International Journal of Information Science and Computing 2(2): December, 2015: pp. 75-84

A Framework for course pathway recommendation system in nonformal E-learning

S. Kavitha and V. Khanaa

Department of Faculty of Computer Applications, Aarupadai veedu Institute of Technology, Chennai, India

Dean - IT, Bharat University, Chennai, India

Corresponding author: kavithasomu_99@yahoo.com

Abstract

Data Mining is the extraction of hidden predictive information from large database which can be used in various commercial applications like bioinformatics, E-commerce etc. Association Rule, classification and clustering are three different algorithms in data mining. Course Recommender System plays an important role in identifying the behavior of students interested in particular set of courses. We collect the data regarding the course enrollment for specific set of data. For collecting this data, we use the learning management system like Moodle. After collecting the data, we apply the different combination of data mining algorithm like classification & association rule algorithm, clustering & association rule algorithm, association rule mining in classified & clustered data, combining clustering & classification algorithm in association rule algorithms or simply the association rule algorithm. Here in this paper we use ADTree classification algorithm, Simple K-means Algorithm & Apriori Association Rule algorithm as different machine learning algorithm. So we propose the five different methods to find the best combination of algorithm in recommending the courses to students in E-learning. We compare the result of this combined approach as well as only the association rule algorithm & present the best combination of algorithm for recommendation of courses in E-learning according to our simulation.

Keywords: Weka, Machine Learning Algorithm, Simple K-meansAlgorithm, ADTree Classification Algorithm, Apriori Association Rule Algorithm, Moodle.

Data mining also known as Knowledge Discovery in Database (KDD) is the extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data. Alternative names to data mining are knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, etc.

Data mining 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 recommend to the student, the best combination of courses in which the students are interested.

The rest of the document is organized as follows. Section 1.1 provides a background of the related research fields covering a brief introduction about each. Section 2 describes the Literature Review. Section 3 discusses the architecture of proposed system. Section 4 & 5 discusses methodology & implementation and best combination of data mining algorithm. The conclusion & future work is presented in Section 6 followed by references.

Background

This research integrates issues from the research field of data mining algorithms such as Classification, clustering & Association Algorithm), Moodle and open source data mining tool, Weka. The following subsections include a brief overview of these topics.

Apriori Association Rule Algorithm

Association rules are used to show the relationship between data items. Association rule generation is usually split up into two separate steps: First, minimum support is applied to find all frequent itemsets in a database. Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

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 [9]. The Apriori association rule algorithm is given in below [9]:

Apriori Association Rule Algorithm

Input **:** *Database of Transactions D= {t_1, t2, …, t_{nj}*
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-
3. Prune candidates who have subsets that are not large.

Count C_i to determine $L_{i;}$

Until no more large itemsets found;

Simple K-means Clustering Algorithms

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different from the objects in another group [8]. Clustering can be considered the most important unsupervised learning technique.

Simple K-means algorithm is a type of unsupervised algorithm in which items are moved among the set of cluster until required set is reached. This algorithm is used to classify the data set, provided the number of cluster is given in prior. This algorithm is iterative in nature.

Algorithm: Simple K-means clustering algorithm Input:

Set of Elements or Database of transaction

D= $\{t_1, t_2, t_3, \ldots, t_n\}$ Number of required Cluster k

Output:

Set of Cluster K

Method:

Make initial guesses for the means $\mathbf{m_1, m_2, ..., m_{k;}}$ Repeat

Assign each element t_i to the cluster having the closest mean.

Calculate the new mean for each cluster. Until there are no changes in any mean

ADTree Classification Algorithm

Classification is a data mining task that maps the data into predefined groups & classes. It is also called as supervised learning.

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.

Learning Management System Moodle

Moodle is an open-source course management learning system to help educators create effective online learning communities [15]. It is also possible to modify the source code of any file of Moodle. It is very easy to add the course in the system. Here are adding 13 course category & near about 82 subjects which are related to Computer Science & Engineering and Information Technology Department. Here we are using the Moodle as it is very user friendly. Also it maintains detailed logs of all activities of students. It stores the record of every click that students make. We can use these logs to find courses in which student are interested. It stores the logs in relational database MYSQL. Moodle is shown in figure 1.

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 [12]. 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 [11]. It is freely available software. Weka has several standard data mining tasks, data preprocessing, clustering, classification, association, visualization, and feature selection. Figure 2 shows Weka 3.5.3 with Explorer window.

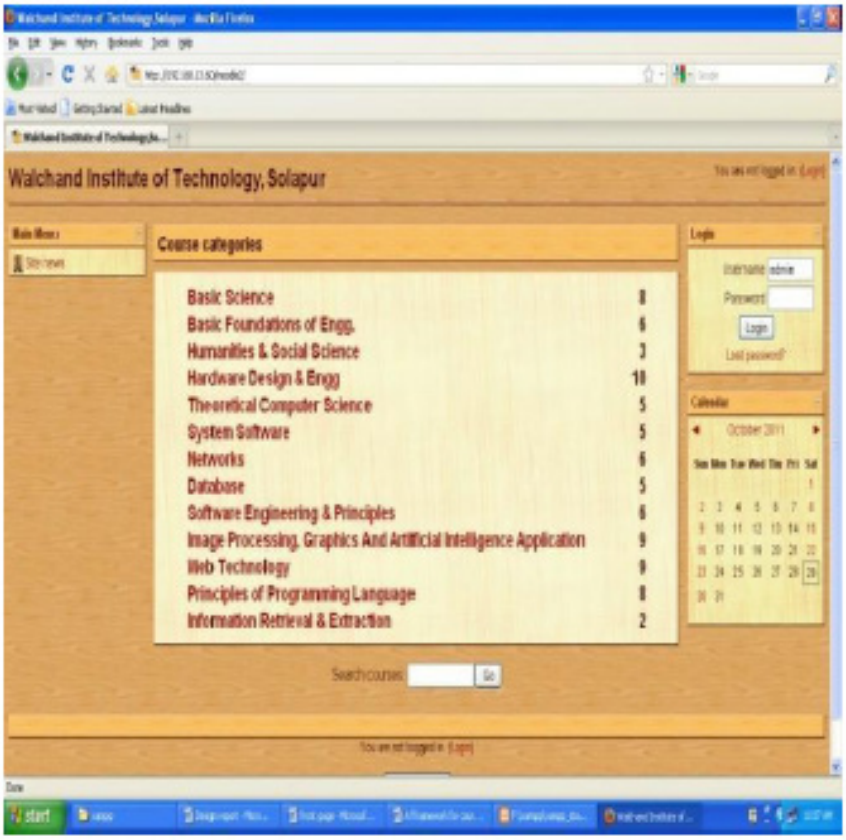

Fig. 1: Moodle

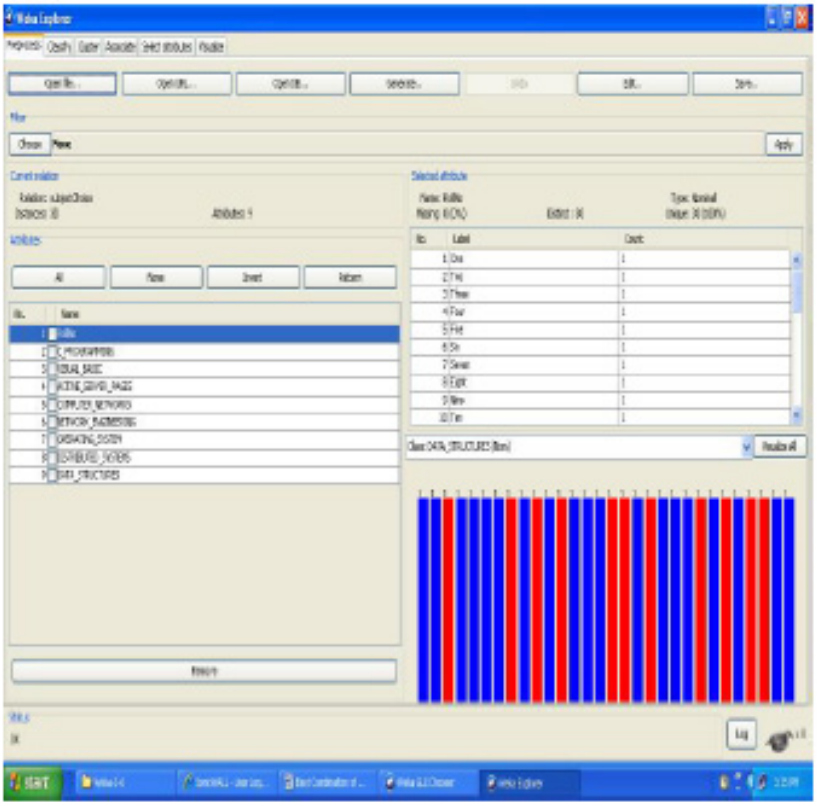

Fig. 2: Weka 3.5.3 with Explorer window

Literature Review

C. Carmona [1], they propose proposed how to use adaptive machine learning algorithms to learn about the student's preferences over time. First they use all the background knowledge available about a particular student to build an initial decision model based on learning styles. This model can then be fine-tuned with the data generated by the student's interactions with the system in order to reflect more accurately his/her current preferences.

Castro [2] provides an up-to-date snapshot of the current state of research and applications of Data mining methods in e-learning. It provides the taxonomy of e-learning problems to which the Data Mining techniques have been applied including, for instance: Students' classification based on their learning performance; detection of irregular learning behaviors; e-learning system navigation and interaction optimization; clustering according to similar e-learning system usage; and systems' adaptability to students' requirements and capacities.

Lili He & Hongtao Bai [3] explained Aspect oriented programming which offers a unique module, an aspect, to encapsulate scattered and tangled code, which made it hopeful to solve the problem of crosscutting concerns. Identification and encapsulation of crosscutting concerns was the key problem in the migration from OO system to AO system. A novel aspect mining method which combined clustering

and association rule technology is provided in this article. Clustering analysis based on the execution traces was provided to find out candidate aspects; while association rule mining based on the execution traces with ordered call was used to find out the crosscuts. Both the aspect code (advice body) and the crosscuts (point cuts) were gotten after the above two processes, which constituted the aspect mining process.

Al³pio Jorge [4] proposed a method for grouping and summarizing large sets of association rules according to the items contained in each rule. They used hierarchical clustering to partition the initial rule set into thematically coherent subsets. This enabled the summarization of the rule set by adequately choosing a representative rule for each subset, and helped in the interactive exploration of the rule model by the user. Rule clusters can also be used to infer novel interest measures for the rules.

B. Ramasubbareddy, A. Govardhan & A. Ramamohanreddy [13] proposed Associative classification which was a classification of a new tuple using association rules. It was a combination of association rule mining and classification. They searched for strong associations between frequent patterns and class labels. The main aim of this paper was to improve accuracy of a classifier. The accuracy can be achieved by producing all types of negative class association rules.

The main hypothesis discussed in the paper [4] was that Web content analysis can be used to improve Web usage mining results. They proposed a system that integrated Web page clustering into log file association mining and used the cluster labels as Web page content indicators. They experimented with several approaches to content clustering, relying on keyword and character n-gram based clustering with different distance measures and parameter settings.

Architecture of Course Recommender System

In Course Recommendation System, we are consider the 13 course category which is shown in following table 1. Under each category there will courses. So there are about 82 subjects.

We have created the student login & gave the access to the student. We have considered the student of two courses Computer Science & Engineering and Information Technology for collecting the data. Student will enroll for those subjects in which they are interested. This data is stored in the moodle database which we use to find out the best combination. The framework for this recommendation system is shown in the figure 3.

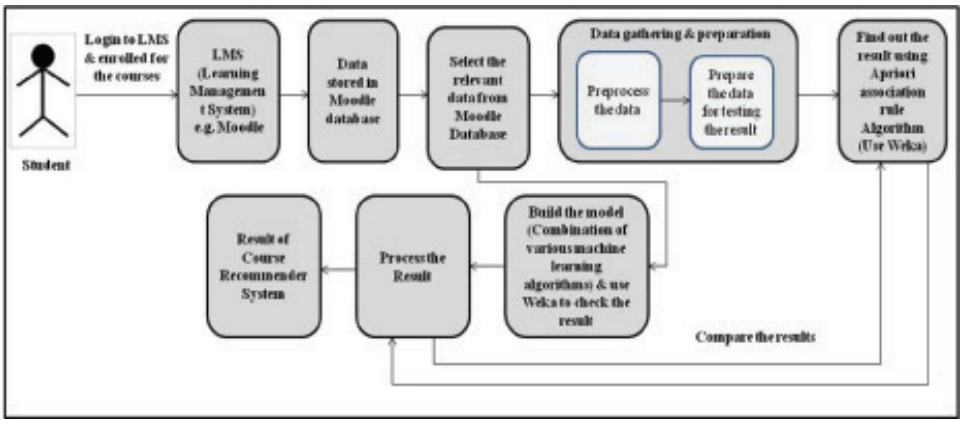

Fig. 3: Course Recommender System in E-learning

Best Combination of Data Mining Algorithm

According to our simulation, the best combination of algorithm is the combination of clustering, classification & association rule mining. Here we preprocess the data to remove the noise or to add the missing value but there is no need to prepare the data as explained subsection 4.1.1. If we are using the Apriori association rule algorithm & increase the support then we are getting the refined association rule but the number of rules, we get, are less & w need to prepare the data also. Association rule obtained using this combination i.e. clustering, classification & association rule algorithm also match with the in general real world interdependencies among the courses.

Figure 4 shows the graph for courses after application of data mining techniques

After Application of Classification Algorithm- ADTree& Association Rule-Apriori Association Rule	
C, VB, ASP, CN, NE, MP, CO, DBE, ADS, OS, DS, FSA, DS-I, SE, STQA	Minimum support: 0.95 Minimum metric <confidence>: 0.9 Best rules found: 1. Operating_System=yes → C_Programming=yes conf:(1) 2. C_Programming=yes → Operating_System=yes conf:(1) 3. Distributed_System=yes → C_Programming=yes conf:(1) 4. C_Programming=yes → Distributed_System=yes conf:(1) 5. Data_Structure-I=yes → C_Programming=yes conf:(1) 6. C_Programming=yes → Data_Structure-I=yes conf:(1) 7. Distributed_System=yes → Operating_System=yes conf:(1) 8. Operating_System=yes → Distributed_System=yes conf:(1) 9. Data_Structure-I=yes → Operating_System=yes conf:(1) 10. Operating_System=yes → Data_Structure-I=yes conf:(1)

After Application of Clustering algorithm-Simple K-means & Association Rule-Apriori Association Rule	
C, VB, ASP, CN, NE, MP, CO, DBE, ADS, OS, DS, FSA, DS-I, SE,	Minimum support: 0.85 Minimum metric <confidence>: 0.9 Best rules found: 1. Visual_Basic=yes → C_Programming=yes conf:(1) 2. C_Programming=yes → Visual_Basic=yes conf:(1)

STQA	3. Active_Server_Pages=yes → C_Programming=yes	conf:(1)
	4. C_Programming=yes → Active_Server_Pages=yes	conf:(1)
	5. Active_Server_Pages=yes → Visual_Basic=yes	conf:(1)
	6. Visual_Basic=yes → Active_Server_Pages=yes	conf:(1)
	7. Visual_Basic=yes Active_Server_Pages=yes → C_Programming=yes	conf:(1)
	8. C_Programming=yes Active_Server_Pages=yes → Visual_Basic=yes	conf:(1)
	9. C_Programming=yes Visual_Basic=yes → Active_Server_Pages=yes	conf:(1)
	10. Active_Server_Pages=yes → C_Programming=yes Visual_Basic=yes	conf:(1)

After Application of Classification algorithm-ADTree, Clustering algorithm-Simple K-means Algorithm & Association Rule-Apriori Association Rule

C, VB,	Minimum support: 0.95	
ASP, CN,	Minimum metric <confidence>: 0.9	
NE, MP,		
CO, DBE,	Best rules found:	
ADS, OS,		
DS, FSA,	1. Visual_Basic=yes → C_Programming=yes	conf:(1)
DS-I, SE,	2. C_Programming=yes → Visual_Basic=yes	conf:(1)
STQA	3. Active_Server_Pages=yes → C_Programming=yes	conf:(1)
	4. C_Programming=yes → Active_Server_Pages=yes	conf:(1)
	5. Operating_System=yes → C_Programming=yes	conf:(1)
	6. C_Programming=yes → Operating_System=yes	conf:(1)
	7. Distributed_System=yes → C_Programming=yes	conf:(1)
	8. C_Programming=yes → Distributed_System=yes	conf:(1)
	9. Finite_State_Automata=no → C_Programming=yes	conf:(1)
	10. C_Programming=yes → Finite_State_Automata=no	conf:(1)

After Application of Clustering algorithm-Simple K-means , Classification Algorithm- ADTree & Association Rule-Apriori Association Rule

C, VB,	Minimum support: 0.95	
ASP, CN,	Minimum metric <confidence>: 0.9	
NE, MP,		
CO, DBE,	Best rules found:	
ADS, OS,		
DS, FSA,	1. Visual_Basic=yes → C_Programming=yes	conf:(1)
DS-I, SE,	2. C_Programming=yes → Visual_Basic=yes	conf:(1)
STQA	3. Active_Server_Pages=yes → C_	conf:(1) Programming = yes

4. C_Programming = yes → Active_Server_Pages = yes conf:(1)
5. Computer_Network = yes → C_Programming = yes conf:(1)
6. C_Programming = yes → Computer_Network = yes conf:(1)
7. Network_Engineering = yes → C_Programming = yes conf:(1)
8. C_Programming = yes → Network_Engineering = yes conf:(1)
9. Operating_System = yes → C_Programming = yes conf:(1)
10. C_Programming = yes → Operating_System = yes conf:(1)

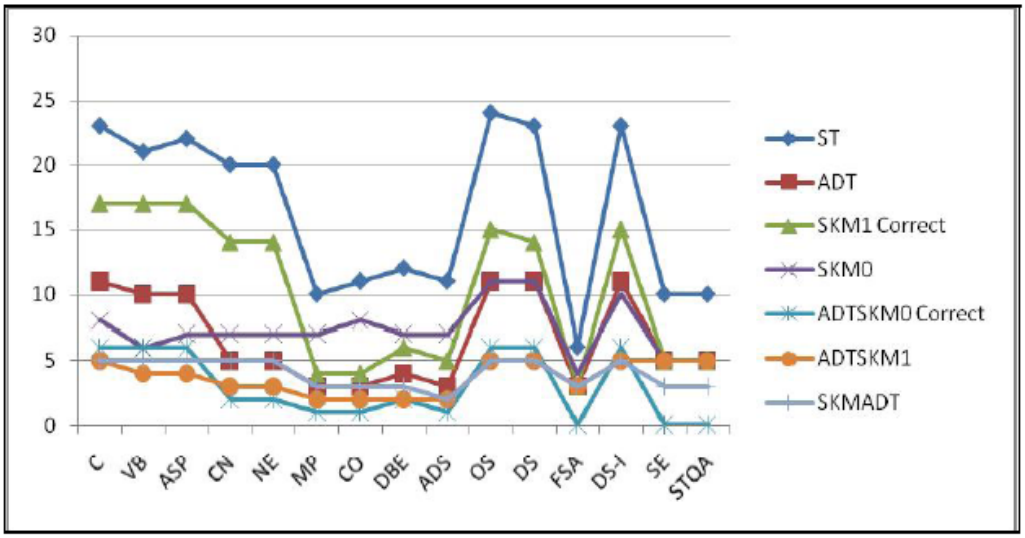

Fig. 4: Graph for courses after application of various data mining algorithms

ADT- Courses after application ADTree classification Algorithm to Sample Table
SKM1 Correct – Courses after application of Simple K-means clustering algorithm (cluster 1 correct result)
SKM0 – Courses after application of Simple K-means clustering algorithm (cluster 0 incorrect result)
ADTSKM0 Correct – Courses after application of Simple K-means algorithm to classified data (Cluster 0 correct result)
ADTSKM0 Correct – Courses after application of Simple K-means algorithm to classified data (Cluster 0 correct result)
SKMADT- courses after application of ADTree classification algorithm to correct (SKM1 Correct) clustered data.

Conclusion and Future Work

In this paper, we compare different combination of data mining algorithms i.e. Classification & association rule algorithm, clustering & association rule algorithm, Association Rule Mining of classified & clustered data, combining Clustering & Classification algorithm into Association Rule algorithm & only association rule algorithms. We consider ADTree classification algorithm, Simple K-means clustering algorithm & Apriori association rule algorithm. We compare the result & found that the combination clustering, classification & association rule algorithm is the best combination. Future work includes the atomization

of this combination algorithm i.e. clustering, classification & association rule algorithm on huge amount of data obtained from Moodle course of a college.

References

C. Carmona, G. Castillo and E. Millán: Discovering Student Preferences in E-learning, EC-TEL07, pp.33-42 (2007)

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.

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

Al¶³pio Jorge:"Hierarchical Clustering for thematic browsing and summarization of large sets of

Association Rules" Supported by the POSI/SRI/39630/2001/Class Project

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 ISSN 0975 – 8887

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

Alaa el-Halees (2009) :"Mining Students Data to Analyze e-Learning Behavior: A Case Study."

Accessed from http://uqu.edu.sa/files2/tiny_mce/plugins/filemanager/ files/30/papers/f158.pdf on 05-03-2012

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

Sunita B Aher and Lobo L.M.R.J.:" Data Preparation Strategy in E-Learning System using Association Rule Algorithm" selected in International Journal of Computer Applications. ISSN 0975 – 8887