

E-Learning Retrieval System Through Advanced Data Mining Clustering Technique

V.Venkateswaran¹, R.Ramakrishnan², A. Ramalingam³

^{2,3}Assistant Professor, Department of M.C.A

^{1,2}Sri Mnakula Vinayagar Engineering College

Abstract : The e-Learning provides the Learners to interact each other as well as with teachers. The new technology virtual learning system (VLE) is used to interact between the student and teacher. The vast improvement in the field of information technique leads to share of information resources. The development in e-Learning system which arise a new way of distributing knowledge around the world without any restrictions. The e-Learning system has huge volume of recourse sharing at various places. The e-Learning retrieval system contains various information's that can be shared and communicate between the Learners in a course and store Resources from the various server can be retrieved. So there is a need to measure the sharing of resource using specific technique. The data mining clustering technique is used to retrieve the recourse of the e-learning system and measure the recourse shared.

Keywords: LMV, VLE, e-Learning, Data Mining, KDD; EDM

I. INTRODUCTION

1.1 E-Learning Systems

The e-Learning system consists of acquiring knowledge and recourse sharing. The e-Learning system can enable better allocation of resources and organize the learning processes in order to improve the learning experience of the student as well as increase their learning Knowledge [1].

The e-Learning systems increase the sharing of recourses providing recourse material to learners. The recourse is being centralized using server is locate worldwide. The e-Learning recourse in much discipline the related recourse are grouped based on the course as per the leaner wish. The learner can select the course material as Text, PPT, video which is related to the course. The e-learning system provides fast easy and efficient of recourse material can retrieved by data mining clustering technique.

1.2 Theory of e-Learning

The e-Learning system accumulates a vast amount of information which is valuable for learner development. The learning management systems accumulate a great deal of student activities can record whatever student activities are involved such as reading writing talking test performing various task and even communicating with other. The e-Learning system facilities of communication between educators, sharing resources, producing content material, preparing assignment, enabling synchronous learning with forms chat, new services, of e-Learning system which improves the knowledge of the learner accessing open resources with waste of resources.

II. Applying Data Mining Technique

Data mining or knowledge discovery in database (KDD) is the automatic extraction of implicit and interacting. Data mining is a multidisciplinary area in which several computing paradigms converge decision tree construction, rule induction, K-means, SVM, Apriori, Page Rank, AdaBoost, KNN, Naïve Bayss, And CART algorithms etc And some of the most useful data mining task and method are static's Visualization, clustering classification, association rule mining ,sequential pattern mining text mining, etc.

2.1 Educational Data Mining

Due to the large quantities of data in e-learning systems, it is very difficult for educators to analyze them manually. Educational Data Mining EDM the area of scientific inquiry centered on the development of methods for making discoveries within the unique kinds of data. That come from educational settings, and using those methods to better understand students and settings which they learn in Educational Data Mining is an emerging discipline, with developing methods for exploring unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in as defined by The Educational Data Mining community.

The Educational Data Mining providing research efforts in the field, review of the many applications of Data Mining to e-learning. The review of the history and current trends of EDM review of most relevant studies carried out in this field. Data mining techniques relation with the fields of Artificial Intelligence (AI) and Machine Learning (ML) have been highlighted in many researches are classified into four main areas 1) improving student models, that provide detailed information about a student's characteristics; 2) discovering models of the knowledge structure of the domain 3) studying the pedagogical support provided by learning software 4) scientific discovery about learning and learners. The first three categories are universal across different fields of data mining are categories particularly related to educational data mining[2].

- Statistics and visualization
- Web mining
- Text mining
- Prediction
- Clustering
- Relationship mining
- Distillation of data for human judgment
- Discovery with models

2.2 Using data mining Technique

In E-learning, clustering has been used for finding clusters of students with similar learning characteristics and to promote group-based learning and to provide learner diagnosis. Rapid Miner system has several clustering algorithms available. The K-Means has been used here. Clustering techniques apply when the instances of data are to be divided into natural groups. In k-means algorithm clusters are specified in advance prior to application of the algorithm provides a good review of different data mining clustering techniques.

The data mining technique is used in e-learning system which can able to download large volume of resources. The clustering technique is one of the data mining technique is used in the e-learning system for probably and distribution.

III. Concept Of Clustering

The Clustering is defined as the division of data in groups of similar object. The clustering is used for searching the relevant data as the linear request when the resource is not found in the related server. The clustering also helps to gather to relevant resource from the various servers with fast speed.

2.1 Clustering Technique

The clustering technique is broadly classified into two partitioning, Hierarchical clustering it sub divided into agglomerative and divisive. The hierarchical algorithms build clusters gradually partitioning algorithm learn cluster directly. They either try to discover cluster by iteratively relocating points between subset, or try to identify cluster as areas highly populated with data.

Partitioning relocation methods are further categorized into probabilistic clustering K-medoids methods. Such methods concentrate on how well points fit into their clusters and tend to build cluster of proper convex shapes.

Partitioning algorithm of the second type is surveyed in the section density-based partitioning. They try to discover dense connected components of the data, which are flexible in term of shape .density –based connectivity is used in the algorithm. They are less sensitive to outliers and can discover clusters of irregular shapes. They usually work with low dimensional data of numerically attributes, know as spatial data. Spatial objects could include not only points also extended objects[18].

IV. K-Mean Clustering Technique

The k-means clustering determine number of cluster k and we assume the center of these clusters can take any random object as the initial centroids. Then the k-means algorithm will do the three steps below until convergence. Iterate stable (=no object move group): 1. determine the centroid coordinate 2. Determine the object based on minimum distance.



Figure 3: K-Mean Clustering Process

3.1 Distance calculation

The distance is the most commonly used of Euclidean distance. In most case when people said about distance they will refer to Euclidean distance. Euclidean distance or simply “distance examines the root of square difference between coordinates of a pair of objects.

For example :

Features ^k

	Cost	Time	Weight	Incentive
Object A	0	3	4	5
Object B	7	6	3	-1

Points a has coordinate (0, 3, 4, 5) and point B has coordinate (7, 6, 3, -1). The Euclidean Distance between point A and B is

$$d_{BA} = \sqrt{(0-7)^2 + (3-6)^2 + (4-3)^2 + (5+1)^2} = \sqrt{49+9+1+36} = 9.747$$

Euclidian distance is a special case of Minkowski distance with $\lambda = 2$

V. K-MEANS ALGORITHM

k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. A set of observations (x₁, x₂, ..., x_n), where each observation is a d-dimensional real vector, k-means clustering aims to partition the observations into k sets (k ≤ n) S = {S₁, S₂, ..., S_k} so as to minimize the within-

$$\arg \min_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

cluster sum of squares

where μ_i is the mean of points in S_i.

Classification - Tree Induction Results

The Tree Induction algorithm used to predict students’ potential performance was focused on using student assessment and e-Learning usage. Pessimistic pruning was applied to insure that the expected confidence levels obtained for predictions on the training data are similar to actual confidence levels obtained from unseen data. A set of rules was generated from the decision tree operator, showing interesting information about the supervised classification of the students. These rules classify at least three main categories of students: students with a low number of assignments are classified as FAIL; students with medium number of assignments quizzes are classified as FAIL or PASS depending on their quiz score, and students with a high number of assignments are classified as FAIL, PASS or EXCELLENT depending on number of actions in the e-Learning system etc. The rules generated from the Tree Induction Algorithm operator.

4.1 K-Means Results

The educator can use the cluster centroid K-Means results as presented in order to group Students into three types of students: very active students (cluster 0), active students (cluster 1) and nonnative students (2). This information helps educator to group students for working together in Collaborative activities. Students were divided in 3 groups based on their activities done in e-Learning Cluster 0 is characterized by most active students in e-Learning, with high assignment number, which participated to the online quiz and have a moderately number of discussion and forum read; Cluster 1 is characterized by moderately active students in e-Learning, with moderately assignment number, who participated or not to the online quiz in e-Learning with a low number of actions in the system.

4.2 Implementation of k-means

The k-means only allow numerical values for attributes in case it may be necessary to convert the data set into the standard spreadsheet format and convert categorical attributes to binary. It may also be necessary to normalize the value of attributes that are measured on substantially different scales. This K-mean algorithm automatically handles a mixture of categorical and numerical attributes. Furthermore the algorithm automatically normalizes numerical attributes when doing distance computations. The simple k-mean algorithm uses Euclidean distance measure to compute distance between instance and clusters.

To perform clustering select the cluster and clustering algorithm In this case we select ;simple K-means te value is used to generate a random number which in tirm used for making the initial assignment of instance to cluster. In general k-means is quite sensitive to how cluster are initially assigned. of instance to cluster.

Once the option have been specified , we can run the clustering algorithm. The result show the centriod of each cluster as well as statistics on the number and percentage of instance assigned to different cluster. The centroids can be used to characterize the cluster in through visualization cluster assignment. We can choose the cluster number and any of the other attributes for each of the three different dimensions available.

VI. DISCUSSION

Several data mining techniques such as: Attribute Clustering (K-Means), Classification, Association Mining (Apriori, FPGrowth, Create Association Rule, GSP) were applied to e-Learning summarization tables. By applying clustering methods, the goal was to split data set in groups of data points that naturally group together. Student actions were clustered together in order to investigate patterns of students behavior in the e-Learning System. KMeans algorithm was used to define clusters, which starts with no prior knowledge about groups in the data. By using prediction techniques, the goal was to develop a model that can infer predicted variables from predictors' variables. Inductive Decision tree algorithm was selected as classification method. The predicted variable was the categorical variable Final Mark. The target was to define variables that significantly affect in the Final Mark. By applying ARM, the goal was to discover relationships between variables. FPGrowth, Create Association Rule and APRIORI algorithms were selected as association rule mining techniques. The rules obtained can be explained in

the form that if some set of variable values is found, another variable will have a great chance to have a specific value.

VII. CONCLUSIONS

In this research, a data mining model for e-Learning data was proposed based on several techniques Attribute Clustering (KMeans), Classification , Association Mining (Apriori, FPGrowth, Create Association Rule, GSP) was proposed. This educational data mining work allowed identifying and locating information about E-learning processes that need improvements or those that perform very well and could be used as good examples. The educational data mining investigated in this research allows analyzing and better understanding the learning and teaching processes by applying data mining techniques. The experimental results have shown that the data mining model presented was able to obtain comprehensible, actionable and logical feedback from the LMS data describing students' learning behavior patterns. This work concentrated on the overall LMS performance at Epoka University and the mining Process of e-Learning data. Mining the e-Learning data allowed identifying the most effective ways to the teaching process that can be used to enhance the education process. To further test the effectiveness of the proposed model and to increase the generality of this research, more extensive experiments should be conducted by using larger LMS data sets.

A reference list **MUST** be included using the following information as a guide. Only *cited* text references are included. Each reference is referred to in the text by a number enclosed in a square bracket (i.e., [3]). References **must be numbered and ordered according to where they are first mentioned in the paper**, NOT alphabetically.

References

- [1] Nichols and schedule "theory for e-Learning Educational technology and soc., Mvol,6,no,2,pp,1-10,2003
- [2] MCSER-Mediterranean Center of Social and ducational Research Rome, Italy, 2011
- [3] Romero, C., Ventura, S.(2010), Educational Data Mining: A Review of the State-of-the-Art. *IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews*.
- [4] Uribe-Tirado, A., Melgar-Estrada, L.-M., & Bornacelly-Castro, J.-A. (2007). Moodle learning management system as a tool for information, documentation, and knowledge management by research groups. *Profesional de la Información*.
- [5] Silva, D., & Vieira, M.(2002). Using data warehouse and data mining resources for ongoing assessment of distance learning. In Proceedings of the 2002 IEEE Conference on Advanced Learning Technologies Kazan, Russia.
- [6] Sison, R., & Shimura, M. ,(1998). Student Modelling and Machine Learning. *International Journal of Artificial Intelligence in Education*.
- [7] Srikant, R., & Agrawal, R. (1996). Mining Sequential Patterns: Generalizations and Performance Improvements. In P. M. G. Apers, M. Bouzeghoub, & G. Gardarin (Eds.) *Proc. 5th Int. Conf. Extending Database Technology,EDBT*, vol. 1057, (pp. 3–17). Springer-Verlag.

- [8] Srivastava J., Cooley R., Deshpande M., & Tan P. (2000, January) Web Usage Mining: Discovery and Applications of usage Patterns form Web Data, *SIGKDD Explorations*.
- [9] Talavera, L., & Gaudioso E. (2004). Mining student data to characterize similar behavior groups in unstructured collaboration spaces. *Workshop on Artificial Intelligence in CSCL. ECAI*.
- [10] Tang, T.Y., McCalla, G. , (2005). Smart Recommendation for an Evolving e-Learning System: Architecture and Experiment. *International Journal on e-Learning*.
- [11] Tsai, C. J., Tseng, S. S., & Lin, C. Y. (2001). A Two-Phase Fuzzy Mining and Learning Algorithm for Adaptive Learning Environment. In ICCS '01: Proceedings of the International Conference on Computational Science- Part London, UK: Springer-Verlag Wang, F., & Shao, H. (2004).
- [13] Effective personalized recommendation based on time-framed navigation clustering and association mining. *Expert Systems with Applications* 27(3). Wu X., (2004).
- [14] Data Mining: An Artificial Intelligence Perspective. The IEEE Intelligent Informatics Bulletin. Yoo, J., Yoo, S., Lance, C., Hankins, J. (2006).
- [15] Student Progress Monitoring Tool Using Treeview. In The 37th Technical Symposium on Computer Science Education., Houston, USA :ACM Press. Zaïane, O. , & Luo J., (2001).
- [16] Web usage mining for a better web-based learning Environment. Proceedings of conference on advanced technology for education, Banff, Alberta . Xu, R., & Wunsch, D. I. I. (2005).