International Journal of Computer Applications (0975 – 8887)
Volume 39– No.1, February 2012

A Comparative Study of Association Rule Algorithms for Course Recommender System in E-learning

Sunita B. Aher
M.E. (CSE) -II
Walchand Institute of Technology
Solapur University India

Lobo L.M.R.J.
Associate Professor, Head, Department of IT
Walchand Institute of Technology
Solapur University India

ABSTRACT
A course Recommender System plays an important role in predicting the course selection by student. Here we consider the real data from Moodle course of our college & we try to obtain the result using Weka. Association rule algorithms are used to find out the best combination of courses in E-Learning. Here in this paper we consider four association rule algorithms: Apriori Association Rule, PredictiveApriori Association Rule, Tertius Association Rule & Filtered Associator. We compare the result of these four algorithms & present the result. According to our simulation result, we find that Apriori association algorithms perform better than the Predictive Apriori Association Rule, Tertius Association Rule, & Filtered Associator in predicting the course selection based on student choice.

KEY WORDS
Weka, Apriori Association Rule, Predictive Apriori Association Rule & Tertius Association Rule, Filtered Associator.

1. INTRODUCTION
Data Mining can be used to extract knowledge from e-learning systems such as Moodle, through the analysis of the information available in the form of data generated by their users. The main objective becomes finding the patterns of system usage by teachers and students and discovering the students' learning behavior [1].

We have used Moodle- Learning management System which is an open-source course management learning system to help educators create effective online learning communities [3].This Moodle is used to collect the data from student about the courses in which they are interested.

In this Course Recommender System, we add 13 course categories & under each course category there are courses. So there are near about 82 courses we have added to Moodle. We consider the two departments: Computer science & Engineering and Information Technology for collection of data. As we are using the Moodle to collect the data, the information about course selection by student is stored in moodle database. We use this database to collect the data which is relevant to find the best combination of courses. After collecting the data from database, we preprocess the data. Preprocessing means that we delete those columns and rows from database having very low course count & student count respectively. After preprocessing of the data we got only 8 courses out of 82. These courses are: STLD (Switching Theory & Logic Design), DS-I(Data Structure-I), DS-II(Data Structure-II), OS-I(Operating System-I), CN-I(Computer Network-I), VB(Visual Basic), CP(C-Programming), JP(Java Programming). After preprocessing of data, next step is to find the result using the open source data mining tool, Weka. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code [8]. The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality [9]. We find out result using four association rule algorithms i.e. Apriori Association Rule, PredictiveApriori Association Rule, Tertius Association Rule and Filtered Associator & compare it. This process is shown in figure1 [10].

2. LITERATURE REVIEW
The research [2] aim is to extract the significant prevention factors for particular types of cancer. They used three association rule mining algorithms, Apriori, Predictive Apriori and Tertius algorithms in order to discover most of the significant prevention factors against these specific types of cancer.

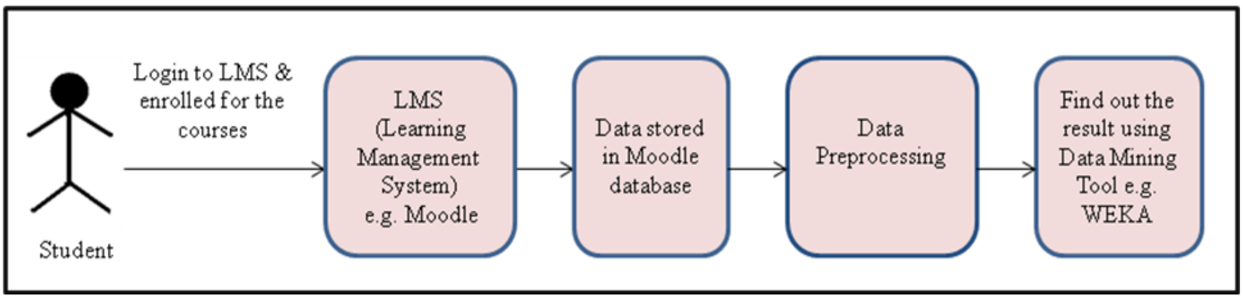

Figure 1: Framework to find the best combination of subject in Course Recommender System[10]

The research [4] compares the performance of three popular association rule mining algorithms, namely Apriori, predictive Apriori and Tertius based on data characteristics. The accuracy measure is used as the performance measure for ranking the algorithms. A meta-learning technique is implemented for a unique selection from a set of association rule mining algorithms. On the basis of experimental results of 15 UCI data sets, this research discovers statistical information based rules to choose a more effective algorithm.

The goal of research [5] is to experimentally evaluate association rule mining approaches in the context of XML databases. Algorithms are implemented using Java. For experimental evaluation different XML datasets are used. Apriori and FP Tree algorithm have been implemented and their performance is evaluated extensively.

In paper [6], they conducted experiment in the WEKA environment by using four algorithms namely ID3, J48, Simple CART and Alternating Decision Tree on the spam email dataset and later the four algorithms were compared in terms of classification accuracy. According to their simulation results the J48 classifier outperforms the ID3, CART and ADTree in terms of classification accuracy.

3. ASSOCITION RULE ALGORITHMS

Association rules are used to find the frequent pattern, association or correlation in transaction database. Association rule mining can be used in Basket Data Analysis, Educational Data Mining, Classification, Clustering etc. Association Rule algorithms are Apriori, Sampling, Partitioning & Parallel Algorithm.
This section describes the Apriori Association Rule, PredictiveApriori Association Rule, Tertius Association Rule & Filtered Associator algorithm briefly.

3.1 Apriori Association Rule

Apriori Association rule is used to mine the frequent patterns in database. Support & confidence are the normal method used to measure the quality of association rule.

- **Support** for the association rule X->Y is the percentage of transaction in the database that contains XUY [9].
- **Confidence** for the association rule is X->Y is the ratio of the number of transaction that contains XUY to the number of transaction that contain X [9].

Terms related to this algorithm are as follows [11]:

- **Frequent Itemsets:** The sets of item which has minimum support & it is denoted by L_i for i^{th} itemset.
- **Apriori Property:** Any subset of frequent itemset must be frequent.
- **Join Operation:** To find L_k , a set of candidate k-itemsets is generated by joining L_{k-1} with itself.
- **Join Step:** Candidate item C_k is generated by joining $L_{k-1}$with itself
- **Prune Step:** Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset

The Apriori association algorithm is given below [7]:

Algorithm: Apriori Association Rule Algorithm
Purpose : To find subsets which are common to at least a minimum number C(Confidence Threshold) of the itemsets.
Input : Database of Transactions D={t_1,t_2, …, t_n}
Set if Items I= {I_1 ,I_2 ,….I_k}
Frequent (Large) Itemset L
Support,
Confidence.
Output : Association Rule satisfying Support & Confidence
Method :

1. C_1 = Itemsets of size one in I;
2. Determine all large itemsets of size 1, L_1;
3. i = 1;
4. Repeat
5. i = i + 1;
6. C_i = Apriori-Gen(L_{i-1});
7. Apriori-Gen(L_{i-1})
    1. Generate candidates of size i+1 from large itemsets of size i.
    2. Join large itemsets of size i if they agree on i-1.
    3. Prune candidates who have subsets that are not large.
8. Count C_i to determine L_i;
9. until no more large itemsets found;

Figure 2 shows the generation of itemsets & frequent itemsets where the minimum support count is 2 [9]. Figure 3 shows the flowchart for the Apriori Association Rule algorithm.
To generate the association rule from frequent itemset we use the following rule:

- For each frequent itemset L, find all nonempty subset of L
- For each nonempty subset of L, write the association rule S → (L-S) if support count of L/support count of S >= Minimum Confidence

The best rule from the itemset L= {2, 3, 5} are calculated as follows:
Consider the minimum support is 2 & minimum confidence is 70%. All nonempty subset of {2, 3, and 5} are: {2,3},{2,5},{3,5},{2},{3},{5}.

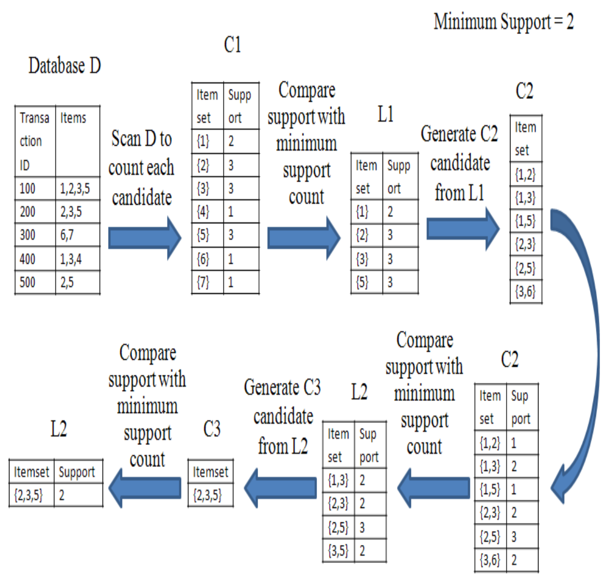

Figure 2: Generation of itemsets & frequent itemsets [9]

Rule 1: {2, 3} → {5} Confidence = Support Count of ({2, 3, 5})/ Support Count of ({2, 3}) = 2/2 = 100%
Rule 2: {2, 5} → {3} Confidence = Support Count of ({2, 3, 5})/ Support Count of ({2, 5}) = 2/3 = 67%
Rule 3: {3, 5} → {2} Confidence = Support Count of ({2, 3, 5})/ Support Count of ({3, 5}) = 2/2 = 100%
Rule 4: {2} → {3, 5} Confidence = Support Count of ({2, 3, 5})/ Support Count of ({2}) = 2/3 = 67%
Rule 5: {3} → {2, 5} Confidence = Support Count of ({2, 3, 5})/ Support Count of ({3}) = 2/3 = 67%
Rule 6: {5} → {2, 3} Confidence = Support Count of ({2, 3, 5})/ Support Count of ({5}) = 2/3 = 67%

Hence the accepted rules are Rule 1 & Rule 3 as the confidence of these rules is greater than 70% [10].

In this paper, we use Apriori Association algorithm to find out the result i.e. best combination of courses, after the preprocessing of data stage. In Weka the option available with Apriori association rule algorithm are car, class Index, delta, lower bound minimum support, metric type, minimum metric, number of rules, output itemsets, remove all missing columns, significance level, upper bound minimum support, verbose.

3.2 Predictive Apriori Association Rule

In predictive Apriori association rule algorithm, support & confidence is combined into a single measure called predictive accuracy. This predictive accuracy is used to generate the Apriori association rule. In Weka, this algorithm generates 'n' best association rule based on n selected by the user.

3.3 Tertius Association Rule

This algorithm finds the rule according to the confirmation measures (P. A. Flach, N. Lachiche 1999). It uses first order logic representation. It includes various option like class Index, classification, confirmation Threshold, confirmation Values, frequency Threshold, horn Clauses, missing Values, negation, noise Threshold, number Literals, repeat Literals, roc Analysis, values Output etc.

3.4 Filtered Associator

This algorithm is a class for running an arbitrary associator on data that has been passed through an arbitrary filter. Like the associator, the structure of the filter is based exclusively on the training data and test instances will be processed by the filter without changing their structure. Here in this algorithm we can consider the Apriori, Predictive Apriori & Tertius association rule algorithm to get the result.

4. EXPERIMENTAL RESULT

We find the result using these four association rule i.e. Apriori Association Rule, Predictive Apriori Association Rule, Tertius Association Rule & Filtered Associator. The result is shown in table 1.

First row in the table 1 represent the result using Tertius Association Rule algorithm. Result in this first row shows the association rule containing "no" also. As we are recommending the course to the student, these results cannot be considered for recommending the course.

Second row in the table represent the result using Predictive Apriori Association Rule algorithm. This rule also contains "no". We cannot consider this rule also for recommending the course to the student.

Third, fourth & fifth row in table 1 represents the result using the Filtered Associator. Third row represents the result when we use the Apriori association rule in Filtered Associator which is same as the sixth row in table 1. Fourth row represents the result when we use the Predictive Apriori association rule algorithm in Filtered Associator which cannot be used to recommend the course as it contains the association rule containing "no". Fifth row represents the result using Tertius in Filtered Associator which cannot be used for course recommender system.

Sixth row in table 1 represent the result using Apriori Association Rule algorithm. As all the rules in the third row contain "yes" only, we can use this Apriori association rule to recommend the course to the student. As we increase the lower support bound, we get the refined rule as shown in sixth row of table1. The meaning of the rule is that if student is interested in Data Structure-II then he/she is interested in Data Structure-I. Due to above rule, we can recommend to new student who has recently enrolled for Data Structure-II course, the Data Structure-I as a course to be opted. Association rule obtained here also match with the in general real world interdependencies among the course.

Table 1: Result of various association rule algorithm using open source data mining tool Weka

Courses considered	**Result of association rule algorithms using open source data mining tool WEKA**		
1. Tertius Association Rule			
1.STLD 2.DS-I 3.DS-II 4.OS-I 5.CN-I 6.VB 7.CP 8.JP	**Best rules found:** 1. CN_I = yes and JP = yes → OS_I = yes or VB = no 2. DS_I = yes and CN_I = no and VB = yes → OS_I = no 3. DS_I = yes and OS_I = no → CN_I = no or JP = no 4. DS_II = yes and OS_I = no → CN_I = no or JP = no 5. CN_I = yes and JP = yes → OS_I = yes or CP = no 6. CN_I = yes and JP = yes → OS_I = yes or STLD = no 7. VB = no → OS_I = yes or CN_I = yes 8. CN_I = yes and JP = yes → OS_I = yes 9. DS_II = yes and VB = yes → OS_I = no or CP = yes 10. DS_I = yes and VB = yes → OS_I = no	**Best rules found:** 1. VB = no → OS_I = yes or CN_I = yes 2. CN_I = yes and JP = yes → OS_I = yes 3. DS_I = yes and VB = yes → OS_I = no 4. OS_I = yes → CN_I = yes or VB = no 5. DS_II = yes → JP = no or STLD = no	**Best rules found:** 1.VB = no → OS_I = yes or CN_I = yes

2. PredictiveApriori Association Rule				
1.STLD 2.DS-I 3.DS-II 4.OS-I 5.CN-I 6.VB 7.CP 8.JP	**Best rules found:** 1. DS_II=yes OS_I=no JP=yes → DS_I=yes 2. OS_I=no CN_I=no → DS_I=yes 3. DS_II=yes OS_I=no → DS_I=yes 4. STLD=no OS_I=no → DS_I=yes 5. STLD=no DS_II=no → JP=yes 6. OS_I=no VISUAL_BASIC=no → DS_I=yes 7. OS_I=no CN_I=no → VB=yes 8. STLD=yes OS_I=no CP=no → DS_I=yes 9. OS_I=no CN_I=no JP=no → STLD=yes 10. OS_I=no CN_I=no VB=yes JP=no → CP=yes	**Best rules found:** 1. OS_I=no CN_I=no → DS_I=yes 2. DS_II=yes OS_I=no → DS_I=yes 3. STLD=no OS_I=no → DS_I=yes 4. STLD=no DS_II=no → JP=yes 5. STLD=no CP=no → DS_I=yes	**Best rules found:** 1. STLD=no OS_I=no → DS_I=yes	
3. A) Filtered Associator Apriori association rule algorithm				
1.STLD 2.DS-I 3.DS-II 4.OS-I 5.CN-I 6.VB 7.CP 8.JP	**1.Minimum support: 0.3** **2.Minimum metric <confidence>:0.9** **3. Number of cycles performed: 14** **Best rules found:** 1. DS_II → DS_I 2. DS_II JP → DS_I 3. DS_II VB → DS_I 4. DS_II CN_I CP → DS_I 5. STLD DS_II CP → DS_I 6. DS_II CP → DS_I 7. DS_II VB CP → DS_I 8. STLD DS_II → DS_I 9. DS_II CN_I → DS_I 10. DS_II CP JP → DS_I	**1.Minimum support: 0.4** **2.Minimum metric <confidence>: 0.9** **3.Number of cycles performed: 12** **Best rules found:** 1. DS_II → DS_I 2. DS_II JP → DS_I 3. DS_II VB → DS_I 4. DS_II CP → DS_I 5. STLD DS_II → DS_I 6. DS_II CN_I → DS_I 7. DS_II OS_I → DS_I	**1.Minimum support: 0.5** **2.Minimum metric <confidence>: 0.9** **3.Number of cycles performed: 10** **Best rules found:** 1. DS_II → DS_I 2. DS_II CP → DS_I	**1.Minimum support: 0.6** **2.Minimum metric <confidence>: 0.9** **3.Number of cycles performed: 8** **Best rules found:** 1. DS_II → DS_I
B) Filtered Associator PredictiveApriori Association Rule				
1.STLD 2.DS-I 3.DS-II 4.OS-I 5.CN-I 6.VB 7.CP 8.JP	**Best rules found:** 1. DS_II=yes OS_I=no JP=yes → DS_I=yes 2. OS_I=no CN_I=no → DS_I=yes 3. DS_II=yes OS_I=no → DS_I=yes 4. STLD=no OS_I=no → DS_I=yes 5. STLD=no DS_II=no → JP=yes 6. OS_I=no VB=no → DS_I=yes 7. OS_I=no CN_I=no → VB=yes 8. STLD=yes OS_I=no CP=no → DS_I=yes 9. OS_I=no CN_I=no JP=no → STLD=yes 10. OS_I=no CN_I=no VB=yes JP=no → CP=yes	**Best rules found:** 1. OS_I=no CN_I=no → DS_I=yes 2. DS_II=yes OS_I=no → DS_I=yes 3. STLD=no OS_I=no → DS_I=yes 4. STLD=no DS_II=no → JP=yes 5. STLD=no CP=no → DS_I=yes	**Best rules found:** 1. STLD=no OS_I=no → DS_I=yes	
C) Filtered Associator Tertius Association Rule				
1.STLD 2.DS-I 3.DS-II 4.OS-I 5.CN-I 6.VB 7.CP 8.JP	**Best rules found:** 1. CN_I = yes and JP = yes → OS_I = yes or VB = no 2. DS_I = yes and CN_I = no and VB = yes → OS_I = no 3. DS_I = yes and OS_I = no → CN_I = no or JP = no 4. DS_II = yes and OS_I = no → CN_I = no or JP = no 5. CN_I = yes and JP = yes → OS_I = yes or CP = no 6. CN_I = yes and JP = yes → OS_I = yes or STLD = no 7. VB = no → OS_I = yes or CN_I = yes 8. CN_I = yes and JP = yes → OS_I = yes 9. DS_II = yes and VB = yes → OS_I = no or CP = yes 10. DS_I = yes and VB = yes → OS_I = no	**Best rules found:** 1. VB = no → OS_I = yes or CN_I = yes 2. CN_I = yes and JP = yes → OS_I = yes 3. DS_I = yes and VB = yes → OS_I = no 4. OS_I = yes → CN_I = yes or VB = no 5. DS_II = yes → JP = no or STLD = no	**Best rules found:** 1.VB = no → OS_I = yes or CN_I = yes	
4. Apriori Association Rule				
1.STLD 2.DS-I 3.DS-II 4.OS-I 5.CN-I 6.VB 7.CP 8.JP	**1.Minimum support: 0.3** **2.Minimum metric <confidence>:0.9** **3. Number of cycles performed: 14** **Best rules found:** 1. DS_II → DS_I 2. DS_II JP → DS_I	**1.Minimum support: 0.4** **2.Minimum metric <confidence>: 0.9** **3.Number of cycles performed: 12** **Best rules found:** 1. DS_II → DS_I	**1.Minimum support: 0.5** **2.Minimum metric <confidence>: 0.9** **3.Number of cycles performed: 10**	**1.Minimum support: 0.6** **2.Minimum metric <confidence>: 0.9** **3.Number of cycles performed: 8**

	3. DS_II VB → DS_I 4. DS_II CN_I CP → DS_I 5. STLD DS_II CP → DS_I 6. DS_II CP → DS_I 7. DS_II VB CP → DS_I 8. STLD DS_II → DS_I 9. DS_II CN_I → DS_I 10. DS_II CP JP → DS_I	2. DS_II JP → DS_I 3. DS_II VB → DS_I 4. DS_II CP → DS_I 5. STLD DS_II → DS_I 6. DS_II CN_I → DS_I 7. DS_II OS_I → DS_I	**Best rules found:** 1. DS_II → DS_I 2. DS_II CP → DS_I	**Best rules found:** 1. DS_II → DS_I

6. COMPARATIVE STUDY OF ASSOCIATION ALGORITHMS

In Course Recommender System, as we are recommending the course to the student, we need all association rule containing "yes" only. Among all these association algorithms i.e. Apriori Association Rule, Predictive Apriori Association Rule, Tertius Association Rule and Filtered Associator, only the Apriori association rule algorithm gives all association rules containing "yes" only. Other algorithms except the Filtered Associator using Apriori association rule algorithm gives the association rules containing "no" also. Filtered Associator using Apriori association rule algorithm gives the association rule containing "yes" only as we are using the Apriori association rule i.e. the same result as Apriori association rule algorithm. Hence we use the Apriori association rule algorithm to recommend the course to the student based on the various student choices.

5. CONCLUSION AND FUTURE WORK

Here we try to compare the four association rule algorithm in predicting the course selection by the student: Apriori Association Rule, Predictive Apriori Association Rule, Tertius Association Rule & Filtered Associator. As we are recommending the course, we need the algorithm where the association rules consist of "yes" only. So we compare these result using these four association rule algorithm & found that Apriori Association Rule algorithm perform better than other association rule algorithms as it gives the association rule containing "yes" only. Future work include finding out the combination of various data mining algorithm to predict the course selection by student.

6. REFERENCES

[1] Castro, F., Vellido, A., Nebot, A., & Mugica, F. (in press). Applying data mining techniques to e-learning problems: A survey and state of the art. In L. C. Jain, R. Tedman, & D. Tedman (Eds.), Evolution of Teaching and learning paradigms in intelligent environment. Studies in Computational Intelligence (Vol. 62). Springer-Verlag.

[2] Nahar J, Tickle KS, Ali AB, Chen YP: "Significant cancer prevention factor extraction: an association rule discovery approach". Journal of Medical Systems Volume 35, Number 3, 353-367, DOI: 10.1007/s10916-009-9372-8 .

[3] C. Romero, S. Ventura and E. Garcia. Data Mining in Course Management Systems: MOODLE Case Study and Tutorial. Computers and Education, 2007. Num. 51. pp. 368-384.

[4] Mazid, M.M. Shawkat Ali, A.B.M. Tickle, K.S. : "Finding a unique Association Rule Mining algorithm based on data characteristics" Electrical and Computer Engineering, 2008. ICECE 2008. International Conference Issue Date: 20-22 Dec. 2008 On page(s): 902 - 908 Location: Dhaka E-ISBN: 978-1-4244-2015-5 Print ISBN: 978-1-4244-2014-8

[5] Gurpreet Kaur1, Naveen Aggarwal2:" Association Rule Mining in XML databases: Performance Evaluation and Analysis" IJCST Vol. 1, Issue 2, December 2010 ISSN : 2 229-4333 (Print) | ISSN : 0976 - 8491 (Online)

[6] Aman Kumar Sharma, Suruchi Sahni:" A Comparative Study of Classification Algorithms for Spam Email Data Analysis" in International Journal of computer science & Engineering ISSN: 0975-3397 Vol. 3 No. 5 May 2011.

[7] "Data Mining Introductory and Advanced Topics" by Margaret H. Dunham

[8] Weka (2007). http://www.cs.waikato.ac.nz/ml/weka/.

[9] Sunita B Aher and Lobo L.M.R.J.. Data Mining in Educational System using WEKA. *IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT)* (3):20-25, 2011. Published by Foundation of Computer Science, New York, USA (ISBN: 978-93-80864-71-13)

[10] Sunita B Aher and Lobo L.M.R.J. Article: A Framework for Recommendation of courses in E-learning System. Selected in *International Journal of Computer Applications* Published by Foundation of Computer Science, New York, USA **ISSN 0975 – 8887**

[11] http://www.cs.sunysb.edu/~cse634/lecture_notes/07apriori.pdf accessed on date 02-02-2012