International Conference on Advanced Computer Technology (ICACT) 2011
Proceedings published by International Journal of Computer Applications® (IJCA)

A multivariate LMS with –Hybrid Recommender System using Association Rules for Peer Learners

M.Ravichandran
Research Scholar, Department ECE
SathiyabamaUniversity,Chennai

Dr.G.Kulanthaivel
National Institute of Technical Teachers
Training Research Institute , Chennai

ABSTRACT
In eLearning system sharing of event experiencehas attracted the cognitive process of researchers to improve the efficacy of learning . Now a days the higher learning environment losses of face to face interaction with various educators, lectures , facilitators and tutors. Recommender systems are increasingly being used in today's world. Collaborative filtering, together with association rules mining are probably themost widely used methods to implement recommender systems.

In this paper we undertake a review of past research conducted in the area of recommender systems with the focus being the use of association rule mining. We propose a novel methodology that combines the use of association mining with the use of distance metrics such as the Jaccard measure to identify effective e-Learners that belong to the same type to recommend appropriate LE'S to peer learners for the improvement of learning. . Our experimental results on the sample learners profile dataset shows that the use of the Jaccard metric improved the coverage of recommendations over the use of the standard association rule mining method.

KEYWORDS
Datamining,Clustering, Classification, Multivariate analysis , e-Learning. ,Experience sharing.

1. INTRODUCTION
The System architecture of the proposed researchwork has been discussed in section I and section II deals with the eLearning experiences of the experienced learners database extracted from the moodle database of the virtual learning environment. After pre processing for the database information like learner profile activity/assessment/content etc , the normalised dataset has been constructed to apply different data mining techniques for the generation of association rule mining to find out the strong rules with frequently referred or conducted events. Such association rule mining technique has been detailed in section III. In section IV the multivariate analysis is taken against the training dataset of the student database for analysing the various attributes related students activities to improve the selection of rules associated in rule mining using information gain technique that help to construct the decision tree to classify the student. In section V the hybrid recommender system is used to find out the similarity between the learner using jaccard metrics concepts that help the peer learners.

2. System Architecture
The schematic diagram is shown in figure 1. gives clear idea about the total system of recommender system for peer learners. Initially the information is retrieved from the moodle LMS which contains the various users (teacher, students, activities, resources, styles and etc.) about a particular course in the form of raw data. From that the learners activities log, the status of each activities carried out by the individual student for a particular time stamp has been collected and also converted in to the normalized machine learning form of data (Attribute relation file format - oops.style.arff)(Learning Experience).. This normalized file dataset is taken up for experimentation input. In order to perform recommender system for the peer learners the dataset is trained with the attribute selection method, clustering technique and classification method. The attribute selection method is also use well known rankers algorithm (information gain) to get best spilton criterion that will taken up for the decision trees root node. , concept, observe, Experiment) based on the activity time slot data. The new peer learners can easily adapt the suitable activities and referencing with the help of hybrid recommender system using jaccardcoeffient.

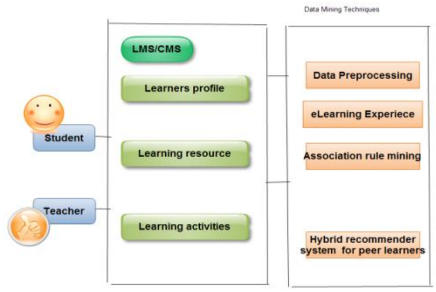


Figure 1. Schematic diagram

3. eLearning Experience
The Experienece sharing is the most common practice and a hot research topic. In elearning especially with the absence of face to-face contact with eductors, lectures, facilitators and tutors capturing and utilizing learning experiences of learners referred as learners experience. According to [1] the three major components of learning paradigm in eLearning paradigm are human, knowledge and technology. The nature of learning process is a process of transfer between tacit and explicit knowledge[2]. Experience is a term which has preoccupied philosophers and which many have avoided[3]. Experience is not a just observation, passive undergoing of something., continuing,complex and meaningful interactions is to the understanding of experience.When applying to learning it is the process, systematic or random of exploring and active or passive cognitive engagement with a domain knowledge with the objectives of gaining skill and wishdom knowledge in the hope of fulfilling the Bloom's Taxonomy of educational objectives[4]. In eLearning , the explicit knowledge is presented to learners in the form of instructional materials ,coursenotes, quizzes etc. due to advances in information communication technologies(ICT) instead the meta information.ie. knowledge of the type of information when it is useful, what to do with it and how to reuse it, is additional knowledge that is vital to learning[5].The eLearning experiences of the experienced learners database extracted from the moodle database of the virtual learning

environment. After pre processing for the database information like learner profile activity/assessment/content etc , the normalised dataset has been constructed to apply different data mining techniques.

4. Association Rule Mining for eLearner's

Association rule mining discovers relationships among various attributes database (learnerid, cia1, cia2, modelexam, ate dance, univexam). An x=>y association rule expresses a close correlation between the above said items in a dataset with values of support and confidence. The confidence of the rule is the percentage of transactions that contains the consequence in transactions that contain the antecedent. The support of the rule is the percentage of transactions that contains both antecedent and consequence in all transactions in the dataset. Association rule mining has been applied to web based education system for building recommender agents that could recommend on line learning activities(Zaiane 2002). Diagnosing student learning problems and offer students advice(Hwang, Hsiao& Tseng2003).Guiding the learners activities automatically and recommending learners materials(LU 2004). Determining which learning materials are the most suitable to recommended to the user(Markellou, Mousourouli, spiros and Tsakalidis 2005). Identifying attributes characterizing patterns of performance disparity between various groups of students(Minaci-Bidgoli,Tan and punch 2004).finding out relationships in learners behaviour patterns (yu,own and lin 2001). Data mining techniques for the generation of association rule mining to find out the strong rules with frequently referred or conducted events. Such association rule mining technique has been detailed in the experiment section of this paper.

5. Multivariate Analysis

An attribute selection measure is a heuristic for selecting the splitting criterion that best separates a give data partition of eLearners dataset of a classlabelled training tuples in to individual classes. We can split dataset in to smaller partitions according to the outcomes of splitting criterion, ideally each partition would be pure. The popular multivariate attribute analysis information gain (Rankers algorithm) is applied to the eLearners dataset to get best split and further decision tree can be generated with the output variable pointed by the multivariate analysis.

6. Hybrid Recommender system

Clustering is the process of grouping objects in to classes of similar objects(Jain, Murty and Flynn 1999). It is unsupervised classification or partitioning patterns in to groups based on their locality and connectivity within an n dimensional space. Here finding clusters of students with in similar learning characteristics and to promote group based collaborative learning(Tang and Mc calla 2005). The K_Means(MacQueen 1967) one of the simplest and most popular algorithm has been used here and it is an algorithm that clusters objects based on attributes in K partitions. In this case our objective is to group students from a specific course object oriented programming in to different clusters depending on the activities done in moodle and the extracted data set stored in the excel sheet. The similarity between learners is identified using jaccard coefficient sim(I,J). D(I,J)=(R+S)/ (Q+R+S).

The measurement calculated from the jaccard coefficient of two learners with high rank value , whose experiences are given as the recommendation for the peer learners.

Experiment

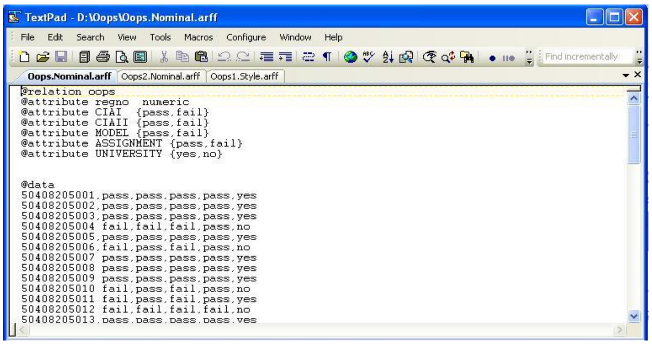
Data mining Approaches for Educational Data

For our research problem, we have taken 66 students of a class of a course "Database Systems". We have organized the course in the virtual learning environment using moodle, the classes and other activities like quiz, test, assignment, group discussion and other things through this moodle. The Snapshot of the moodle course is shown in the Fig 3.

The main activities taken by the students are listed here

1. Student Continual Internal Assessment Test (CIA)
2. Technical Quiz (Quiz)
3. Assignment
4. Attendance
5. University Result

The sample dataset of the course is listed here.

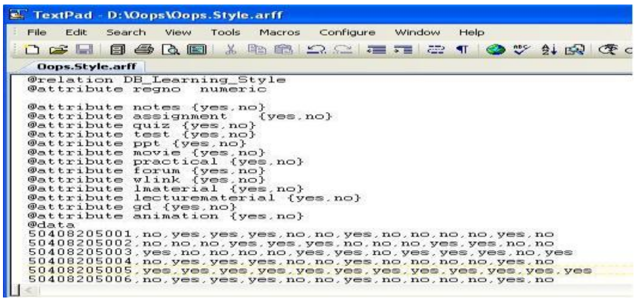


```
@relation oops
@attribute regno  numeric
@attribute CIAI  {pass,fail}
@attribute CIAII {pass,fail}
@attribute MODEL {pass,fail}
@attribute ASSIGNMENT {pass,fail}
@attribute UNIVERSITY {yes,no}


@data
50408205001,pass,pass,pass,pass,yes
50408205002,pass,pass,pass,pass,yes
50408205003,pass,pass,pass,pass,yes
50408205004 fail,fail,fail,pass,no
50408205005,pass,pass,pass,pass,yes
50408205006,fail,pass,fail,pass,no
50408205007 pass,pass,pass,pass,yes
50408205008 pass,pass,pass,pass,yes
50408205009 pass,pass,pass,pass,yes
50408205010 fail,pass,fail,pass,no
50408205011 fail,pass,fail,pass,yes
50408205012 fail,fail,fail,fail,no
50408205013,pass,pass,pass,pass,yes
```

Fig -3 Sample Dataset of a Student Class

The following dataset is used for predicting the learning style as described in the Fig -2 of the learners collected form the Moodle Database Fields which are updated based on the resources referred by the learner and based on the various activities carried out by the students during their learning activities.



```
@relation DB_Learning_Style
@attribute regno  numeric

@attribute notes {yes,no}
@attribute assignment   {yes,no}
@attribute quiz {yes,no}
@attribute test {yes,no}
@attribute ppt {yes,no}
@attribute movie {yes,no}
@attribute practical {yes,no}
@attribute forum {yes,no}
@attribute wlink {yes,no}
@attribute lmaterial {yes,no}
@attribute lecturematerial {yes,no}
@attribute gd {yes,no}
@attribute animation {yes,no}
@data
50408205001,no,yes,yes,yes,no,no,yes,no,no,no,no,yes,no
50408205002,no,no,no,yes,yes,yes,no,no,no,yes,yes,no,no
50408205003,yes,no,no,no,no,yes,yes,no,yes,yes,yes,no,yes
50408205004,no,yes,yes,yes,no,no,yes,no,no,no,no,yes,no
50408205005,yes,yes,yes,yes,yes,yes,yes,yes,yes,yes,yes,yes
50408205006,no,yes,yes,yes,no,no,yes,no,no,no,no,yes,no
```

Fig -4 Sample Dataset of a Student Class

Data Clustering (KSimpleMeans: Square Error Criterion)

The Learning Style of the invidual students is predicted and clustered based on the KSimpleMeans Clustering Technique to cluster the different learning style based on the various learning resources and the activities based on the learning activities of the peer learners.

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2$$

Where 'E' is the sum of the square error for all objects in the given dataset (DB_Learning_Style)

'p' is the point in the space representing the given objects (ie) the various activities tuples of student

'mi' is the mean of cluster Ci.Both 'p' and 'm' are multidimensional array.

We can see in Fig. 5 that there are 4 clusters of students based on Learning Styles. Cluster 0 is characterized by learning style of students as concrete (27%). Cluster 1 is characterized by students as Concept (23%).Cluster 2 is characterized by students with learning style observe. Cluster 3 is classified as Experimenter.

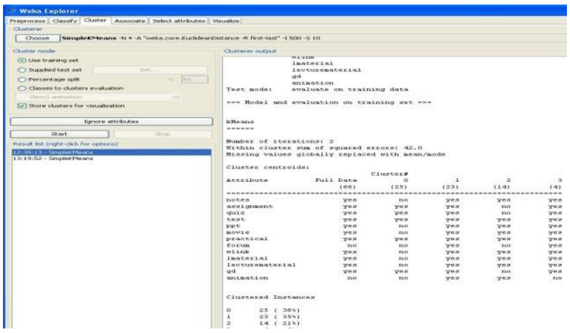


Fig -5 Clustering of Learning Styles

Data Classification-Multivariate analysis (ID3-Decision Tree)

A common application for classifications algorithms often involves decision-based classification and adaptive learning over a training set of a real world databases. The decision tree is a popular utility for implementing such tactics. A decision tree is a decision-modeling tool that graphically displays the classification process of a given input for given output class labels. This paper will discuss the algorithmic induction of decision trees, and how varying methods for optimizing the tree, or pruning tactics, affect the classification accuracy of a testing set of data.

The Effective construction of a decision tree measures such as statistical significance information gain and gini index can be used to access the goodness of split. For our problem the goodness of the split has the root node possesses the highest info gain value calculated from the following along with the learning resource type and learning activities.

$$\mathrm{Info}(D) = -\sum_{J=1}^{m} P_i \log_2(p_i)$$

$$\mathrm{Indo}_A(D) = \sum_{J=1}^{v} |D_j|/|D| * \mathrm{info}(D_j)$$

$$\mathrm{Gain}(A) = \mathrm{info}(D) - \mathrm{info}_A(D)$$

The partitioning the tuble at the node would result in a split that falls below a pre specified threshold of the given subset is halted(Learning Style as the Leaf nodes)

Based on the Resources refereed or accessed by the individual learner, this decision tree is built to prune the learning style of the indented learner.

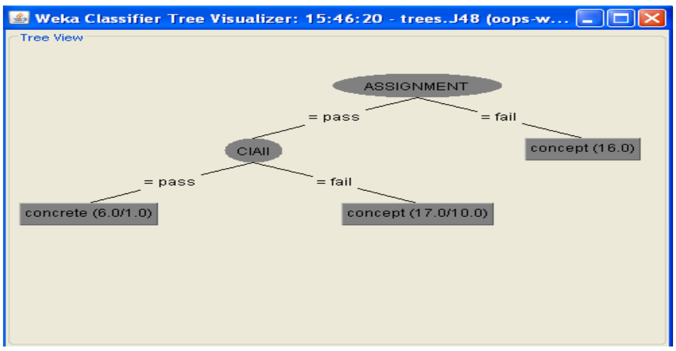


Fig -6 J48 classification Decision Tree

The J48 algorithm is used to characterize students who passed or failed the course. The J48 is an algorithm for generating decision trees and inducing classification rules from the tree. In this case, our objective is to classify students into different groups with equal final result depending on the activities carried out in Moodle. We have executed the j48 with the default parameters over the student_summarization_discretized file (in which the learning style is the last attribute) and k-fold cross validation with k = 3 (the original sample is partitioned into K subsamples, and of the K subsamples, a single sub-sample is retained as the validation data for testing the model, and the remaining K-1 subsamples are used as training data.). On executing the algorithm in Keel, a decision tree is obtained (see Fig. 6) as well as a summary with a number of nodes and a number of leaves on the tree, number and percentage of correctly and incorrectly classified instances.

7. SUMMARY

7.1. Clustering

Clustering high dimensional data is particularly important task in cluster analysis (e-Learning Application requires the analysis of objects containing a large no of features and dimensions. Ex. Various resources of learning materials ,activities and learning styles. This can be clustered by learning styles as per Kolb Taxonomy. The classical partitioning method centroid based technique: KMeans Method with 4 clusters .

Clustered Instances
0 25 (38%) – Learning Style I (Concrete)
1 23 (35%)– Learning Style II (Concept)
2 14 (21%)– Learning Style III (Observe)
3 4 (6%)– Learning Style I (Experiment)

7.2 Multivariate Analysis

The heuristic for selecting the splitting criterion that best separates a given e-learning data partition. The multi variate analysis is done with the help of "GainRationEval" using Rankers algorithms provides the root node selection of our decision tree generation

Ranked attributes:
0.72531 9 wlink
0.72531 1 notes
0.49398 6 movie
0.49398 11lecturematerial
0.49398 10lmaterial
0.35374 8 forum
0.26983 4 test
0.20552 7 practical
0.15883 5 ppt
0.00216 12gd
0.00216 3 quiz
0.00216 2 assignment

Selected attributes: 9,1,6,11,10,8,4,7,5,12,3,2 : 12

The highest rank choosen 0.72531 (The related attribute ranked high is 'wlink' from the Learning Style Database-Oops.style.arff).

3.The scalability of decionson

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances	62	93.9394 %
Incorrectly Classified Instances		6.0606 %
Kappa statistic	0.8791	
Mean absolute error	0.1033	
Root mean squared error	0.2272	
Relative absolute error	20.8179 %	
Root relative squared error	45.6321 %	
Total Number of Instances	66	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0.111	0.882	1	0.938	0.957	yes
	0.889	0	1	0.889	0.941	0.957	no
Weighted Avg.	0.939	0.051	0.947	0.939	0.94	0.957	

=== Confusion Matrix ===

a b<-- classified as
30 0 | a = yes
4 32 | b = no

8. CONCLUSIONS

In this work we have shown how useful the application of data mining techniques in learning management systems can be for web based instructors and students. Although we have shown these techniques separately, they can also be applied together in order to obtain interesting information in a more efficient and faster way. First, instructors can use visualization techniques to obtain a general view of the student's usage data. And for example, if they find something strange or irregular in the plots, then they can obtain more detailed information about these events by viewing statistical values. Or, if they find some similar groups of students in graphs, then they can apply clustering techniques in order to obtain the exact groups students can be divided into. And these groups can also be used to create a classifier in order to classify students.

At present, we are developing a specific Moodle data mining tool oriented for use e-learning instructors which would obviate the need for LMS administrators to help these instructors to pre-process or apply mining techniques. It has an intuitive and user-friendly interface to do data mining and automatically pre-processes Moodle data, making it easier to configure and execute data mining techniques due to its parameter-free data mining algorithms. Likewise, they can directly apply feedback and results obtained by data mining into Moodle courses.

9. REFERENCES

[1] A TW sdiq and Xueli "Learning Experience can e-Learning technology be used as a vehicle? In 4th international conference in e-learning 2009 pp 32-39

[2] M. Polany The tacit dimension New york 1967

[3]I Nonaka "A dynamic theory of organizatinal knowledge creation pp 114-37 1994.

[4] D Boudcohen R walker D using Experience for learning open university press 1993

[5] D R Krathwohl A revision of Blooms Taxonomy an overview 2002

[6] O Marjanovie "sharing and reusing learning experiences the knowledge management perspective. 5th IEEE ICALT 05 707-709 2005.

[7] Cristo Bal Romero, SebastiaN Ventura 2008, Data Mining in course management systems : study and tutorial. Computers and education Elsevier

[8] Damjanovic, V., Kravcik, M. &Devedzic, V. (2005). An approach to the realization of personalized adaptation by using eq agent system. In Paper presented at workshop on personalization on the semantic Web: Held in conjunction with the 10th international conference on user modeling (UM'05), 24–30 July, Edinburgh, UK.,

[9] Damjanovic, V., Kravcik, M. &Devedzic, V. (2005). An approach to the realization of personalized adaptation by using eq agent system. In Paper presented at workshop on personalization on the semantic Web: Held in conjunction with the 10th international conference on user modeling (UM'05), 24–30 July, Edinburgh, UK.,