

DISCERNING LEARNER'S ERUDITION USING DATA MINING TECHNIQUES

S. Anupama Kumar¹ and Dr. Vijayalakshmi M.N²

¹ Research Scholar, PRIST University, ¹ Assistant Professor,
kumaranu.0506@gmail.com

² Associate Professor, Dept of MCA,
Dept of M.C.A. ^{1,2} R.V.College of Engineering, Bangalore
mnviju74@gmail.com

ABSTRACT

Association Rule mining can be used in various areas of education data to bring out the interesting rules about the learner's records. It can be used to bring out the hidden facts in understanding the behavior of the learner in a learning environment, learning style, examination pattern and assessment etc. In this paper Apriori and Tertius rule mining algorithms are used to understand the behavior and attitude of the students towards diverse learning styles. Both the algorithms revealed interesting rules depending on the confidence factor of the dataset. Item sets were generated using the Apriori algorithm giving best rules. Using Tertius algorithm, the most excellent rules were generated based on the number of hypothesis, confidence value, true and false positive rates.

KEYWORDS

Learning styles, EDM, Apriori, Tertius

1. INTRODUCTION

Within a decade, the Internet has become a pervasive medium that has changed the education system completely. The way through which information and knowledge are transmitted and shared throughout the world has undergone tremendous change in the new era of technology. The learning style has not limited to auditory learners but has grown towards visual and tactile learners. Students in primary and secondary education are also participating in e-learning systems and are sometimes ahead of trainers. Data Mining can be used to extract knowledge from educational domains through the analysis of the information available in the form of data generated by their users. The main objective of applying the techniques is to find the patterns of system usage by teachers and students and discovering the students' learning behaviour patterns. Association analysis can be applied on the education data to find out the interesting rules that are present in the data and bring out new knowledge out of it. These rules can be utilised by the educator to understand the need of the learner and improve the learning skills. The paper is organized into four sections. Section II explains the background work, section III describes the application of Apriori and Tertius algorithm on the data set and section IV discusses the result and conclusion.

2. BACK GROUND WORK

There are three main learning styles in which a student can be categorized. If a teacher is capable of understanding his/her learning style it becomes easy to employ the skills and increase the information that can be learned. It also helps in increasing the amount of information to be learned.

2.1. Visual Learners

Visual learners learn new skills by seeing and learn new information through visuals. They learn and absorb information through the use of hand-outs, charts and visual information. The visual learner take notes when visual information is being presented, and create their own charts and graphics while learning. This type of learning style is popularly used in primary and secondary education rather than higher education.

2.2. Auditory Learners

Auditory learners learn through the use of verbal communication and lectures. The information may be revealed to them through written work but they listen only through verbal communication. They listen to the tone of the voice, the pitch of the voice and the speed of the voice. They get the benefit from creating recordings of the information taught. This method was best used in the previous generation by our ancestors than the current era. This is also used by blind people.

2.3. Tacitile Learners

Some people learn better by completing the actual. These people are referred to as “tactile” or “kinesthetic” learners. They learn only when have a hands on experience. Tactile learners may suffer from a short attention span which can make it difficult to learn new skills unless the hands-on approach is taken.

It is very essential for a facilitator to understand the learning method of the student so as to make them learn better. It reduces the amount of reviewing and study time using methods of learning that are not compatible with learning ability.

In [1] the author has described a hybrid approach which uses EDM and regression analysis to analyse live video streaming (LVS) students' online learning behaviours and their performance in their courses. Students' participation and login frequency, as well as the number of chat messages and questions that they submit to their instructors, were analysed, along with students' final grades. Oladipupo O.O. and et al [2] have applied association rule mining to predict the student's failure in the examination. In [3] M.A.Anwar and et al have explained the application of association mining in finding interesting patterns that help educational administrators and decision makers in higher education sector for the improvement and revision of teaching methodology, restructuring of curriculum, and modifying the prerequisites requirements of various courses. Aher and et al have experimented the association rule mining algorithms like Apriori, Tertius, FP-Growth etc in predicting the course selection by student and found Apriori algorithm to be best suited for his data set.

Association Rule Mining is one of the most popular data mining methods for discovering interesting relationships between variables in transaction databases or other data repositories. An association rule is an implication $X \rightarrow Y$, where X and Y are disjoint itemsets (i.e., sets with no items in common). The intuitive meaning of such a rule is that when X appears, Y also tends to appear. The two traditional measures for evaluating association rules are support and confidence. The confidence of an association rule $X \rightarrow Y$ is the proportion of the transactions containing X which also contain Y . The support of the rule is the fraction of the database that contains both X and Y . Association rule mining is divided into two sub tasks.

(i) To discover those itemsets whose occurrences exceed a predefined support threshold, and which are called frequent itemsets.

(ii) to generate association rules from those large items constrained by minimal confidence.

Alaa El-Halees[6] has used association rule mining as a tool to analyse the behaviour of the student using the academic records and related the rule with the result of the student .In this paper we propose Apriori algorithm and Tertius algorithm to understand the behavaiou of the students towards a particular learning style.

3.1 IMPLEMENTATION OF APRIORI ALGORITHM

Association analysis is one of the important task in data mining which generates interesting rules for a given data set. The author in [10] have used Apriori algorithm to generate a common rules and also rare rules and compared the results for a moodle data base.

3.1.1 Data Sets used for the work.

Data set comprises of questions comprising of students interest towards their learning style, learning environment, their eagerness to learn and apply current technology, their attitude towards self learning etc. The questionnaire was circulated among the students and the responses for the same were collected. The data set was then pre-processed and the two algorithms were applied on it.

Apriori is an unsupervised rule mining algorithm which generates frequent item sets depending upon support and confidence values. It iteratively reduces the minimum support until it finds the appropriate rules. The algorithm starts identifying the frequent individual items in the database and extends the iteration 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. The algorithm uses minimum support uniformly and changes the confidence level.

The minimum support has been assigned as 0.75 for the given data set and Minimum confidence metric is set as 0.9 for the first iteration. Five cycles were performed to find the large item sets. The size of the item set is 11 during the 1st iteration. The item sets are described below:

Table 1: Frequent Itemset-Apriori

S.No	Question Number and value	Frequent item set total
1	Q1=no	127
2	Q2=No	128
3	Q6=Never	131
4	Q8=Week	138
5	Q9=Week	142
6	Q10=Week	131
7	Q11=Week	126
8	Q13=Week	142
9	Q14=Dictionary	144
10	Q15=Google	161
11	Q16=No	142

During the second iteration the number of frequent item sets were 19 and third iteration it is 6. The best rules found after the generation of itemset is shown in table 2. The confidence level of

the rules is 0.98. the table clearly reveals that Q15 is one of the most frequently generated item where it is ruled with all other questions. The rules framed can be explained as below :

The other rules are listed in table 2. They can also interpreted in the similar way.

Table 2 : Learning rules from Apriori algorithm

S.No	Rule
1	Q10=Week ==> Q15=Google
2	. Q1=no ==> Q15=Google
3	Q11=Week ==> Q15=Google
4	Q14=Dictionary ==> Q15=Google
5	Q9=Week ==> Q15=Google
6	Q13=Week ==> Q15=Google
7	Q16=No ==> Q15=Google
8	Q8=Week ==> Q15=Google
9	Q9=Week Q14=Dictionary ==> Q15=Google
10	Q6=Never ==> Q15=Google

The association rules states that the students are more comfortable with the use of new technology and self learning than the traditional learning style. From table 2, it is clear that learning style has got shift in paradigm from being auditory and visual learners to visual and tactile learners. The combination of the questions framed as a rule is a proof for the same.

3.2 IMPLEMENTATION OF TERTIUS ALGORITHM

The Tertius association rule algorithm is presented in Flach & Lachiche. The algorithm is designed to extract first order rules using unsupervised learning^[11]. The algorithm used for the research is a supervised learning classification algorithm. The rules will be generated depending on the number of literals considered for hypothesis. The algorithm outputs three values namely (i) True Positive rate (TPR), (ii) False Positive rate (FPR) and (iii) ROC value.

The hypothesis is set depending of the number of literals appearing the data set. The maximum number of literals taken for the purpose of research is 4 and the experiment is performed till it reaches the minimum. The experiment is repeated by keeping the classification "ON" and the minimum confirmation value as 10. The following observations are made

- (i) As the number of literals are decreased, the number of hypothesis considered and explored also reduced.
- (ii) The rules generated are not repeated

The table3 gives us the range of the hypothesis selected and explored to get a set of 10 rules for varying literals

Table 3 : Analysis of Hypothesis

No. of literals used	No. of hypothesis considered	No. of hypothesis explored
4	7280	3887
3	1137	767
2	110	109

From table3 it is clear that as the number of literals is reduced, the number of hypothesis considered and explored are linear.

The set of rules generated by tertius algorithm when the number of literals is 2 is listed below:

- ✓ Q19 = No ==> Q1 = no
- ✓ Q19 = Yes ==> Q1 = yes
- ✓ Q17 = Never ==> Q1 = no
- ✓ Q17 = Week ==> Q1 = yes
- ✓ Q4 = Yes ==> Q1 = yes
- ✓ Q4 = No ==> Q1 = no
- ✓ Q12 = Week ==> Q1 = no
- ✓ Q12 = day ==> Q1 = yes
- ✓ Q3 = Monthly ==> Q1 = no
- ✓ Q3 = Weekly ==> Q1 = yes

From the rule set it is clear that the rules are reversible depending the Yes/No values in the questionnaire. This reveals the behaviour of the student in diverse situations. The interpretation of the rule clearly states that the students of today's community are more divergent towards the internet technologies and new self learning methods. The same rules are conjuncted with each other when the numbers of literals are increased. Therefore it is clear that these questions are more frequent in the item set than the others. The table4 shows the output of the tertius algorithm.

Table4: Ouput of Tertius algorithm

True Positive Rate	False Positive Rate	Receiver Operating Characteristic
0.35945	0.75591	0.08108
0.32095	0.7027	0.24409
0.24648	0.74016	0.40541
0.22157	0.58268	0.2973
0.21507	0.56757	0.25984
0.20973	0.7027	0.41732
0.20407	0.79528	0.51351
0.19846	0.20472	0.48649
0.18308	0.40158	0.18919
0.1793	0.81081	0.59843

The algorithm displays the TPrate, FPrate and ROC values of each of the rule generated. The ROC value of all the rules is less than 1 implies that the rules generated are very efficient.

4. CONCLUSIONS

Association rule mining techniques have been identified as one of the easiest and interesting method to apply on educational data. The algorithms are capable of generating frequent and rare association mining rules depending on the necessity of the work. Apriori algorithm has been found as one of the best methods to bring out frequent item sets as well as interesting rules in finding out the behaviour of the students towards the learning style. It is found that the students are moving from traditional / auditory type of learning towards visual as well as tactile learning. Tertius algorithm is capable of finding out rules and conjuncting them when the number of literals are increased. They are also capable of classifying the rules according to the values present in the questionnaire. The output in the form of TPR, FPR and ROC values also explains the efficiency of the algorithm. Therefore these algorithms can be used in other domains of education to bring out interestingness among the data present in the repository.

REFERENCES

1. Abdous, M., He, W., & Yen, C.-J. (2012). Using Data Mining for Predicting Relationships between Online Question Theme and Final Grade. *Educational Technology & Society*, 15 (3), 77–88.
2. Oladipupo O.O. & Oyelade O.J. and et al, Knowledge Discovery from Students' Result epository: Association Rule Mining Approach, *International Journal of Computer Science & Security (IJCSS)*, Volume (4) : Issue (2), pp 199-208
3. M. A. Anwar, Naseer Ahmed, and Wajahatullah Khan, Analysis of Students' Grades in Mathematics, English, and Programming Courses: A KDD Approach, *International Journal of Future Computer and Communication*, Vol. 1, No. 2, August 2012
4. Marie Bienkowski and et al, Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, *Center for Technology in Learning SRI International*, October 2012.
5. Aher, Sunita B.; Lobo, L.M.R.J, A Comparative Study of Association Rule Algorithms for Course Recommender System in E-learning, *International Journal of Computer Applications*; Feb 2012, Vol. 39, p48
6. ALAA EL-HALEES, Mining Students Data To Analyze Learning Behavior: A Case Study
7. Dr. Varun Kumar, Anupama Chadha, Mining Association Rules in Student's Assessment Data, *IJCSI International Journal of Computer Science Issues*, Vol. 9, Issue 5, No 3, September 2012, pp 211-216
8. Anwar M. A. and et al, Information Mining in Assessment Data of Students' Performance, *International Journal of Engineering Science and Innovative Technology (IJESIT)* Volume 1, Issue 2, November 2012.
9. Peter a. Flach and et al, Confirmation-Guided Discovery of First-Order Rules with Tertius, *Machine Learning*, 42, 61–95, 2001 © 2001 Kluwer Academic Publishers. Manufactured in The Netherlands.
10. Cristóbal Romero and et al, Mining Rare Association Rules from e-Learning Data, *proc EDM 2011*, pp 171-181.
11. Agathe Merceron and et al, Interestingness Measures for Association Rules in Educational Data, *Proc EDM*, 2011.
12. Samia Michail, Understanding School Responses to Students' Challenging Behaviour, 2011 *UnitingCare Children, Young People and Families*
13. Weka Tertius Explanation, October 19, 2010.
14. John Joseph And Kym Brown, Student Behaviour Management, Using The Emotional Rooms, This article is an excerpt from the book *Starting Out, The Beginning Teacher's Companion*, © Focus Education, 2001.