Cluster Analysis of Behavior of E-learners

Mandeep Kaur, Kewal krishan

Abstract- E-learning is a modern way of learning in which teachers and students don't have actual contact. It's a web based or online learning in which e-learners enroll in educational courses and can learn online. Unlike traditional classrooms in which some students don't put their queries to teachers due to hesitation e-learners can put any kind of queries to teachers because they don't have face to face contact. In E-learning system students have different kind of behavior. Though e-learning courses are designed on the basis of "same content fits all" yet students feel difficulty because every students' learning ability depends upon their individual learning ability. This study proposes the analysis of students' behavior using data mining tools and techniques. Classification and clustering techniques are used to analyze the relationship between usage of courses and performance of students. Students' performance depends upon their grades, how much time they spend in learning, usage of courses as well as richness of course quality. The study uses data from previous approach, E-learning data from Greek University. This paper uses same approach with different data mining tools and techniques.

Keywords- E-learning, data mining, classification, clustering.

I. INTRODUCTION

In the modern world, popularity of personal computers and internet has significant trend, therefore the web-based education or E-learning have grown exponentially in the last years. E-learning is a new education concept different from old educational concept, it deliver digital content and provides learner-orient environment for the teachers and students. E-learning gives a condition for us to realize the life-long learning principle and help us to build a more real learning society. The learning process needs tools and techniques to present the knowledge, interact with it and shares it with others. In this context e-learning become an important tool to support the learning system to achieve its goals. These technologies are used to create and deliver individualized, comprehensive, dynamic learning content that facilitates learning, anytime and anywhere. Students are able to participate in online learning using PC's, mobile phones and other handheld devices. E-learning has many definitions it includes:

- ☐ Online learning/education
- ☐ Distance education/learning
- ☐ Technology based training
- ☐ Web based learning/training
- ☐ Computer based training/ learning from a CD-ROM

Mandeep Kaur M.tech in Information Technology, Lovely professional University, Phagwara, India.

Kewal krishan, Department of CSE/IT , Lovely professional University, Phagwara, India.

E-learning systems have same contents for all the students who have enrolled in courses but every student have their individual learning capability so in this case some students feel difficulty to perform well. Students' performance depends upon their grades. This study focuses on the cluster analysis of students data on the basis of their grades.

1.1 Cluster analysis

Cluster analysis is the statistical methods of partitioning the data into homogeneous classes for classify the data. It divides data into meaningful or useful groups known as clusters. In data mining, clustering is data analysis tool for solving various classification problems. It results into classification of set of observations into two or more groups so that the degree of association is strong between same group or cluster and weak between members of different group. Various data mining clustering algorithms are available like k-means clustering, hierarchical clustering, and fuzzy c-means clustering algorithm. This study uses the E-learning data from Greek university and follows the previous approach with different tools and techniques.

The objective of this paper is to analyze the log file of students data of Greek university then preprocess it. Various index and metric computations are used (Valsamidis, Kazanidis, Kontogiannis, & Karakos, 2011).

In the end classification and clustering of data is used with different tools and techniques of data mining to analyze the comparative results. The paper initially makes a literature review and follows with the related work and comparative study and then conclusion.

II. LITERATURE REVIEW

Traditionally, statistical analysis is used to analyze the behavior of students like students' grades and performance. But nowadays, some teachers don't have knowledge of statistics. Cluster analysis is the technique for teachers who have no statistical background. [1]. E-learning systems are designed for students as well as workers to learn online at anytime, anywhere. Students have their own web portfolios which comprises their learning performance and online activities. The learning portfolio analysis system helps teachers to examine students' behavior, motives and achievements. Different students have different behavior and interests. The clustering of the students is done on the basis of their performance. Their performance helps us to analyze their behavior[2]. In e-learning there are various kind of learners named regular, worker, casual, bad and absent. Various fuzzy clustering techniques are used such as fuzzy c-means and kernelized fuzzy c-means to find learners'

Cluster Analysis of Behavior of E-learners

categories and deals with the

work how to return bad students to regular ones. Expectation Maximization algorithm is used to group users into different clusters.[3]. E-learning systems are of different types which are usually known as Learning Management Systems (LMS). These LMSs proved very helpful tool for students as teaching-learning process (Chellatamilan T & Suresh R.M, 2011). Students can enroll in various courses using LMSs.

It is investigated that if employee of an IT industry and student of a college are in best of their behavioral attitudes in their domain, they produce the best results. clustering techniques are great use in organizations and educational institutes where each employee and student have different behavior so to improve their performance and work efficiency they should given work according to their behavior which leads to best results (Charu Nath, 2011).

Various approaches use existing techniques in different way to perform e-learning. The study used metrics and index to measure the web applications like qualitative and quantitative measurements. The study describe the usage of various data mining techniques such as classification, clustering and association to analyze log files of e-learning platform and deduce conclusion (Ioannis Kazanidis and Alexandros Karakos, 2011.

III. RELATED WORK

In this section we use data mining techniques to analyze the behavior of students using e-learning data from Greek university. Classification and clustering is done using different approach by taking students' Grade as a class. The methodology consists of five steps:

[1] Logging the data.

[2] Data preprocessing.

improve the contents of courses their richness must be known to us.

Classifier algorithm classifies courses in three stages: The first stage aims to identify richness of educational content. Enrichment metric is used and courses classify into high enrichment value and low enrichment value indicates rich and poor content respectively. The courses whose Enrichment is higher than the average E-nrichment of all courses are characterized as high Enrichment courses while the others as low Enrichment courses. This stage results into two clusters of courses courses with rich and courses with poor educational content respectively.

The second stage of algorithm identifies Homogeneity value, how often course information is added or updated by educators. The higher the Homogeneity value the more frequently the course updates or the more dynamic the course content, depending on Enrichment value. The lower Homogeneity value results in static content or of poor content updates. The classification of the courses in this stage depends on both the average Enrichment value and the average Homogeneity value of the high and low Enrichment clusters accordingly. Therefore, this stage of the algorithm results in four clusters.

In third algorithm take UPCS value in order to check whether users find each course useful or not. Therefore, the previous four clusters were further split into high and low UPCS courses according to UPCS average value results into eight clusters. This paper classifies these online courses

using Decision tree classifier algorithm.

[3] Index computation. [4] Metrics calculation. [5] Classification.

[6] Clustering.

6. **Clustering:** In clustering section, Hierarchical agglomerative

clustering is used to cluster the courses taking grade as class attribute. The similar items group into one cluster and different in other. This study clusters the courses in two groups: high activity and low activity.

1. **Logging the data:** In this step log files are analyzed. These log

files stores users' information.

2. **Data preprocessing:** This step used for removal of noise such as missing values and outlier detection from data. In this section outlier detection of log files is done to filter the log files for data analysis. The produced log file is filtered and include following fields:

CourseID- identification string of each field,

SessionID- identification string of each session,

Page Uniform Resourse Locator (URL) - request of each page of E-learning platform that user visited.

3. **Index Computation:** Indexes such as Sessions, Pages, Unique pages, UPCS are used for the facilitation of the course usage evaluation.

4. **Metrics Calculation:** In this step two metrics are calculated: Enrichment and homogeneity. We use two metrics for evaluation of course usage: metric quality which is mean of Enrichment and Homogeniety and metric final which is product of quality with UPSC. These two metrics used to classify and group the courses depends upon their usage. Initially we use UPSC for evaluation of courses, it is quantitative index. The courses with high value of UPSC are popular among students. In some cases where courses have same UPSC value, we use one qualitative metric: Quality, which combine the Enrichment and Homogenity with equal weights. The final result is derived from product of Quality with UPSC.

5. **Classification:** In this section classifier algorithm is used to classify online courses according to their metrics. In order to

IV.CONCLUSION

The study finds that there is relation between course usage and students' grades. The students have fewer grades due to less richness of course content as well as with less frequent updating course contents. So, due to this low activity of course usage the students' grades are affected. The proposed work is used with different tools and techniques of data mining to analyze the relation between course contents and students' grades. Classification of courses categorized the rich content courses from poor content courses, with the help of which authors can improve their course content.

REFERENCES

[1] Chien-Ming Chen, Chao-Yi Li, Te-Yi Chan, Bin-Shyan Jong, and Tsong Wuu Lin, "Diagnosis of Students' Online Learning Portfolios", ASEE/IEEE Frontiers in Education Conference, October 10 – 13,2007, 37th.

[2] Chien-Ming Chen, Cheng-Hao Ma, Bin-Shyan Jong, Yen-Teh Hsia and TsongWuu Lin, "Using Data Mining to Discover the Correlation between Web Learning Portfolios and Achievements", 38th ASEE/IEEE Frontiers in Education Conference, October 22 – 25, 2008, Saratoga Springs, NY.

[3] Mofreh A. Hogo, "Evaluation of E-Learners Behaviour using Different Fuzzy Clustering Models: A Comparative Study", (IJCSIS) International Journal of Computer Science a nd Information Security, Vol. 7, No. 2, 2010.

[4] Felix Castro, Alfredo Vellido, Àngela Nebot, and Francisco Mugica, "Applying Data Mining Techniques to e-Learning Problems".

[5] Chun-Hsiung Lee, Gwo-Guang Lee ,Yungho Leu, "Analysis on the Adaptive Scaffolding Learning Path and the Learning Performance of e-Learning", Issue 4, Volume 5, April 2008.

[6] Stavros Valsamidis and Sotirios Kontogiannis,"E-Learning Platform Usage Analysis", Interdisciplinary Journal of E-Learning and Learning Objects Volume 7, 2011.