IJICTR International Journal of Information & Communication Technology Research ITRC

Volume 4- Number 2- March 2012 (71-81)

Technical Note

Data Mining in the E-Learning Systems: A Virtual University Case Study

Behrouz Minaei-Bidgoli
Dept. of Computer Engineering
Iran University of Science & Technology Tehran, Iran
b_minaei@iust.ac.ir

Seyed Hassan Hani
Dept. of Computer Engineering
University Of Qom
Qom, Iran
sh-hani@qom.ac.ir

Vahid Ghanbari
Dept. of Computer Engineering
University Of Qom
Qom, Iran
vghanbari@stu.qom.ac.ir

Received: September 9, 2011-Accepted: December 31, 2011

***Abstract*— Virtual learning environments provide the opportunity for the students to learn educational materials in educational institutions from different parts and places and with no requirement to physical presence. Over the time, a host of different information related to students, the content of teaching and learning and the interactions among them are recorded in virtual learning systems database that one of the most important of such information is how the students use the system. In this article, we will review the articles that have paid attention to this subject and a classification of the performed activities in this field and would be provided; in continue, some examples of these activities have been applied to the data from Imam Khomeini virtual university.**

Keywords- Data mining; Educational systems; Web mining; Web-based educational systems

I. Introduction

Over the past three decades, e-learning has been widely used among human populations. The wave of investments in this field in recent years has witnessed tremendous growth and now many institutions are working and providing services in this field area. In the past decade, data mining in education and learning systems has come to the view as a growing field of scientific research related to computer science that deals with development of methods for exploring and extracting knowledge from the unique data of educational systems. Most of the existing systems in e-learning centers use to maintain information about the activities and interactions of students engaged in the learning environment, but this unstructured information is far less effective for the teachers and administrators of these centers due to high volume, weakness and lack of analysis and reporting tools. Data mining area and extraction knowledge from database are capable of proper performance to confront these problems, and this caused that so much research have been performed in the area of deployment of the tools and data mining techniques and knowledge extraction in the training and learning systems in the past ten years. Up to now, some articles have been written to review the activities carried out in the field of educational data mining that some of them can be mentioned including [1, 2, 3, 4]. The current study intends to provide a more comprehensive review of activities conducted in this area from the beginning up to the present time by totaling the mentioned papers as well as reviewing newer articles, and study the implementing a few of these methods by using the step-by-step method. It is arranged in the following way: Section 2 deals with reviewing important articles in the field of data mining in e-learning and classify them and the amount of published articles in each of the categories are examined. The third section deals with investigating the status of conducted researches in the field of educational data mining and concentration on different issues in various time periods are examined. Section 4 describes the general process of applying data mining to e-learning data. Then details the preprocessing step necessary for adapting the data to the appropriate format. In the following some of the mentioned methods will be applied on the data of Imam Khomeini virtual University. Finally, the conclusions and further research are outlined.

II. CONDUCTED ACTIVITIES IN THE FIELD OF DATA MINING IN E-LEARNING

Activities and research conducted in this area are divided into the following groups:

- statistical analysis and visualization
- Web mining of information related to the students' use of virtual learning environment
- Text mining

Table 1 and Fig. 1 show briefly the performed activities extent in this area:

In continue, we will further study the articles in various areas:

Table. 1 The performed studies extent in the fields of educational data mining

Branch	Percentage
Statistical analysis and visual modeling	**32.2**
Statistical analysis	20.34
Visualization	11.86
Web mining of information related to the students' use of virtual learning environment	**67.8**
Extraction of the association rules	16.95
Classification	11.86
Development of Tools	11.86
Extraction of sequential patterns	10.17
Clustering	8.47
Outlier analysis	8.47
Social network analysis	3.39

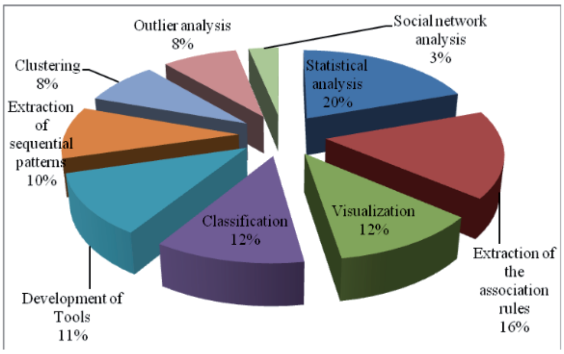

Fig. 1 The performed studies extent in the fields of educational data mining

A. *Statistical analysis*

Statistical methods can be used to summarize and express the descriptive characteristics of a data set. Also, using these methods, modeling the existing trends in data and deduction of processes and latent patterns can be dealt with. Different statistical analysis methods have the ability to be used on data in education and learning systems. Among these methods, descriptive analyzes such as mean, median, standard deviation, frequency tables, histograms, etc can be noted. Also, the deductive advanced techniques such as correlation analysis, regression, statistical hypothesis test and the time sequences have the capabilities to be used on this information.

Usually, the information related to how students use e-learning systems is the beginning to apply these methods. This information can be analyzed through standard tools of statistical analysis of web servers such as Access Watch, Analog, Gwstat, WebStat, etc [5]. Also, the Synergo / ColAT can be mentioned as a specific tool for statistical analysis of students' information in a designed learning management system [6]. In the article [7], a simple analysis of the number of students referrals and number of page visits have been presented. In [8], the time distribution of the students' attending in virtual environment and the courses visited more have been considered. In [9], the students' behaviors within a semester have been reviewed and the key words in searching them have been analyzed. Also in [10], weekly and monthly reports of changes in student behavior patterns have been presented.

In [11], the difficulty of presented problems and exercises in the virtual education system has been analyzed using the statistical analyses. The results of these analyses have been also used to improve the e-learning environment. The correlation analysis has been used to study the students' trends in learning [12] and predicting the final scores [13]. Also in [14], the regression analysis has been used for the modeling of student knowledge according to the final scores and review of criteria that cause disruption in this modeling. In [15], using regression it is predicted whether a student chooses the right answer to a question. Also in [16], using the student response to questions during the semester, his/ her final scores is predicted.

B. *Visualization*

Visualization is considered as a field of computer graphics and user interface design knowledge that deals with the representation of information as digital interactive or dynamic images [17].

The main application of these methods is to facilitate the analysis of huge volume of data by presenting them in a graphical format. Typically, the raw data is displayed in the form of tables, charts and graphs, which are not so useful while dealing with high volumes of information. In contrast, using visualization, complex and multi-dimensional information that has been produced from tracking of students in virtual learning environments can be well studied. In [18], visualization methods have been used for analyzing the accountability status of students to exercises. Some researchers in this field have studied the development of learning management system and adding visualization modules and their on-line surveying. For example in [19], the CourseVis tool has been developed for analyzing the students' interactions with existing courses in the WebCT learning management system. In [20], the GISMO tool uses the information related to Moodle open source learning management system as the main source and provides the graphic charts about different subjects and how to review their educational content by the students. Also, the TADA-Ed tool [21] integrates different methods of visualization to facilitate the monitoring process by the professor. In [22], a tool for studying high volume of interactions between professors and students in the LISTEN project has been provided, which is a reading tutorial program. Using these tools, courses professors and educational administrators can achieve good knowledge about the students and about how they learn in virtual learning environment. As a step forward in [23], a business intelligence tool by the name MATEP has been provided for observing and analysis how the students interact on line. In recent years, using data warehouse and business intelligence tools in e-learning systems have received a wide welcome among who are involved in this area.

C. Clustering

Clustering methods deal with data grouping according to similarities among them. In e-learning, clustering methods are used for grouping students according to various features. In [24], the students have been grouped based on the characteristics related to learning style and have been used in order to promote interactions level among similar groups and prevent the students problems. In [25], behavioral patterns of students in an interactive learning environment have been extracted using clustering methods. Also in [26], using clustering of the students, how to provide educational content and their learning path have been personalized. In [27], tests and questions have been grouped based on how the students answer to them. Also in [28], the users have been classified in terms of attending in the virtual learning environment and how to interact with the system. The main advantage of clustering methods includes revealing the latent patterns in the data, which can be extracted using analysis and statistical methods that often leads to valuable results to promote the quality of learning and education.

D. Classification

Classification methods deal with data modeling using the set of learning and predicting the class labels of test data sets based on their features. In electronic learning, these methods are usually used to predict a specific feature in terms of previous models. In [29], classification methods have been used to study the responds of different groups of students to different instructional strategies. In [30], the performance of students and their final scores have been anticipated using a combination of classification methods. In [31], the students' misuse of learning environment and their irresponsible behaviors have been detected using classification methods. Also in [32], the students' performance in a learning environment along with impressive features to improve such a performance has been studied. In [33], the students have been categorized to two groups of prone to error and prone to correctness based on the use pattern, and using this classification results, the common causes of students' mistakes have been studied. In [34], the classification methods have been used to identify students with low motivation and finding some solutions to treat and prevent their dropping out of courses. Also in [35], these methods have been used to predict the success rate of the subjects. One of the common methods of classification is to use the decision tree generating methods. The decision tree due to the simplicity and high interpretation capability is very useful in the field of education and learning. The generated decision tree can be also used to extract rules as IF THEN ELSE, which may lead to the extraction of interesting information about the way of students use and their results.

E. Extracting the association rules

The association rules are applied to extract the relationship among different features in the database. The results of these algorithms are usually provided as a set of $X \rightarrow Y$ rules, where X and Y are sets of attributes. Different studies have worked to extract the association rules in the databases of web-based training systems that we will review them in continue. In [36], the association rules have been applied for making the recommending agents to the students. These agents recommend various educational activities suitable to the students' use patterns and suggest them some shortcuts for elimination of unnecessary educational resources. In [37], the association rules have been used for resolving the students' problems in the learning environment and providing consultation for them. In [38], these rules have also been used to guide the activities of students and suggesting them educational content. In [39], the most appropriate educational content is determined specifically to provide to the students, and in [40], the extracted rules have been used to optimize the virtual learning environment based on items that are interesting from the perspective of students. In [41], a method for determining the main distinguishing features of the students has been presented from the perspective of efficiency. In [42], interesting rules according to the pattern of observing educational content by students has been extracted to provide feedback to the content producers. In [43], the existing communications among the behavioral patterns of students in the learning environment has been studied, and in [44], the errors of students that usually happen together have been extracted.

F. Extracting sequential patterns

Sequential patterns are considered as a particular form of association rules in which the time and the arrangement of objects in the database are considered as a parameter. These methods have been widely used in e-learning and specifically to study the behavioral patterns of students in the learning environment. In [45], extraction of sequential patterns has been used for assessment of students' activities and personalizing the training content. In [46], these methods have been also used for the production of educational activities suitable to the learning pattern of different groups of students. In [47], these methods have been applied to extract the learning behavior of the students and its comparison with the designed learning path (ideal form). In [48], using the sequential patterns extraction, the learning path of successful students and sequences of the indicators for doing educational activities have been determined. Some researchers have evaluated the virtual learning environment and its improvement by extracting the learning path of the students. In [49], these methods have been used for evaluating and improving the design of educational websites. Also in [50], these methods have been used to create a standard for designing the best structure for online learning environment.

G. Outlier analysis

The outlier analysis methods search for records in the database, which are largely different to the expected values. These methods are usually used for data purification and removing noise and inaccurate information and data, but sometimes they can be used to identify growing trends and find rare cases among the data sets. For example, these methods can be used to find unusual students in e-learning. In [51] and [52], these methods have been used to detect abnormal behavior of students in the learning environment. Also in [53], in addition to recognize the behavior deviation of students from the normal route, abnormal behaviors of professors and teachers have been also analyzed.

H. Social network analysis

Social network analysis, which is considered as one of the new areas in data mining, deals with the extraction of patterns among existing units in a social environment. In fact, these methods consider the relation between these two units as a dependency and try to extract communicating patterns among different units and using them to discover knowledge. E-learning is also a context for interaction between students, professors and educational content. In fact most of the new educational contexts have developed based on social constructivism theory, in which the student's learning process occurs through a series of his/her interactions with educational content, learning activities, course instructor and other students. In [54], these methods have been used to extract the group activities of students and analysis of sociogram and the correlation of its components. In [55], these methods have been used for analysis and interpretation of educational content structure in online learning communities.

I. Tool development

Some researchers have designed specific tools to study the way of users' use of learning environment in addition to providing algorithms. Among these tools, some can be mentioned as following : the MultiStar [56] for the extraction of association and classification rules, the Tool [57] to study and quantitative analysis of students' performance, the EPRules [58] to extract association rules, the KAON [59] for clustering and source searching, the O3R [60] for discovering sequential patterns, the MINEL [61] to determine a framework for analyzing virtual learning systems logs and finding learning path of students, the CIECoF [62] to extract association rules and the Simulog [63] to extract the unexpected behavior of the students.

III. STUDY THE STATUS OF RESEARCH ON EDUCATIONAL DATA MINING AREA

After review of research on the area of web searching for information about the students ' use of learning environments and to learn more about other areas of educational data mining, we will review the main topics in the field in early years and modern trends in this area according to the presented papers in the first and second conferences and two training in continue.

A. The main topics in the early years

In the past, the communication extraction and analysis methods have allocated the major part of related studies to educational data training. In [1], 60 articles in this area have been reviewed from 1995 to 2005 that among them 26 articles (43 percent) are devoted to the extraction of communication. Also, 17 articles (28 percent) are devoted to various types of forecasting methods and other items in this category have devoted a small ratio to themselves. Fig. 2 shows total distribution of papers provided in each of the areas. It should be noted that the articles used several methods have been counted in all relevant areas.

B. The main topics in recent years

As mentioned, studies in this area in the early years were about communications extracting and prediction, but the pattern trend changed in the first three years of holding data mining conference. Fig.2 shows total distribution of papers provided in each of the areas.

As Fig.3 shows, the rank of communication extraction in presented papers in these conferences has plummeted to fifth rank and only nine percent of the articles have devoted to the topic. In contrast, the prediction area using learning data mining that was in the second rank, has achieved the first rank by devoting 42 percent of the papers to itself. Other items have changed a little than before, but the new method with public interest is exploration of knowledge by using modeling, so that allocating 19 percent of articles has placed in second rank.

It is true that all articles in the field of educational data mining are not limited to the articles presented in the first and second conferences, but it should be noted that the number of presented articles in these two conferences are approximately equal to the number of published papers in the first decade of this area. This shows the rapid growth of research in this area and the impossibility of reviewing all the papers.

IV. CASE STUDY

The application of data mining in e-learning systems is an iterative cycle [1]. The mined knowledge should enter the loop of the system and guide, facilitate and enhance learning as a whole, not only turning data into knowledge, but also filtering mined knowledge for decision making. The e-learning data mining process consists of the same four steps in the general data mining process. Fig.4 shows them.

- Collect data: The learning management system is used by students and the usage and interaction information is stored in the database. In this paper we have used the students' usage data in the Imam Khomeini virtual university system.
- Preprocess the data: The data is cleaned and transformed into an appropriate format to be mined. In order to preprocess the data, we can use a database administrator tool or some specific preprocessing tool.
- Apply data mining: The data mining algorithms are applied to build and execute the model that discovers and summarizes the knowledge of interest to the user (instructor, student and administrator). To do so, either a general or a specific data mining tool, or a commercial or free data mining tool can be used.
- Interpret, evaluate and deploy the results: The results or model obtained are interpreted and used by the instructor for further actions. The instructor can use the information discovered to make decisions about the students' and university course activities to improve the students' learning.

A. *Preprocessing of data*

Actual data of the virtual learning system of the Imam Khomeini educational and research institute have been used in implementation of the article.

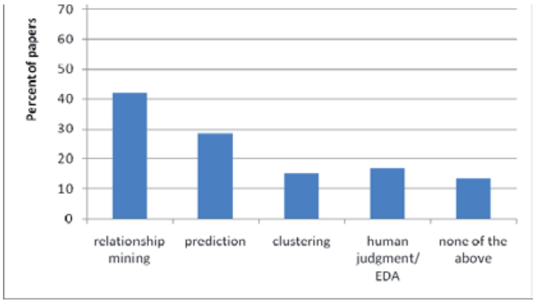

Fig. 2 Distribution of articles in the field of educational data mining according to [1]

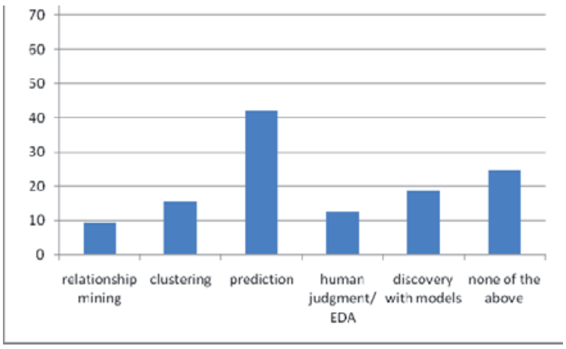

Fig.3 Distribution of articles in the educational data mining conferences

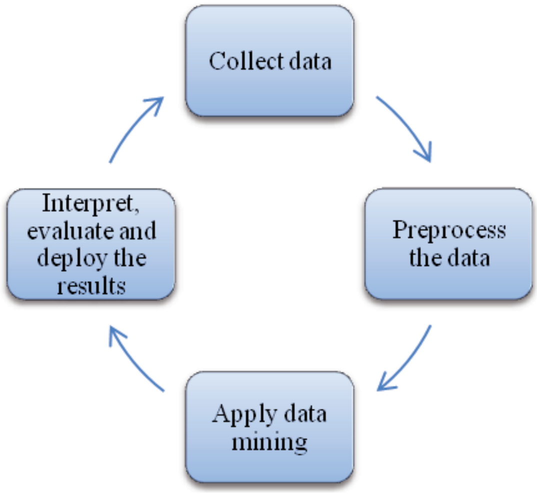

Fig.4 process of data mining

This system keeps detailed logs of all activities that students perform. It logs every click that students make for navigational purpose. It stores the logs in some MySQL relational database (a database with 62 tables for each course and 3 other database for logs , user's information and etc with totally 88 tables). But we do not need all this information and it is also necessary to convert it into the required format used by data mining algorithms. Therefore a number of general data preprocessing tasks like data cleaning, data transformation and enrichment, data integration and data reduction applied in to data. Although the amount of work required in data preparation is less, the following tasks also need to be done:

- Selecting data:

The project was performed on the data related to the first semester of 2010. The data includes 1076 students and 35 courses titles that totally contain 3674 students - course.

- Preparation of summary table (summarizing information of a course):

Table. 2 shows characteristics have been extracted and used from various databases.

- Data discretization and normalization:

Performing a discretization of numerical values may be necessary to increase interpretation and comprehensibility. Discretization divides the numerical data into categorical classes that are easier to understand for the instructor (categorical values are more user-friendly for the instructor than precise magnitudes and ranges). Some numerical values of the summarization table have been discretized (like for mark attribute: pass if value is>12 and fail if value is<12).

- Transferring prepared data to the data mining software:

The data must be transformed to the required format of the data mining algorithm or framework. In our case, the summary table have been exported to SQL Server database.

B. Applying data mining techniques to data

In the following, three techniques introduced in the previous section, for example, have been applied to the data of Imam Khomeini virtual University. This has been done through SQL SERVER software, these three cases are: Classification and association rules and clustering that each of them will be described in the following:

- Classification:

In this section by the help of decision tree algorithm, a tree was developed in order to predict the student status to pass or not to pass the term.

The instructor can use the knowledge discovered by these tree for making decisions about course activities and for classifying new students. For example, it is very logical that the number of quizzes passed was the main discriminator of the final marks. But there are some others that can help the instructor to decide to promote the use of some types of activities to obtain higher marks. Or on the contrary, to decide to eliminate some activities related to low marks. The instructor can also detect new students with learning problems in time (students classified as FAIL). The instructor can use the decision tree model in order to classify new students and detect in time if they will have learning problems (students classified as FAIL) or not (students classified as PASS).

Fig. 5 shows the decision tree constructed from the data. In the figure, three levels of the created tree are visible, as indicated in the figure, percentage of correct responses to quizzes, the percentage of exercises and amount of educational content are the most important factors of students ranking. The next levels are parameters such as participation in chats and forums and the number of delivered messages in each of them.

Table. 2 Fields used in data mining

Fields
Course
The number of studied content sessions
The number of answered tests
The number of correct answers on tests
The number of wrong answers on tests
The number of posts in chat
The number of posts on forum
The time spent for the study the content
The number of performed tasks
The number of visiting the chat
The number of visiting the educational content
The number of visiting the forum
The final score

To measure the accuracy of the constructed model, the classification matrix and k-fold cross validation methods have been used that the results are shown on Table.3 and Table.4.

- Association rules

Table.5 shows some part of the extracted association rules. Apriori algorithm with min support of 4 has been used for extracting those rules. In the rules shown in the table, the amount of forum visit, for example is right hand side rules. By applying the algorithm on the data, a high volume of rules are extracted that only part of it has been shown in table 5. Some of the extracted rules are obvious such as rule 1; others are part of other rules for example rule 10 is part of rule 26; but generally, these rules help the instructor to understand strengths and weaknesses of students and efficiency and impact of each of instruments on learning rate of students. For example, rule 29 is one of these rules.

- Clustering

K-means algorithm was applied on the data with different number of clusters for data clustering. The purpose of implementing this algorithm is to classify students, such that according to the amount of their activity, they can be attributed to categories of strong, moderate and weak students. The best algorithm results were obtained by selecting three as the number of clusters. Fig.6 shows the obtained clusters.

Table.6 shows characteristics of the three clusters; as indicated in the table, students have been included in three clusters, and the amount of their activity in the virtual learning environment determines their membership in the relevant cluster.

The instructor can use this information in order to group students into three types of students: very active students , active students and non-active students . Starting from this information, for example, the instructor can group students for working together in

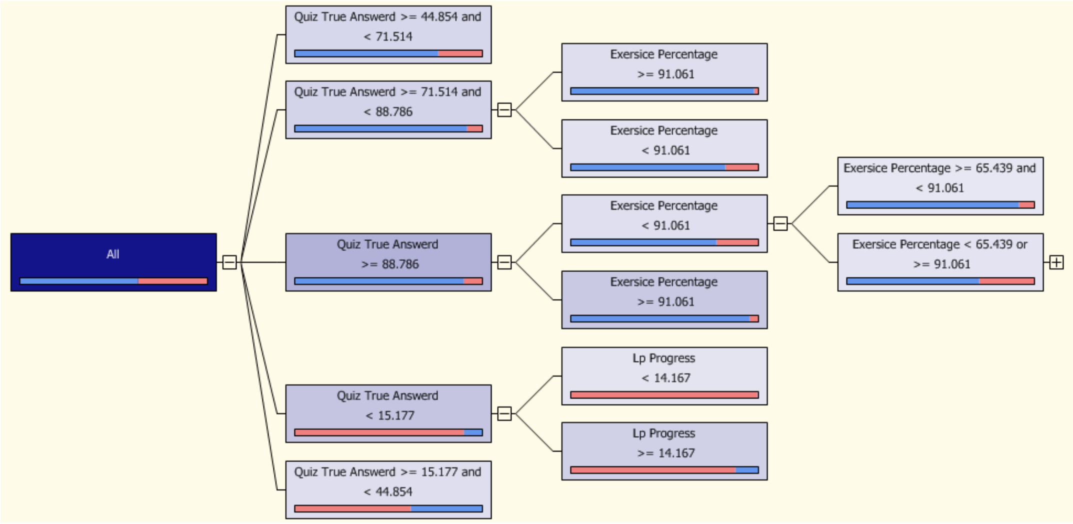

Fig. 5 The decision tree constructed from the data

Table. 3 Classification matrix

Counts for Dtree on Grade Pass Fail		
Predicted	0 (Actual)	1 (Actual)
0	56	17
1	14	134

Table. 4 10-fold cross validation

Partition Index	Partition Size	Test	Measure	Value
1	199	Classification	Pass	168
2	200	Classification	Pass	175
3	200	Classification	Pass	174
4	200	Classification	Pass	161
5	200	Classification	Pass	171
6	200	Classification	Pass	168
7	200	Classification	Pass	164
8	199	Classification	Pass	175
9	199	Classification	Pass	172
10	199	Classification	Pass	169

Table. 5 The extracted association rules

	Probability	Rule
1.	1	Chat Access Count = Normal, Lp Progress = Normal - > Forum Accsess Count = Normal
2.	1	Quiz Access Count = High, Numberofm On Chat = Low -> Forum Accsess Count = Normal
3.	1	Quiz Access Count = High, Chat Access Count =Normal -> Forum Accsess Count = Normal
4.	1	Chat Access Count = High, Numberofm On Chat =Low -> Forum Accsess Count = Normal
5.	1	Chat Access Count = High, Categorical Grade = Pass -> Forum Accsess Count = Normal
6.	0.8	Chat Access Count = High, Quiz Access Count = Normal -> Forum Accsess Count = Normal
7.	0.778	Learnpath Access Count =Very High, Chat Access Count = Normal -> Forum Accsess Count = Normal
8.	0.7	Chat Access Count = Normal, Quiz Access Count = High -> Forum Accsess Count = Normal
9.	0.667	Chat Access Count = High, Lp Progress = Very High -> Forum Accsess Count = Normal
10	0.667	Chat Access Count = High -> Forum Accsess Count = Normal
11	0.667	Chat Access Count = High, Quiz Total Answerd =Very High -> Forum Accsess Count = Normal
12	0.667	Chat Access Count = High, Learnpath Access Count = High -> Forum Accsess Count = Normal
13	0.615	Chat Access Count = Normal, Categorical Grade = Pass -> Forum Accsess Count = Normal
14	0.6	Chat Access Count =Very High, Quiz Total Answerd =Very High -> Forum Accsess Count = High
15	0.6	Numberofm On Chat = Normal, Chat Access Count = Normal -> Forum Accsess Count = Normal
16	0.6	Chat Access Count = Normal, Quiz Access Count =Low -> Forum Accsess Count = Normal

17	0.6	Chat Access Count =Very High, Quiz True Answerd =Very High -> Forum Accsess Count = High
18	0.571	Learnpath Access Count = Very High, Numberofm On Chat = Normal-> Forum Accsess Count = Normal
19	0.524	Learnpath Access Count = Very High, Quiz True Answerd =Very High -> Forum Accsess Count = Normal
20	0.5	Chat Access Count =Very High, Lp Progress = Very High -> Forum Accsess Count = High
21	0.5	Learnpath Access Count = Very High, Quiz Access Count =Normal -> Forum Accsess Count = Normal
22	0.5	Quiz Access Count = High, Learnpath Access Count = Normal -> Forum Accsess Count = Normal
23	0.5	Chat Access Count =Very High -> Forum Accsess Count = High
24	0.481	Chat Access Count = Normal, Quiz True Answerd =Very High -> Forum Accsess Count = Normal
25	0.469	Learnpath Access Count= Very High, Quiz Total Answerd =Very High -> Forum Accsess Count = Normal
26	0.462	Learnpath Access Count = Very High, Categorical Grade = Pass -> Forum Accsess Count = Normal
27	0.462	Quiz Access Count = Very High, Learnpath Access Count = Very High -> Forum Accsess Count = Normal
28	0.462	Learnpath Access Count = Very High, Exersice Percentage =Very High -> Forum Accsess Count = Normal
29	0.462	Learnpath Access Count = Low, Categorical Grade =Fail -> Forum Accsess Count = Low
30	0.455	Learnpath Access Count = Very High, Lp Progress = Very High -> Forum Accsess Count = Normal
31	0.441	Learnpath Access Count = Very High -> Forum Accsess Count = Normal
32	0.439	Chat Access Count = Normal, Quiz Total Answerd =Very High -> Forum Accsess Count = Normal

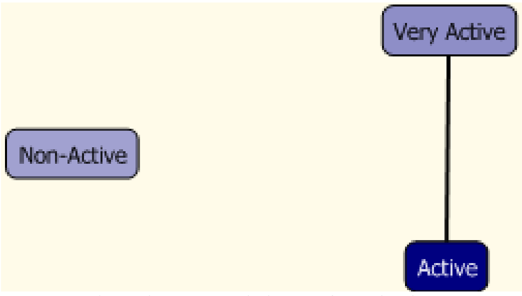

Fig. 6 the extracted clusters from the data

Table. 6 The Clusters' characteristics

Cluster name	Cluster centroids
Very active	Lp Progress>=87.5 , Numberofm On Chat >= 22 , Chat Access Count >= 13 , Forum Accsess Count >= 44 Learnpath Access Count >= 85 , Quiz Access Count >= 40, Exersice Percentage >= 91 , Quiz True Answerd >= 71 , Quiz Total Answerd >= 91, Exersice Percentage >= 91.
Active	Lp Progress=55 - 62 , Numberofm On Chat < 2 - 22, Quiz Access Count=15 - 40, Chat Access Count < 4 , Quiz Total Answerd=65 - 91 , Quiz True Answerd=44- 71 , Learnpath Access Count=20 - 48, Exersice Percentage=35 – 71, Forum Accsess Count < 4 , Forum Accsess Count=4- 18
Non-active	Lp Progress<55 Quiz Total Answerd < 8, Quiz True Answerd < 15 , Exersice Percentage < 9 Quiz Total Answerd=8 - 33.4128176704 , Quiz Access Count < 15 , Learnpath Access Count<20 ,Forum Accsess Count < 4 , Chat Access Count < 3, Numberofm On Chat < 2

collaborative activities (each group with only students of the same cluster or each group with a similar number of students of each cluster). The instructor can also group new students into these clusters depending on their characteristics.

CONCLUSIONS

In this paper, the useful applications of data mining in the learning management system were introduced. However, each of them was discussed individually. To learn more attractive information, a combination of introduced applications can be used. For example, if there was a special case in the graphs, further details can be obtained by studying calculated statistical values. Or if we find some similar groups of students in the graph, using the clustering techniques, the groups can be separated from one another, and so on.

REFERENCES

[1] C. Romero and S. Ventura, "Educational Data Mining: A Survey from 1995 to 2005",*Expert Systems with Applications*, vol. 33, pp. 125-146, 2007.

[2] C. Romero, S. Ventura, and E. Garcia, "Data mining in course management systems:Moodle case study and tutorial", *Computers &Education* , vol. 51, no. 1, 2008, pp.368–384.

[3] R. S. J. B. and K. Yacef, "The State of Educational Data Mining in 2009: A Reviewand Future Visions", *Journal of Educational Data Mining*, vol. 1, no. 1, 2009, pp. 1-15.

[4] F. Castro, A. Vellido, A. Nebot, and F. Mugica, "Applying Data Mining Techniques toe-Learning Problems", in *Studies in Computational Intelligence*. Berlin, Heidelberg:Springer-Verlog, 2007, pp. 183-221.

[5] O. Ziane, M. Xin, and J. Han, "Discovering web access patterns and trends by applyingOLAP and data mining technology on web logs",*Advances in Digital Libraries*, 1998,pp. 19-29.

[6] N. Avouris, V. Komis, G. Fiotakis, M. Margaritis, and E. Voyiatzaki, "Why logging offingertip actions is not enough for analysis of learning activities.", in *Workshop onUsage analysis in learning systems at the 12th International Conference on ArtificialIntelligence in Education*, Amsterdam, Netherland, 2005, pp. 1-8.

[7] C. Pahl and C. Donnelan, "Data mining technology for the evaluation of web-basedteaching and learning systems",in*Proceedings of the Congress E-learning*, Montreal,Canada, 2003, pp. 1–7.

[8] K. Nilakant and A. Mitrovic, "Application of data mining in constraintbasedintelligenttutoring systems", in *Proceedings of artificial intelligence in education*, Amsterdam,Netherlands, 2005, pp. 896–898.

[9] H. L. Grob, F. Bensberg, and F. Kaderali, "Controlling open source intermediaries – aweb log mining approach", in *Proceedings of the international conference oninformation technology interfaces*, Zagreb, Croatia, 2004, pp. 233–242.

[10] D. Monk, "Using data mining for e-learning decision making",*Electronic Journal of e-Learning*, vol. 3, no. 1, 2005, pp. 41–54.

[11] M. E. Zorilla, E. Menasalvas, D. Marin, E. Mora, and J. Segovia, "Web usage miningproject for improving web-based learning sites", in *Proceedings of Web miningworkshop*, Cataluna, Spain, 2005, pp. 1–22.

[12] I. Arroyo, T. Murray, B. Woolf, and C. Beal, "Inferring unobservable learning ariablesfrom students' help seeking behavior", in *Proceedings of Intelligent tutoring systems*, Alagoas, Brazil, 2004, pp. 782–784.

[13] D. E. Pritchard and R. Warnakulasooriya, "Data from a web-based homework tutor can predict student's final exam score", in *Proceedings of World conference on educationalmultimedia, hypermedia and telecommunications*, Canada, 2005, pp. 2523–2529.

[14] M. Feng, N. Heffernan, and K. Koedinger, "Looking for sources of error inpredictingstudent's knowledge", in *Proceedings of AAAI workshop on educational data mining*, California, USA, 2005, pp. 1–8.

[15] J. E. Beck and B. Woolf, "High-level student modeling with machine learning", in *Proceedings of the International Conference on Intelligent tutoring systems*, 2000, pp. 584-593.

[16] N. Anozie and B. W. Junker, "Predicting end-of-year accountability assessment scores from monthly student records in an online tutoring system", in *Proceedings of the AAAIworkshop on educational data mining*, California, USA, 2006, pp. 1–6.

[17] R. Spence, "Information Visualization",*Addison-Wesley*, 2001.

[18] R. Shen, F. Yang, and P. Han, "Data analysis center based on e-learning platform", in *Proceedings of the Workshop on The Internet Challenge: Technology and Applications*, Berlin, Germany, 2002, pp. 19–28.

[19] R. Mazza and V. Dimitrova, "CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses",*International Journal of Human-Computer Studies*, vol. 65, no. 2, 2007, pp. 125–139.

[20] R. Mazza and C. Milani, "Exploring usage analysis in learning systems: Gaining insights from visualisations", in Proceedings of Workshop on Usage analysis inlearning systems at 12th International Conference on Artificial Intelligence inEducation, New York, USA, 2005, pp. 1-6.

[21] A. Merceron and K. Yacef, "TADA-Ed for educational data mining",*InteractiveMultimedia Electronic Journal of Computer-enhanced Learning*, vol. 7, no. 1, 2005,pp. 267-287.

[22] J. Mostow et al., "An educational data mining tool to browse tutor-student interactions: Time will tell!", in *Proceedings of the Workshop on Educational Data Mining*, Pittsburgh, USA, 2005, pp. 15–22.

[23] M. E. Zorrilla and E. Alvarez, "MATEP: Monitoring and Analysis Tool for E-learning Platforms", in *Proceedings of the 8th IEEE International Conference on AdvancedLearning Technologies*, Santander, Spain, 2008, pp. 611-613.

[24] T. Tang and G. McCalla, "Utilizing Artificial Learners to Help Overcome the Cold-Start Problem in a Pedagogically-Oriented Paper Recommendation System", in *Proceedings of the International Conference on Adaptive Hypermedia*, 2004, pp. 245-254.

[25] L. Talavera and E. Gaudioso, "Mining Student Data to Characterize Similar Behavior Groups in Unstructured Collaboration Spaces", in *Proceedings of the Workshop inArtificial Intelligence in Computer Supported Collaborative Learning in conjunctionwith 16th European Conference on Artificial Intelligence, ECAI'2003.*, Valencia, Spain, 2004, pp. 17–22.

[26] E. Mor and J. Minguillon, "E-learning personalization based on itineraries and longtermnavigational behavior", in *Proceedings of the 13th international world widewebconference*, New York, USA, 2004, pp. 264–265.

[27] J. Spacco, T. Winters, and T. Payne, "Inferring use cases from unit testing", in *Proceedings of the AAAI Workshop on Educational Data Mining*, New York, USA, 2006, pp. 1-7.

[28] F. H. Wang and H. M. Shao, "Effective personalized recommendation based on timeframednavigation clustering and association",*Expert Systems with Applications*, vol. 27, no. 3, 2004, pp. 365–377.

[29] G. Chen, C. Liu, K. Ou, and B. Liu, "Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology",*Journal of Educational Computing Research*, vol. 23, no. 3, 2000, pp. 305–332.

[30] BehrouzMinaei-Bidgoli and Bill Punch, "Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System",*Genetic and EvolutionaryComputation*, vol. 2, 2003, pp. 2252–2263.

[31] R. S. Baker, A. T. Corbet, and K. Koedinger, "Detecting Student Misuse of Intelligent Tutoring Systems", in *Proceedings of the 7th International Conference on IntelligentTutoring Systems*, Maceio, Brazil, 2004, pp. 531-540.

[32] S. B. Kotsiantis, C. J. Pierrakeas, and P. E. Pintelas, "Predicting Students' Performance in Distance Learning Using Machine Learning Techniques",*AppliedArtificialIntelligence*, vol. 18, no. 5, 2004, pp. 411–426.

[33] M. V. Yudelson et al., "Mining Student Learning Data to Develop High Level Pedagogic Strategy in a Medical ITS", in *Proceedings of the AAAI Workshop onEducational Data Mining*, Boston, MA, USA, 2006, pp. 1-8.

[34] M. Cocea and S. Weibelzahl, "Can Log Files Analysis Estimate Learners' Level of Motivation?", in *Proceedings of the 14th Workshop on Adaptivity and User Modelingin Interactive Systems*, Hildesheim, Germany, 2006, pp. 32-35.

[35] W. Hamalainen and M. Vinni, "Comparison of machine learning methods for intelligent tutoring systems", in *Proceedings of the 8th international*

conference inintelligent tutoring systems, Taiwan, 2006, pp. 525–534.

[36] O. Zaïane, "Building a recommender agent for e-learning systems", in *Proceedings of the International Conference on Computers in Education*, 2002, pp. 55–59.

[37] G. J. Hwang, C. L. Hsiao, and C. R. Tseng, "A computer-assisted approach to diagnosing student learning problems in science",*Journal of Information Science andEngineering*, vol. 19, 2003, pp. 229–248.

[38] J. Lu, "Personalized e-learning material recommender system", in *Proceedings of the International conference on information technology for application*, Utah, USA, 2004,pp. 374–379.

[39] P. Markellou, I. Mousourouli, S. Spiros, and A. Tsakalidis, "Using semantic web mining technologies for personalized e-learning experiences", in *Proceedings of theweb-based education conference*, Grindelwald, Switzerland, 2005, pp. 461–826.

[40] A. A. Ramli, "Web usage mining using apriori algorithm: UUM learning care portal case.", in *Proceedings of the International Conference on Knowledge Management*, Malaysia, 2005, pp. 1-19.

[41] BehrouzMinaei-Bidgoli, P. Tan, and William Punch, "Mining interesting contrast rules for a web-based educational system", in *Proceedings of the international conference onmachine learning applications*, Louisville, KY, 2004, pp. 1–8.

[42] C. Romero, S. Ventura, and P. D. Bra, "Knowledge discovery with genetic programming for providing feedback to courseware author", *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, vol. 14, no. 5, 2004, pp. 425–464.

[43] P. Yu, C. Own, and L. Lin, "On learning behavior analysis of web based interactive environment", in *Proceedings of the Workshop on Implementing Curricular Change inEngineering Education*, Oslo, Norway, 2001, pp. 1-10.

[44] A. Merceron and KalinaYacef, "Mining student data captured from a web-based tutoring tool: Initial exploration and results",*Journal of Interactive Learning Research*, vol. 15, no. 4, 2004, pp. 319–346.

[45] O. Zaïane and J. Luo, "Web usage mining for a better web-based learning environment", in *Proceedings of conference on advanced technology for education*, Banff, Alberta, 2001, pp. 60-64.

[46] W. Wang, J. Weng, J. Su, and S. Tseng, "Learning portfolio analysis and mining in scorm compliant environment," in *Proceedings of the ASEE/IEEE Frontiers inEducation Conference*, Savannah, USA, 2004, pp. 17–24.

[47] C. Pahl and C. Donnellan, "Data mining technology for the evaluation of web-based teaching and learning systems", in *Proceedings of the Congress E-learning*, Montreal, Canada, 2003, pp. 1-7.

[48] J. Kay, N. Maisonneuve, K. Yacef, and O. R. Zaiane, "Mining patterns of events in students' teamwork data", in *Proceedings of the educational data mining workshop*, Taiwan, 2006, pp. 1–8.

[49] L. Machado and K. Becker, "Distance education: A web usage mining case study for the evaluation of learning sites", in *Proceedings of the International conference onadvanced learning technologies*, Athens, Greece, 2003, pp. 360–361.

[50] S.Ha, S.Bae, and S. Park, "Web mining for distance education", in *Proceedings of the IEEE international conference on management of innovation and technology*,Singapore, 2000, pp. 715–719.

[51] M. Ueno, "Online outlier detection system for learning time data in e-learning and its evaluation", in *Proceedings of the International conference on computers andadvanced technology in education*, Beijing, China, 2004, pp. 248–253.

[52] F. Castro, A. Vellido, A. Nebot, and J. Minguillon, "Detecting a typical student behaviour on an e-learning system", in *Proceedings of the SimposioNacionaldeTecnologı´as de la Informacin y lasComunicaciones en la Educacion*, Granada, Spain, 2005, pp. 153–160.

[53] M. Muehlenbrock, "Automatic action analysis in an interactive learning environment", in Proceedings of the workshop on usage analysis in learning systems at the 12thinternational conference on artificial intelligence in education, Amsterdam, Netherlands, 2005, pp. 73–80.

[54] P. Reyes and P. Tchounikine, "Mining learning groups' activities in forum-type tools",in*Proceedings of the 2005 conference on computer support for collaborative learning*, Taiwan, 2005, pp. 509–513.

[55] R. Rallo, M. Gisbert, and J. Salinas, "Using data mining and social networks to analyze the structure and content of educative online communities", in *Proceedings of theInternational conference on multimedia and ICTs in education*, Caceres, Spain, 2005, pp. 1–10.

[56] D. Silva and M. Viera, "Using data warehouse and data mining resources for ongoingassessment in distance learning", in *Proceedings of the IEEE International Conferenceon Advanced Learning Technologies*, Kazan, Russia, 2002, pp. 40–45.

[57] F. C. I. Chang, L. P. Hung, and T. K. Shih, "A new courseware for quantitativemeasurement of distance learning courses",*Journal of Information Science andEngineering*, vol. 19, 2003, pp. 989–1014.

[58] C. Romero, S. Ventura, and P. D. Bra, "Knowledge discovery with geneticprogramming for providing feedback to courseware author",*User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, vol. 14, no. 5, 2004, pp.425–464.

[59] J. Tane, C. Schmitz, and G. Stumme, "Semantic resource management for the web: Anelearning application", in *Proceedings of the World Wide Web Conference*, New Yorke,USA, 2004, pp. 1–10.

[60] K. Becker, M. Vanzin, and D. A. Ruiz, "Ontology-based filtering mechanisms for webusage patterns retrieval", in *Proceedings of the International conference on ecommerceand web technologies*, Mu¨ nchen, Germany, 2005, pp. 267–277.

[61] A. Bellaachia, E. Vommina, and B. Berrada, "Minel: A framework for mining elearninglogs", in *Proceedings of the fifth IASTED international conference on Webbasededucation*, Mexico, 2006, pp. 259–263.

[62] E. Garcia, C. Romero, S. Ventura, and C. Castro, "Using rules discovery for thecontinuous improvement of e-learning courses", in *Proceedings of the Internationalconference intelligent data engineering and automated learning*, Burgos, Spain, 2006,pp. 887–895.

[63] J. Bravo and A. Ortigosa, "Validating the evaluation of adaptive systems by user profilesimulation", in *Proceedings of the workshop on user-centred design and evaluation ofadaptive systems*, Dublin, Ireland, 2006, pp. 52–56.

Behrouz Minaei obtained his Ph.D. from Michigan State University, Michigan, USA in Computer Science and Engineering Department. He is an assistant professor in Computer Engineering Department of Iran University of Science & Technology, Tehran, Iran. He is also leading at a Text Mining research gr-oup in NOOR Co. developing large corpora for Farsi and Arabic languages.

Seyed Hassan Hani Tabaei obtained his M.Sc. degree in Computer Engineering from Sharif University of Technology, Tehran, Iran. He is the head of Computer Engineering Department of University of Qom. He also worked as a IT manager and IT consultant for organizations such as municipality of Qom and Isfahan. His specialties are decision support systems (DSS), e-payment, e-readiness and e-government.

Vahid Ghanbari received his B.Sc. degree with (first class) Honors in computer science from University of Qom, Iran, in 2009 and currently he is a M.Sc. student in the Department of Information Technology at University of Qom, Iran. He is also a lecturer at Taali University of Information Technology, Qom, Iran. His specialties are e-learning, data mining in e-learning systems and temporal pattern mining.