USAGE OF ASSOCIATION RULE MINING IN COURSE SELECTION FOR INDUSTRIAL TRAININGS

Pratiyush Guleria[1], Manu Sood[2]

[1, 2] *Department of Computer Science, Himachal Pradesh University, Shimla, Himachal Pradesh, (India)*

ABSTRACT

This paper focuses on usefulness of Association Rule Mining in Education and Councelling.Knowledge extracted using ARM will be helpful in decision making for students to determine courses chosen for industrial trainings. In this paper, we are deriving preferable courses for pursuing trainings for students based on course combinations. Here, two measures are used for deciding the usefulness of an association rule i.e. support and confidence and only those rules are selected which satisfy both a minimum support and a minimum confidence threshold.

Keywords: ARM, Councelling, Confidence, Decision, Education, Support.

I. INTRODUCTION

In data mining, association rule learning is a method for discovering interesting relations between variables in large databases [1]. ARM task is to discover the hidden association relationship between the different item sets in transaction database [2]. An X⇒Y type association rule expresses a close correlation between items in a database [3]. Association rules provide information in the form of if-then statements and they are probabilistic in nature. If part is the antecedent and then part is the consequent. Association rules analyzes the antecedent and consequent for set of items called item set that are disjoint having no items in common and for examining each row in database, user has to set two threshold values i.e. the first value is called the support for the row and the second is called the confidence for the row."Support" is simply the number of transactions that includes all the items in the antecedents and consequent part of the row. It can sometimes be expressed as percentage of total number of records in the database. In ARM [4], rules are selected only if they satisfy both a minimum support and a minimum confidence threshold.

Support of the rule A=>B is shown in Eq .1

support (A=>B[s, c]) = p(AUB) = support ({A, B}) ---------- (1)

s,c represents support and confidence. Eq. 1 denotes the frequency of the rule within all transactions in the database i.e. the probability that a transaction contains both A and B and Eq. 2 denotes the percentage of transactions containing A which also contains B i.e. the probability that a transaction containing A also contains B [5].

Confidence of the rule A=>B is shown in Eq.2

confidence (A=>B[s, c]) = p (B|A) = p(AUB)/p(A)

= support ({AUB})/support ({A}) ---------- (2)

Table I shows the hypothetical list of course combinations taken by students for their industrial trainings. Here each row is considered as transaction, each comprising a combination of variables or item sets. From Table I, strong rules are derived shown in Table II with Rule Support and Rule Confidence. Threshold values assumed for Support is 0.3 and Confidence is 0.75.

TABLE I. Hypothetical List of Course Combinations

SID	Courses Combination(X)	Preferable Course(Y)
1	Java,J2EE	Android
2	Java,J2EE,HTML	Android
3	HTML,Javascript	PHP
4	C,C++	Asp.Net
5	J2EE,Java	Android
6	C,C++	Java
7	C,C++,VB.Net	Java
8	C,C++	Java
9	HTML	PHP
10	HTML,Asp.net	PHP
11	----------	--------

TABLE II Rule Support and Confidence

Rules	Rule Support	Rule Confidence
Java,J2EE=> Android	0.3	1
HTML=>PHP	0.3	0.75
C,C++=>Java	0.3	0.75

II. LITERATURE REVIEW

In [6], author has discussed about image classification using Association Rule Mining with decision tree algorithm. Association Rule Mining task is to find out hidden relationships between different item sets. Enrique Garcia et.al [3] describes a collaborative educational data mining tool based on association rule mining. This tool helps in improvement of e-learning courses and allows teachers to analyze and discover hidden information based on interaction between the students and the e-learning courses. According to Ruijuan Hu [1], Data Mining based on Association Rules is playing important role in Medical Field. Using Association Rule learning interesting relations between variables in large databases are discovered and Apriori is the best known algorithm to mine association rules. In [4], author has applied ARM approach in social-science related fields such as education and Councelling. Mirela Danubianu, Stefan Gheorghe Pentiuc and Iolanda Tobolcea [7] conducted a case study on mining association rules inside a Relational Database. In [8], author presents association rule mining for Students Assessment Data. Student's performance can be analyzed using Association Rule Mining. In [9], author presents data mining in education environment that identifies student's failure patterns using ARM technique. Author in [10] has done Frequent Pattern Mining with WEKA tool using Apriori Algorithm. Stefan Mutter, Mark Hall, and Eibe Frank [11] have used classification approach to evaluate the output of Confidence-

Based Association Rule Mining. In [12], author has worked on a problem of mining association rules between items in a large database of sales transactions using PHP as front-end and MySQL database and analyzed buying habits of customers for improving sales.

III. MINING USING APRIORI ALGORITHM

Apriori algorithm given by Agrawal & Srikant is the best known algorithm to mine association rules. Apriori Algorithm is having a property which states that all nonempty subsets of a frequent item set must also be frequent and an item set is any subset of all the items in the database. Table III shows the steps followed by Apriori algorithm.

TABLE III. Steps Followed by APRIORI Algorithm

Step 1	Initially make a single pass over the dataset to determine the support of each item.
Step 2	Find all frequent 1-itemsets.
Step 3	Iteratively generate new candidate set for 2-itemset, 3 –item set and so on from the frequent item sets found in the previous iterations.
Step 4	Repeat step 2 for finding all frequent 2-itemsets, 3 item sets.
Step 5	Support for each candidate is then counted and checked against the minsup threshold value given.
Step 6	Algorithm eliminates or prunes all candidate item sets whose support count is less than minsup threshold value given.
Step 7	Finally algorithm terminates when there are no new frequent item sets generated.

Apriori algorithm traverses the item set one level at a time, from frequent 1-itemsets to the maximum size of frequent itemsets.Secondly, new candidate item sets are generated from the frequent item sets found in the previous iterations and support of each is then counted and checked against the minimum threshold values [13].

IV. RESULTS AND DISCUSSION

In this paper, dataset consisting of different course combinations(X) and preferable courses(Y) shown in Table I are fed into WEKA tool where Apriori Association Rule Mining Algorithm is applied.

TABLE IV. APRIORI Associator Output

Minimum support: 0.1 (1 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18
Generated sets of large item sets:
Size of set of large item sets L (1): 20
Size of set of large item sets L (2): 13

Table IV shows that minimum support taken for the dataset is 0.1 and confidence is 0.9.Number of cycles performed on the dataset is 18 and number of frequent 1-Itemset and 2 item sets are 20 and 13.

TABLE V. Best Rules Found By APRIORI

Course Combination(X)=Java,J2EE 1	==> Preferable Course(Y)=Android 1	conf:(1)
Course Combination(X)=Java,J2EE,HTML 1	==> Preferable Course(Y)=Android 1	conf:(1)
Course Combination(X)=HTML,Javascript 1	==> Preferable Course(Y)=PHP 1	conf:(1)
Preferable Course(Y) = Asp.Net 1	==> Courses Combination(X)=C,C++ 1	conf:(1)
Course Combination(X)=J2EE,Java 1	==> Preferable Course(Y)=Android 1	conf:(1)
Course Combination(X)=C,C++,VB.Net 1	==> Preferable Course(Y)=Java 1	conf:(1)
Course Combination(X)=HTML 1	==> Preferable Course(Y)=PHP 1	conf:(1)
Course Combination(X)=HTML,Asp.net 1	==> Preferable Course(Y)=PHP 1	conf:(1)
Preferable Course(Y)=C++ 1	==> Courses Combination(X)=C 1	conf:(1)
Course Combination(X)=C 1	==> Preferable Course(Y)=C++ 1	conf:(1)

Best rules generated by Apriori and Predictive Apriori Associator in Weka tool on dataset in Table II are shown in Table V and VI.Table V shows the preferable courses for students based on their course combinations.E.g: "Android" is the preferable course for industrial training for those students who have done java and J2EE in their semesters.

TABLE VI. Best Rules Found By Predictive APRIORI

Course Combination(X)=C,C++ 3 ==> Preferable Course(Y)=Java 2 acc:(0.47888)
Preferable Course(Y)=Java 3 ==> Courses Combination(X)=C,C++ 2 acc:(0.47888)
Course Combination(X)=Java,J2EE 3 ==> Preferable Course(Y)=Android 3 acc:(0.47888)

Predictive Apriori Algorithm uses pruning strategy where it searches for the best rules and highly accurate rules which are included by less accurate ones remain part of the output. Predictive Apriori uses an increasing support values and is able to mine a high quality set of association rules. Predictive Apriori algorithm performs well when it is used to generate a small set of rules [11].

V. CONCLUSION

In this paper we have discussed about ARM and its usage in Educational field. Using ARM technique, students seek help in choosing right course for their industrial trainings based on different course combinations.ARM is used to find associations between frequently occurring variables. Association rules are generated based on the frequent variables in datasets. Apriori is the algorithm used for mining of frequent patterns from the transaction database. Rules discovered using Apriori Algorithm not only help students but also help teachers to find student's interest towards industry oriented courses in an e-learning environment and enhances the effectiveness of academic planning, decision-making. In future, Association Rule Mining would be helpful in finding rules related to industry demanding courses to be introduced into syllabi.

REFERENCES

[1] Ruijuan Hu, *Medical Data Mining Based on Association Rules*, Computer and Information Science, ISSN 1913-8989 E-ISSN 1913- 8997, *Vol. 3, No. 4;* November 2010.

[2] B. Liu and C.K. Wong, *" Improving an association rule based classifier",* journal In Principles of Data Mining and Knowledge Discovery, p. 504–509, 2000.

[3] Enrique García, Cristobal Romero, Sebastián Ventura, Carlos de Castro, "*A collaborative educational association rule mining tool"*, Internet and Higher Education 14 (2011) 77–88, doi:10.1016/j.iheduc.2010.07.006.

[4] Dion H. Goh, Rebecca P. Ang,*"An introduction to association rule mining: An application in counseling and help-seeking behavior of adolescents",* Behavior Research Methods 2007, 39 (2), 259-266.

[5] Pirjo Moen, "*Data mining methods*", Spring 2005, referred from http://www.cs.helsinki.fi/u/ronkaine/tilome/luentomateriaali/TiLoMe-170105.pdf.

[6] P. Rajendran, M.Madheswaran,"*Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm*", Journal of Computing, ISSN 2151-9617, *Vol 2, Issue 1*, January 2010.

[7] Mirela Danubianu, Stefan Gheorghe Pentiuc, Iolanda Tobolcea,"*Mining Association Rules Inside a Relational Database – A Case Study*",ICCGI 2011 : The Sixth International Multi-Conference on Computing in the Global Information Technology, Copyright (c) IARIA,ISBN: 978-1-61208-139-7, 2011.

[8] Varun Kumar, Anupama Chadha,"*Mining Association Rules in Student's Assessment Data*", IJCSI International Journal of Computer Science Issues, ISSN (Online): 1694-0814, *Vol. 9, Issue 5*, No. 3, September 2012.

[9] Oladipupo, I.C O.O., Oyelade O.J.,"*Knowledge Discovery from Students' Result Repository: Association Rule Mining Approach*", International Journal of Computer Science & Security (IJCSS), *Vol 4: Issue (2).*

[10] Paresh Tanna, Yogesh Ghodasara,"*Using Apriori with WEKA for Frequent Pattern Mining*", International Journal of Engineering Trends and Technology (IJETT), ISSN: 2231-5381, *Vol 12 No.3*, Jun 2014.

[11] Stefan Mutter, Mark Hall, Eibe Frank, "*Using Classification to Evaluate the Output of Confidence-Based Association Rule Mining*".

[12] K.S. Adewole, A.G. Akintola & A.R. Ajiboye,"*Frequent Pattern and Association Rule Mining from Inventory Database Using Apriori Algorithm*", African Journal of Computing & ICT, ISSN 2006-1781, *Vol 7. No. 3* - September, 2014.

[13] Tan, Pang-Ning, Vipin Kumar, "Chapter 6. *Association Analysis: Basic Concepts and Algorithms."* Introduction to Data Mining. Addison-Wesley. ISBN321321367, 2005