International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 4, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Association Rule Mining of Classified Data for Predicting Courses Selected by Students in E-Learning

Sunita B Aher
Walchand Institute of Technology,
Solapur University, India
Email: sunita_aher@yahoo.com

Mr. LOBO L.M.R.J.
Department of IT,
Walchand Institute of Technology, Solapur University, India
Email: headitwits@gmail.com

Abstract **- Data mining is the process which discovers new pattern in large database. Classification & association rule are the techniques of data mining. Classification is supervised machine learning technique which predicts the group membership for data instances. The ADTree (Alternating Decision Tree) is a classification technique that combines decision trees with the predictive accuracy into a set of classification rules. Association rule algorithms are used to show the relationship between data items. Here in this paper we combine these two algorithms & apply it to data obtained from Moodle courses of our college for the Course Recommender System which predicts the course selected by the students. First we apply only association rule to the data & then we consider this combined approach. Here we present the advantage of applying the combined approach to Course Recommender System as compare to the result of application of only the association rule algorithm.**

Keywords **- ADTree Classification algorithm, Apriori Association Rule algorithm, Weka**

I. INTRODUCTION

Data mining which is the extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data [8] can also be used to extract the knowledge from E-learning system such as Moodle. The course recommendation system in e-learning is a system that suggests the best combination of courses in which the students are interested [7]. For collecting the data we consider the student of three years of engineering course i.e. Second year, Third year, & Final year of Computer Science & Engineering and Information Technology. Here we are using Moodle as Learning Management System to collect the data regarding the course selection by student. We have created the student login & gave the access to the student. The framework for Course Recommendation System is shown in figure 2.

In this framework, student first logs in the Learning Management System e.g. Moodle. The system verifies the username & password. After verifying the username & password, student will search the course category & courses. Students will enroll for subject in which they like. This enrollment information is stored in database [7].

After collecting the data from student which is stored in Moodle database, the next stage is to select the relevant data from Moodle database. After selecting the data next stage is to prepare it which is explained in paper [9]. To test the result we use the open source data mining tool WEKA. Since our project is to find the best combination of subject, we use the Apriori machine algorithm for testing the result.

Now we consider the combined approach i.e. classification & association rule algorithm. First we apply the classification algorithm ADTree to the data selected from Moodle database. After classifying the data, we apply the Apriori Association Rule algorithm to find the result. We compare the result of combined approach to the result obtained using only the Apriori Association Rule algorithm. If we use the combined approach then there is no need to preprocess the data.

II. LITERATURE REVIEW

In paper [1], a novel combination strategy for multi-class classification (CSMC) based on multiple rules is proposed. In CSMC, rules are regarded as classification experts, after the calculation of the basic probability assignments and evidence weights; Yang's rule of combination is employed to combine the distinct evidence bodies to realize an aggregate classification. The comparison with popular methods like CBA, C4.5, RIPPER and MCAR indicates that CSMC is a competitive method for classification based on association rule.

In paper [2], they studied a new technique called post-bagging, which consists in resampling parts of a classification model rather than the data. They did this with a particular kind of model: large sets of classification association rules, and in combination with ordinary best rule and weighted voting approaches.

In paper [3], they have explored and applied the combinatorial mathematics in class association rule mining. Their algorithm is based on producing combinations of itemsets including classes using combinatorial mathematics and subsequently finds an associative classifier.

In paper [4], they proposed Associative classification which is a classification of a new tuple using association rules. It is a combination of association rule mining and classification. The main aim of this paper is to improve accuracy of a classifier.

International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 4, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

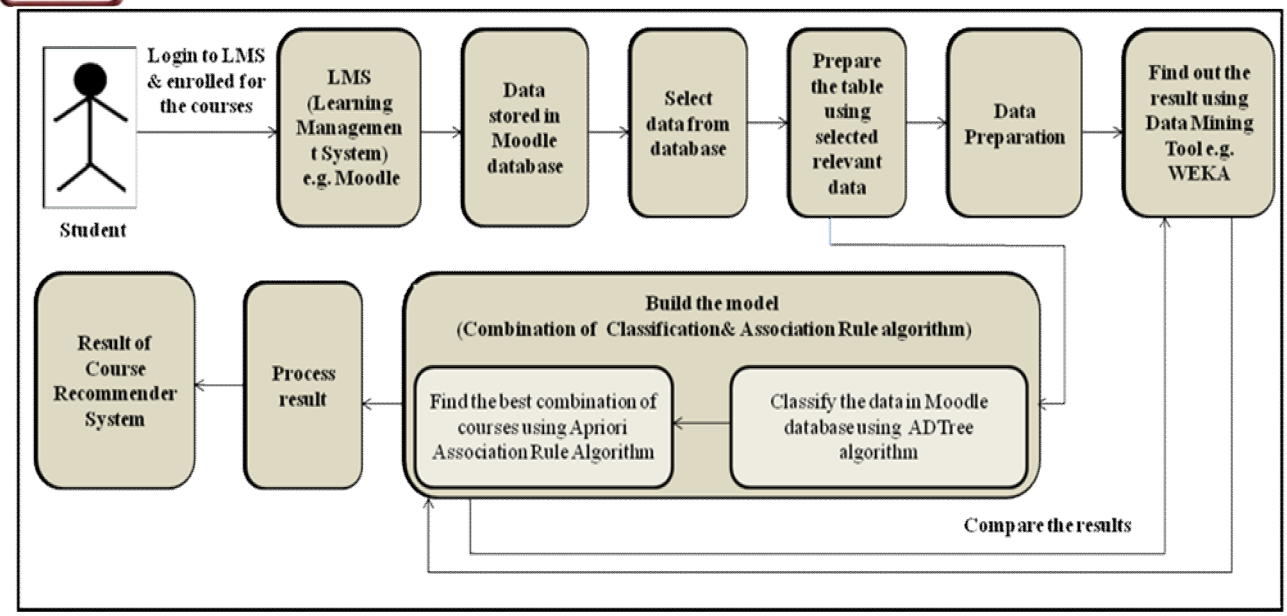

Fig.1. Framework for Course Recommender System

In paper [5], they proposed a novel aspect mining method which combines clustering and association rule technology is provided in this article. Clustering analysis based on the execution traces is provided to find out candidate aspects; while association rule mining based on the execution traces with ordered call is used to find out the crosscuts.

III. ADTREE CLASSIFICATION ALGORITHM

An alternating decision tree (ADTree) is a machine learning method for classification which generalizes decision trees. An alternating decision tree consists of two nodes:

- Decision nodes: specify a predicate condition.
- Prediction nodes: contain a single number. ADTree always have prediction nodes as both root and leaves.

An instance is classified by an ADTree by following all paths for which all decision nodes are true and summing any prediction nodes that are traversed.

IV. APRIORI ASSOCIATION RULE ALGORITHM

Association rule learning in data mining is a technique for discovering interesting relations between variables in large databases.

An example rule for the supermarket could be "{butter, bread} -> milk" meaning that if butter and bread is bought, customers also buy milk.

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. 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 [6].

The algorithm for Apriori association rule algorithm is shown in below.

Apriori Association Rule Algorithm

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 :

C_1 = Itemsets of size one in I;
Determine all large itemsets of size 1, L_1;
i = 1;
Repeat
i = i + 1;
C_i = Apriori-Gen (L_{i-1});
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.

Count C_i to determine L_i;
Until no more large itemsets found;

V. RESULT & IMPLEMENTATION

In this Course Recommendation System, we have considered the 13 course category. Under each category there will courses. So there are about 82 courses. Out of which here we consider the data of 36 courses shown in table 1 where the course count is above 100 & student have to enroll for at least five subjects.

Table1: Course category & courses

Sr No	Course Category Course Name & Code
1	Basic Science(BS) 1.Discrete Mathematical Structure (DMS)
2	Humanities & Social Science(HSS) 1.Communication Skills (COMM)

3	Hardware Design & Engineering(HDE) 1.Switching Theory And Logic Design (STLD) 2.Microprocessor (MP) 3.Computer Organization (CO) 4.Advanced Computer Architecture (ACA) 5.Mobile Computing (MC)
4	Theoretical Computer Science (TCS) 1.Data Structures –I (DS-I) 2.Data Structures –II (DS-II) 3.Design & Analysis of Algorithm (DAA) 4.Formal System And Automata (FSA)
5	System software 1.System Programming (SP) 2.Operating System – I (OS-I) 3.Operating System – II (UNIX) (OS-II)
6	Networks(NT) 1.Computer Networks – I (CN-I) 2.Computer Networks – II (CN-II) 3.Network Security (NS)
7	Database(DB) 1.Database Engineering (DBE) 2.Oracle (SQL)
8	Software Engineering & Principles (SEP) 1.Software Engineering (SE) 2.Software Testing & Quality Assurance (STQA)
9	Image Processing, Graphics And Artificial Intelligence Application(IP&AIP) 1.Computer Graphics (CG) 2.Artificial Intelligence (AI) 3.Human Computer Interfaces (HCI)
10	Web Technology(WT) 1.Hyper Text Markup Language (HTML) 2.JavaScript (JS) 3.Java Server Pages (JSP) 4.ASP.NET (ASP.NET)
11	Principles of Programming Language(PPL) 1.Visual Basic (VB) 2.C-Programming (CP) 3.Advanced C (AC) 4.Object Oriented Design & Programming (OODP) 5.Java Programming (JP) 6.Advanced Java (AJ) 7.C-Sharp (CS)
12	Information Retrieval & Extraction (IRE) 1.Data Mining (DM)

Here we are considering the data extracted from Moodle database of a college after collection of data for course enrollment by student as shown. The table we formed is shown in figure 2. In this table yes represent that the student is interested in that particular course & no represent that student does not like that course.

Course → Roll No of Student↓	PHY	M-I	M-II	AM-I	AM-II	.....	DBE	ADD	ADS	SQL	BIO	.....	AJ	VC++	CS	DW	DM
Becse01	Yes	Yes	Yes	No	No	.....	Yes	Yes	Yes	Yes	Yes	.....	Yes	Yes	No	No	No
Becse0	No	No	No	No	No	.....	Yes	Yes	Yes	Yes	No	.....	Yes	Yes	No	No	No
Beit01	Yes	No	No	No	No	.....	Yes	Yes	Yes	No	No	.....	No	No	No	Yes	Yes
Beit02	No	Yes	Yes	Yes	Yes	.....	No	No	No	Yes	No	.....	Yes	Yes	No	No	No
Tecse01	No	No	No	No	No	.....	Yes	Yes	Yes	Yes	Yes	.....	No	No	No	No	No
Tecse02	Yes	Yes	Yes	Yes	Yes	.....	No	No	No	No	No	.....	Yes	No	No	No	No
Teit01	Yes	No	No	No	No	.....	Yes	No	No	Yes	No	.....	Yes	Yes	No	Yes	Yes
Teit02	No	Yes	Yes	Yes	Yes	.....	Yes	No	No	No	No	.....	No	No	No	Yes	Yes
Secse101	Yes	Yes	Yes	Yes	Yes	.....	No	No	No	Yes	No	.....	Yes	No	No	No	No
Secse102	Yes	Yes	Yes	Yes	Yes	.....	No	No	No	No	No	.....	Yes	Yes	Yes	No	No
Seit01	No	Yes	Yes	Yes	Yes	.....	No	No	No	No	No	.....	Yes	Yes	Yes	No	No
Seit02	Yes	Yes	Yes	Yes	Yes	.....	No	No	No	Yes	No	.....	Yes	No	No	No	No
Secse201	Yes	No	No	No	No	.....	No	No	No	Yes	No	.....	Yes	No	No	No	No
Secse202	Yes	Yes	Yes	Yes	Yes	.....	No	No	No	No	No	.....	Yes	Yes	Yes	No	No

Fig.2. Final table formed to find the result of Course Recommendation System

Before data preparation stage, the result of this system using Apriori association rule contains the rules having “no” only which is shown in third row of table 2. As we are recommending the course, we prepare the data In data preparation step, we prepare the table after deleting those rows & columns from sample table shown in table 1, having very less student count & less course count. After preparation of data we got 8 courses. To test the result we use the open source data mining tool Weka. 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 [6]. The result of applying Apriori association rule after data preparation step is shown in fifth row of table 3. Before Now the association rule contains only “yes”.

The meaning of the association rule “DS=yes -> OS=yes” is that we can recommend to new student who has recently enrolled for DS course, the operating system as a course to be opted. If we increase the support then we get the refined rule which is shown in second row of table 3. For combination of classification & association rule,

there is no need to preprocess the data. First we apply the ADTree classification algorithm & then the Apriori association rule algorithm.

Table 2: Result after application of machine learning algorithms

Courses considered	Parameter considered	Result
Before preprocessing		
36 courses considered in table 1.	Minimum support: 0.6 Minimum confidence: 0.9	Best rules found: 1. Oracle=no ASP_DOT_NET=no → C_Sharp=no conf:(0.94) 2. Design_&_Anaylisis_of_Algorithm=no System_Programming=no → Finite_State_Automata=no conf:(0.93) 3. ASP_DOT_NET=no Advanced_C=no → C_Sharp=no conf:(0.92) 4. ASP_DOT_NET=no Data_Mining=no → C_Sharp=no conf:(0.92) 5. Computer_Organisation=no System_Programming=no → Finite_State_Automata=no conf:(0.92) 6. Design_&_Anaylisis_of_Algorithm=no Oracle=no → C_Sharp=no conf:(0.92) 7. Oracle=no Java_Server_Pages =no → C_Sharp=no conf:(0.92) 8. System_Programming=no Java_Server_Pages=no → Finite_State_Automata=no conf:(0.92) 9. Oracle=no JavaScript=no → C_Sharp=no conf:(0.92) 10. Oracle=no Data_Mining=no → C_Sharp=no conf:(0.92)
After preprocessing		
STLD, DS-I, DS-II, OS-I, CN-I, VB, CP, JP	Minimum support: 0.3 Minimum confidence: 0.9	Best rules found: 1. Data_Structure_II →Data_Structure_I conf:(0.93) 2. Data_Structure_II Java_Programming →Data_Structure_I conf:(0.93) 3. Data_Structure_II Visual_Basic →Data_Structure_I conf:(0.93) 4. Data_Structure_II Computer_Network_I C_Programming → Data_Structure_I conf:(0.93) 5. Switching_Theory_&_Logic_Design Data_Structure_II C_Programming → Data_Structure_I conf:(0.93) 6. Data_Structure_II C_Programming → Data_Structure_I conf:(0.92) 7. Data_Structure_II Visual_Basic C_Programming →Data_Structure_I conf:(0.92) 8. Switching_Theory_&_Logic_Design Data_Structure_II → Data_Structure_I conf:(0.92) 9. DATA_STRUCTURE_II COMPUTER_NETWORK_I →DATA_STRUCTURE_I conf:(0.92) 10. Data_Structure_II C_Programming Java_Programming → Data_Structure_I conf:(0.92)
	Minimum support: 0.4 Minimum confidence: 0.9	Best rules found: 1. Data_Structure_II →Data_Structure_I conf:(0.93) 2. Data_Structure_II Java_Programming →Data_Structure_I conf:(0.93) 3. Data_Structure_II Visual_Basic → Data_Structure_I conf:(0.93) 4. Data_Structure_II C_Programming →Data_Structure_I conf:(0.92) 5. Switching_Theory_And_Logic_Design Data_Structure_II → Data_Structure_I conf:(0.92) 6. Data_Structure_II Computer_Network_I →Data_Structure_I conf:(0.92) 7. Data_Structure_II Operating_System_I →Data_Structure_I conf:(0.9)
	Minimum support: 0.5 Minimum confidence: 0.9	Best rules found: 1. Data_Structure_II →Data_Structure_I conf:(0.93) 2. Data_Structure_II C_Programming →Data_Structure_I conf:(0.92)
	Minimum support: 0.6 Minimum confidence: 0.9	Best Rules 1. Data_Structure_II → Data_Structure_I conf:(0.93)

After application of Apriori Association Rule algorithm on clustered data using Simple K-means algorithm		
36 courses considered in table 1.	Minimum support: 0.95 Minimum confidence: 0.9	Best rules found: 1. Data_Structure_I=yes →Switching_Theory_&_Logic_Design=yes conf:(1) 2. Switching_Theory_&_Logic_Design =yes →Data_Structure_I =yes conf:(1) 3. Operating_System_I =yes → Switching_Theory_&_Logic_Design =yes conf:(1) 4. Switching_Theory_&_Logic_Design =yes →Operating_System_I =yes conf:(1) 5. Visual_Basic=yes → Switching_Theory_&_Logic_Design =yes conf:(1) 6. Switching_Theory_&_Logic_Design =yes 37 →Visual_Basic=yes conf:(1) 7. C_Programming=yes → Switching_Theory_&_Logic_Design =yes conf:(1) 8. Switching_Theory_&_Logic_Design =yes 37 →C_Programming=yes 37 conf:(1) 9. Java_Programming=yes → Switching_Theory_&_Logic_Design =yes conf:(1) 10. Switching_Theory_&_Logic_Design =yes →Java_Programming=yes conf:(1)

VI. Conclusion and Future Work

As ADTree classification algorithm & Apriori association rule algorithm are the popular technique of data mining, we try to use the combination of these two algorithms to find the best combination of courses in educational system. The advantage of applying this combination is that there is no need to prepare the data as we have to prepare it if we are using only the association rule mining & number of rules is also more. Future work includes the atomization of this combination algorithm.

VII. References

[1] Ye-Zheng Liu, Yuan-Chun Jiang , Xiao Liu , Shan-Lin Yang : "CSMC: A combination strategy for multi-class classification based on multiple association rules" Knowledge-Based Systems 21 (2008) 786–793

[2] Al´ipio M. Jorge, Paulo J. Azevedo:" An experiment with association rules and classification: post-bagging and conviction?" Supported by the POSI/SRI/39630/2001/Class Project

[3] P. R. Pal, R.C. Jain:"Combinatorial Approach of Associative Classification" Int. J. Advanced Networking and Applications

[4] B.Ramasubbareddy, A.Govardhan & A.Ramamohanreddy:" Classification Based on Positive and Negative Association Rules" International Journal of Data Engineering, (IJDE), Volume (2): Issue (2) : 2011 84

[5] *Lili He, Hongtao Bai:"*Aspect Mining Using Clustering and Association Rule Method" IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.2A, February 2006

[6] 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)

[7] Sunita B Aher and Lobo L.M.R.J. Article: A Framework for Recommendation of courses in E-learning System. *International Journal of Computer Applications* 35(4):21-28, December 2011. Published by Foundation of Computer Science, New York, USA ISSN 0975 – 8887

[8] Han,J. and Kamber, M., "Data Mining: Concepts and Techniques", 2nd edition.

[9] Sunita B Aher and Lobo L.M.R.J.: "Preprocessing Technique for Association Rule Based Course Recommendation System in E-learning" selected in ICECT-12, proceeding published by IEEE

Author's Profile

Sunita B Aher
received the B.E degree in Computer Science & Engineering in 2000 from Amravati University, Amravati, India and pursuing the M. E. degree in Computer Science in Walchand Institute of Technology, Solapur, India. She is doing the dissertation work under the guidance of Mr. Lobo l.M.R.J, Associate Professor & Head, Department of IT, Walchand Institute of Technology, Solapur, Maharashtra, India.

Mr. Lobo L.M.R.J.
received the B.E degree in Computer Engineering in 1989 from Shivaji University, Kolhapur, India and the M. Tech degree in Computer and Information Technology in 1997 from IIT, Kharagpur, India. He has registered for Ph.D. in Computer Science and Engineering at SGGS, Nanded of Sant Ramanand Teerth Marathawada University, and Nanded, India.Under the guidance of Dr. R.S. Bichkar. He is presently working as an Associate Professor & Head, Department of IT Walchand Institute of Technology, Solapur, Maharashtra, India. His research interests include Evolutionary Computation, Genetic Algorithms and Data Mining.