasks will become much easier if all calculations for class position (ranking) can be automatically done when the teacher enter assignments, practical and exams marks. Students result is the criteria for the

measurement of the student's capability in terms of academic performance in the any academics. It is used to measure a student's capability in each subject offered by the student, because the student result is very important to the student and his/her parents, there is need to avoid common mistakes made during the result processing, this can easily be achieved by an automated result processing system, The system is an effective, efficient and error free results processing system designed and implemented for any college or institution.

II.LITRATURE SURVY

Nur Hani Zulkifli, Jamalahyahaya, Aziz Deraman has proposed the empirical study which has been done to discover requirement analysis process in Business Intelligence and Analytics implementation. The approach involves interviewing business intelligence experts in various domains. The proposed model can be used as a guideline for practitioners in analyzing requirements for BIAPM[1].

Mohiuddin Ali Khan, WajeGharibi, Sateesh Kumar Pradhan has proposed the analysis of real data collected from several institutions, a sample study has been conducted to show how the Apriori Algorithm can be used in educational field and the results have been observed. Data mining can be used effectively in educational institutes for leading education activities in an effective way, for watching students' performances continuously and directing students in course and profession choosing. Thus, the level of student's success can be raised, and we can concentrate on students weakness at selected courses thus improving the standard of students[2].

Leena Khanna,Dr.Shailendra NarayanSingh, Dr. MansafAlam has proposed the systematic review on EducationalData Mining technique which helps in predicting thefuture and changing the future. It containsapplicationareas of EDM,various methods of EDMand the factors affecting the Student's academicperformance and the teaching learning process[3].

Oswaldo Moscoso-Zea1, Andres-Sampedro, and Sergio Luján-Mora has proposed the differences between existing methodologies for DW design. Looking at differences between a business DW design and an educational DW design we found out that the value chain is the most important characteristic that changes the vision of the implementation project. Kimball and Inmon are the two most popular and recommended methodologies in the literature[4].

Lida Shams, Hassan Rashidi has proposed the BI architecture based on the service oriented concept (i.e. the flexible services that can be accessed on demand) is used for analyzing the current situation of educational courses and predicting their near future in order to support decision making. Thus the proposed SOA for BI considerably improves the decision making process[5].

Mohammed I.Al. Twijri, Amin Y. Noaman has proposed a new Data Mining model to be applied in higher education institutions. The Suggested model assists in decision modeling process in the strategic levels of higher institutions as well as regulates the disciplines of student's admission[6].

Ms.TismyDevasia ,Ms.Vinushree T P, Mr.VinayakHegde has proposed the web based application which makes use of the Navie Bayesian mining technique for the extraction of useful information. The system aims at increasing the success graph of students using Naive Bayesian and the system which maintains all student admission details, course details, subject details, student marks details, attendance details, etc[7].

III. EDM SURVEY OF PAPERS

Name	Source of publication	year	Description
A Review on Predicting Student's Performance using Data Mining Techniques.	Conference	2015	Study distillates the gaps in existing prediction methods, variables that define the academic performance and the prediction methods that can be used to

			determine the students' performance. Techniques utilized in previous researches : Neural Network, decision Tree, SVM, K.NN, Naïve Bayes with Neural Network having the highest accuracy and Classification method the most frequent in EDM.
Modelling Key parameters in Higher Education using Logistic Regression: an Indian case based Data Analysis	Conference	2015	In this study the author developed an effective model to predict the effect of various identified parameters in CGPA of the students using Logistic Regression. Parameters like Academic background, Family Closeness and Freedom to make Choices etc does have an impact on the students overall CGPA. The model could not make a distinction between the Low and Medium CGPA and High and medium CGPA.
Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue	Journal	2015	Comprehensive literature review of researches in EDM from 2002 to 2014 with main focus on papers published in Educational Data Mining, Predicting Academic Performance with Pre/Post Enrolment Factors, DM techniques to predict learning performance, association between Pre/Post Enrolment Factors and Employability, DM applications in supplementary areas of education like Faculty Evaluation, Analysis of the Curriculum Chosen by Students etc. Need of Integrated Models/Frameworks for all the stakeholders to ensure Sustainable Growth
Educational Data Mining and its role in Educational field	Journal	2014	This paper highlights Evaluation of Students and Teachers performances using different Data mining methods. Attributes considered in the

			paper are: Attendance, Assignment, GPA, Sessional marks and final grade. The main limitation of this study is that limited attributes are studied.
Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence	Journal	2014	Primary research objective included identification of factors such as grades, demographic characteristics, students' portfolios, students' participation and enrolment in various activities, multimodal skills, students' mood etc. and evaluating the affect of these parameters on his/her performance.
Educational data mining: A survey and a data mining based analysis of recent works	Journal	2014	Main areas of EDM focussed in this paper is Students modelling approaches which includes behaviour, performance, assessment, student support and feedback, domain knowledge, sequencing, and teachers support etc.. Analysis and discovery of useful patterns from these approaches identified.
Educational data mining applications	Journal	2014	Broad areas of application in which EDM can be applied to e-learning are discussed. Major application areas include: User Modelling, User Knowledge Modelling, User Behaviour Modelling , User Profiling, Domain Modelling, Trend Analysis etc.
Educational data mining: A review	Conference	2013	Targeted 9 latest studies to which the data mining methods are applied in educational setting ranging from the year 2004 until 2012.
Techniques focused : Classification, Prediction, Association Rule Analysis, Sequential Pattern Increasin	Journal	2013	Data mining is necessary in organizations to enhance competitive advantage and decision making. Educational data Mining helps to analyze the educational data and developing model for

g Quality of Education Using Educational Data Mining			improving institutional effectiveness. Educational Data Mining is a key area in mining students performance and helpful in predicting educational institutions performance taking parameters like Teaching Skills, Course content, Infrastructure etc.
A Survey and Future Vision of Data mining in Educational Field	Conference	2012	This paper focuses on importance of need of data mining in various Educational System which includes: traditional , web-based , e-learning and intelligent tutoring system etc. This paper describes the use of various techniques such as prediction and classification, association rule mining, clustering, and social area networking to educational data.
Data Mining for Education	Journal	2012	Classification and prediction, relationship mining, clustering, discovery with models and distillation of data for human judgment are the main techniques uses in EDM with the fourth and fifth kind having prominence within educational data mining. Author discussed the main application areas of EDM which includes enhancing the existing student models that present comprehensive information about a student's individuality, discovering or improving the domain models, studying the pedagogical provision stipulated by education software, systematic investigation of learners and their learning.
Educational Data Mining: A Review of the State of the Art	Journal	2010	Peer review of 15 years from 1993 to 2009 which focuses on EDM an upcoming research area which take account of various fields like data mining, learning analytics,

			e-learning, adaptive hypermedia, intelligent tutoring systems, web mining etc.

V. IV. TECHNIQUES USED IN EDUCATIONAL DATA MINING

Some of the important techniques used in EDM are discussed below:

a) Linear Regression: this Predication technique predicts a number. Various attributes like income, sales, age, weight etc. can be predicted using Regression. For example in Educational Data Mining to predict the CGPA of Student various parameters like Age, Gender, Attendance, Family Income, Occupation, Family Qualification can be used as predictors.

b) Clustering: Clustering is the method that helps to group similar records together. This is unsupervised learning approach, which mainly focuses on high-dimensional data. Clustering can be Hierarchical or Non-Hierarchical. K-Means is the most common method of Clustering. For example in Educational Data Mining, clustering can be used to group students based on their Learning styles as Visual, Aural, and Kinesthetic. Cluster model has a disadvantage that there are no clear rules to characterize each cluster.

c) Classification: Classification helps in classifying data based on the Training set and then uses that pattern to classify the new data which is also known as the training set. This is a supervisedlearning technique because the classes are predefined before extracting patterns on the target data. Some popular Classification methods used in EDM are Decision Tree, Neural Network, Naive Bayes Classifier, SVM etc.

d) Association Rule Mining: Association rules in Educational Data Mining are used to determine remarkable and strong association rules from Educational databases using support and confidence as the predefined measures. Association rules further identify which parameters are closely related to each other and have relations among them.

Apriori algorithm is one of most common and widely used methods in Association Mining. Association rules are if/then statements that help to uncover relationships between unrelated data in a database, relational database or other information repository. Association rules are used to find the relationships between the objects which are frequently used together. Applications of association rules are basket data analysis, classification ,cross-marketing, clustering, catalog design, and loss-leader analysis etc.

V.COMPARSION OF ASSOCIATION RULE MINING ALGORITHM

Characteristics	AIS	SETM	Apriori	Aprioritid	Apriori hybrid	FP-Growth
Data Support	Less	Less	Limited	Often suppose large	Very large	Very large
Speed in initial phase	Slow	Slow	High	Slow	High	High
Speed in later phase	Slow	Slow	Slow	High	High	High
Accuracy	Very less	Less	Less	More accurate than Apriori	More accurate than Aprioritid	More accurate

Apriori algorithm requires large amount of memory space due to large number of candidates are generated. It need multiple scans for generating candidate sets. In this algorithm execution time is more as time is wasted in producing candidates set every time. Hence, we are using FP-Growth algorithm which is particularly more efficient than apriori algorithm when it comes to long patterns.

FP (Frequent Pattern) Growth Algorithm

The FP-Growth algorithm, proposed by Han. In his study, Han proved that his method outperforms other popular methods for mining frequent patterns, e.g. the apriori algorithm and the tree projection.

The FP-Growth algorithm is an alternative way to find frequent itemsets without using candidate generations, thus improving performance. For so much it uses a divide-and-conquer strategy. The core of this methods is the usage of a special data structure named frequent-pattern tree (FP tree), which retains the itemset association information.

The algorithm extracts frequent item sets that can be extract association rules. This is done using the support of an item set.

By using divide-and-conquer strategy, it compress the database which provides the frequent sets; then divide this compressed database into a set of conditional databases, each associated with a frequent set and apply data mining on each database. This algorithm performs mining on FP-tree recursively. There is a problem of finding frequent itemsets which is converted to searching and constructing trees recursively.

The key points of FP-Growth algorithm are:

1. FP growth improves Apriority to a big extent
2. Frequent Item set Mining is possible without candidate generation

Two step approach used in this algorithm:

1: Build a compact data structure called the FP-tree. Built using 2 passes over the data set.

2: Extracts frequent itemsets directly from the FP-tree

FP-Tree is constructed using 2 passes over the data-set:

Pass 1

1.Scan data and find support for each item.

2. Discard infrequent items.

3. Sort frequent items in decreasing order based on their support.

4. Use this order when building the FP-Tree, so common prefixes can be shared.

Pass 2:

Nodes correspond to items and have a counter

1. 1. FP-Growth reads 1 transaction at a time maps it to a path
2. 2. Fixed order is used, so paths can overlap when transactions share items (when they have the same prefix).

– -In this case, counters are incremented

– 3. Pointers are maintained between nodes containing the same item, creating singly linked lists

– -The more paths that overlap, the higher the compression. FP-tree may fit in memory.

3. 4. Frequent itemsets extracted from the FP-Tree.

VI. CONCLUSION

There are various association rule-mining algorithms. In this paper, we have discussed the association rule mining algorithms used in education system. Educational Data Mining can be used effectively in institutes for learning education activities in an effective way, for monitoring student's performances continuously. Comparison is done based on the above performance criteria. Each algorithm has some advantages and disadvantages. From the above comparison, we can conclude that, FP-growth is more suitable than all other algorithms discussed here in terms of large data set support and speed.

REFERENCES

[1] Nur Hani Zulkifli, Jamalahyahaya, Aziz Deraman (Article August2016), "Business Intelligence and Analytics in Managing Organizational Performance: The Requirement Analysis", Journal of Advances in Information Technology Vol. 7, No.3,August 2016.

[2] Mohiuddin Ali Khan, WajebGharibi, Sateesh Kumar Pradhan (2014) "Data Mining Techniques for Business Intelligence in Educational System: A CaseMining"978-14799-3351 8/14/$31.00,2014 IEEE.

[3] Leena Khanna , Dr.ShailendraNarayan Singh,Dr.MansafAlam , "Educational Data Mining and its Role in Determining Factors Affecting Student AcademicPerformance: A Systematic Review, 978 1-4673-6984/8/16/$31.00 © 2016 IEEE.[4]Oswaldo Moscoso-Zea1, Andres Sampedro,andSergio Luján-Mora (2016),"Data warehouse design for Educational Data Mining", International Conference on Enterprise Information System (ICEIS)2016, pp. p. 244–249.

[5] Lida Shams, Hassan Rashidi (2016), "Factors affecting the Increase and Decrease student achievement in Primary School with Business Intelligence Approach", International Journal of Computer & Information Technologies(IJOCIT).

[6] MohammedI .Al.Twijri, Amin Y. Noaman, "A New Data Mining Model Adopted for Higher Institutions", Procedia Computer Science 65 (2015).

[7] Ms.TismyDevasia, Ms.Vinushree T P, Mr.VinayakHegde, "Prediction of Students Performance using Educational Data Mining", International Symposium on Educational Technology (ISET),2015 IEEE.

[8]Amjad Abu Saa, "Educational Data Mining & Students' Performance Prediction", International Journal of Advanced Computer Science and Applications, Vol.7, No.5, 2016.

[9] S. M. Merchan and J. A. Duarte, "Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic Performance", IEEE Latin America Transactions, Vol. 14, No. 6, June 2016.

[10] Trupti A. Kumbhare, Prof. Santosh V. Chobe, "An Overview of Association RuleMining Algorithms", IJCSIT, Vol. 5(1), 2014, 927-930.

[11] 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, pp. 601–618, Nov. 2010.

[12] Thakar, Pooja. "Performance Analysis and Prediction in Educational Data Mining: A ResearchTravelogue." arXiv preprint arXiv: 1509.05176 (2015).

[13] B. Kumar and S. Pal, "Mining educational data to analyse students' performance," International Journal of Advanced Computer Science and Applications, Vol.7, No.5, 2016.

[14] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," Procedia Computer Science, vol. 72, pp. 414–422, 2015.

[15] P. Baepler and C. Murdoch, "Academic Analytics and data mining in higher education," International Journal for the Scholarship of Teaching and Learning, vol. 4, no. 2, Jul. 2010.

[16] S. Harikumar, "A study on educational data mining," International Journal of Computer Trends and Technology, vol. 8, no. 2, pp. 90– 95, Feb. 2014.

[17] M. Berland, R. S. Baker, and P. Blikstein, "Educational data mining and learning Analytics: Applications to Constructionist research," Technology, Knowledge and Learning, vol. 19, no. 1-2, pp. 205–220, May 2014.

[18] M. S. Bhullar and A. Kaur. "Use of data mining in education sector". Lecture Notes in Engineering and Computer Science 2200(1), pp. 513-516. 2012.

[19] M. Mayilvaganan,D. Kalpanadevi ," Comparison of Classification Techniques for predicting the performance of Students Academic Environment," in International Conference on Communication and Network Technologies (ICCNT), 2014.

[20] BipinBihariJayasingh, "A Data Mining Approach to Inquiry Based Inductive Learning Practice In Engineering Education", in IEEE 6th International Conference on Advanced Computing 2016.