E-Learning Using Data Mining

Shimaa Abd Elkader Abd Elaal
E-learning using data mining

Shimaa Abd Elkader Abd Elaal

Abstract

Educational Data Mining (EDM) is the process of converting raw data from Educational systems to useful information that can be used by educational software Developers, students, teachers, parents, and other educational researchers.

Currently there is an increasing interest in data mining and educational systems, making educational data mining as a new growing research community. This work provides the application of data mining to traditional educational systems, particular web-based courses, well-known learning content management systems, and adaptive and intelligent web-based educational systems. Each of these systems has different data source and objectives for knowledge discovering. After preprocessing the available data in each case, data mining techniques can be applied: statistics and visualization; clustering, classification and outlier detection; association rule mining and pattern mining; and text mining. The success of the plentiful work needs much more specialized work in order for educational data mining to become a mature area.

Keywords: Data mining and E-learning, Educational Data Mining, Data mining functionalities for e-learning domain, Process of Data Mining in E-Learning
**1 Introduction**

Sun et al (2007) defines e-learning as “a web-based system that makes information or knowledge available to users or learners and disregards time restrictions or geographic proximity”.[10]

**E-learning** (also referred to as web-based education and e-teaching), a new context for education where large amounts of information describing the continuum of the teaching-learning interactions are endlessly generated and ubiquitously available. This could be seen as a blessing: plenty of information readily available just a click away. But it could equally be seen as an exponentially growing nightmare, in which unstructured information chokes the educational system without providing any articulate knowledge to its actors, Data Mining was born to tackle problems like this. As a field of research, it is almost contemporary to e-learning. It is, though, rather difficult to define. Not because of its intrinsic complexity, but because it has most of its roots in the ever-shifting world of business. At its most detailed, it can be understood not just as a collection of data analysis methods, but as a data analysis process that encompasses anything from data understanding, pre-processing and modeling to process evaluation and Implementation. It is nevertheless usual to pay preferential attention to the Data Mining methods themselves. These commonly bridge the fields of traditional statistics, pattern recognition and machine learning to provide analytical solutions to problems in areas as diverse as biomedicine, engineering, and business, to name just a few. An aspect that perhaps makes Data Mining unique is that it pays special attention to the compatibility of the modeling techniques with new Information Technologies (IT) and database technologies, usually
focusing on large, heterogeneous and complex databases. E-learning databases often fit this description. [3]

**Data mining** “is a process that uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and subsequent knowledge from large databases”[9].

Data Mining can be used to extract knowledge from e-learning systems through the analysis of the information available in the form of data generated by their users. In this case, the main objective becomes finding the patterns of system usage by teachers and students and, perhaps most importantly, discovering the students' learning behavior patterns.

Several studies have demonstrated that Data Mining techniques could successfully be incorporated into E-learning environments. The application of data mining techniques and concepts in e-Learning systems helps to support educators to improve the e-Learning environment. Data mining techniques have also been addressed as complementary systems to LMS, and in particular to Moodle, where results are achieved through the use of associates, classifiers, clusters, pattern analyzers, and statistical tools. The scope of data mining is to discover useful knowledge using a variety of techniques such as prediction, classification, association rule mining, clustering, fuzzy logic, etc. [1]

“data mining is the process of analyzing data in order to discover implicit, but potentially useful information and uncover previously unknown patterns and relationships hidden in data.”[8]
2 Data mining and E-learning Aims to provide an up-to-date snapshot of the current State of research and applications of Data Mining methods in e-learning. The Cross-fertilization of both areas is still in its infancy, and even academic References are scarce on the ground, although some leading education-related Publications are already beginning to pay attention to this new field. In order to Offer a reasonable organization of the available bibliographic information According to different criteria, firstly, and from the Data Mining practitioner Point of view, references are organized according to the type of modeling Techniques used, which include: Neural Networks, Genetic Algorithms, Clustering and Visualization Methods, Fuzzy Logic, Intelligent agents, and Inductive Reasoning, amongst others. From the same point of view, the Information is organized according to the type of Data Mining problem dealt with: clustering, classification, prediction, etc. Finally, from the standpoint of the e-learning practitioner, we provide taxonomy of e-learning problems to Which Data Mining techniques have been applied, including, for instance: Students’ classification based on their learning performance; detection of Irregular learning behaviors; e-learning system navigation and interaction Optimization; clustering according to similar e-learning system usage; and systems’ adaptability to students’ requirements and capacities. [2]

3 Educational Data Mining (called EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. A key area of EDM is mining computer logs of
student performance. Another key area is mining enrollment data. Key uses of EDM include predicting student performance, and studying learning in order to recommend improvements to current educational practice. EDM can be considered one of the learning sciences, as well as an area of data mining. A related field is learning analytics. A list of the primary applications of EDM is provided by Cristobal Romero and Sebastian Ventura. In their taxonomy, the areas of EDM application are: [6]

- Analysis and visualization of data
- Providing feedback for supporting instructors
- Recommendations for students
- Predicting student performance
- Student modeling
- Detecting undesirable student behaviors
- Grouping students
- Social network analysis
- Developing concept maps
- Constructing courseware
- Planning and scheduling

Ryan Baker classifies the areas of EDM as follows: [7]

- Prediction
  - Classification
  - Regression
  - Density estimation
- Clustering
- Relationship mining
  - Association rule mining
  - Correlation mining
  - Sequential pattern mining
  - Causal data mining
- Distillation of data for human judgment
- Discovery with models
Fig. 2. The cycle of Applying Data Mining in Educational systems

4 Data mining functionalities for e-learning domain: in the e-learning domain, we are interested in managing mainly two groups of users: the learners as well as the learning providers, whether private training companies, governmental organizations and local authorities providing training for their employees or universities who aim to publish their courses and make them accessible online via the Internet. As for learners, databases should store all personal details including name, age, gender, address, postcode, and educational-relevant details such as qualifications. Moreover having information like work experience, career objectives, income range, previous courses taken and courses of interest would be of great value to be able to predict future behavior of different classes of employed professional people. Also other information such as personal interests and hobbies would be very valuable for data mining tool in order to discover hidden patterns by building intelligent models based on the huge amount of data. [4]
Web usage mining can be applied to e-learning domain as the site records information: learner profiles, web access information, academic details of students, and evaluation results. Web usage mining can track learning activities and identifies web access patterns and user behaviors. Web usage mining has a lot of contributions in e-learning domain such as [8],

(i) Dynamic personalization like providing real-time recommendations for e-learners
(ii) Commonly referenced web pages are cached in proxy servers.
(iii) Structuring or organizing the site structure according to learner’s interest.
(iv) Creating access shortcuts for interested pages to enhance user friendliness.
(v) Updating course content of web site according to previous usage information.
(vi) Identifying groups of learners of similar interest and sending personalized course materials to interested groups.

![Fig.3.Data mining systems components](image-url)
5 Processes of Data Mining in E-Learning

Cristóbal Romero et. al., [5] gave a detailed description about the application of data mining in e-learning systems is an iterative cycle in which 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 application of data mining in educational systems has specific requirements, mainly the need to take into account learners' specific behavior, including pedagogical aspects. The application of data mining in E-learning systems can be described as an iterative cycle where data mining applications contribute in enhancing learning, and also using mined knowledge for decision making.

The e-learning data mining process consists of the same four steps in the general data mining process (see Figure 1) as follows:
1. **Collect data.** The LMS system is used by students and the usage and interaction information is stored in the database. In this work we are going to use the students’ usage data of the Moodle system. Interaction information is stored in the database of the LMS.

2. **Preprocess the data.** The data is cleaned and transformed into an appropriate format to be mined. In order to preprocess the Moodle data, we can use a database administrator tool or some specific preprocessing tool. The data is transformed into an appropriate format.

3. **Apply data mining.** The data mining algorithms are applied to build and execute the model that discovers and summarizes the knowledge of interest for the user (teacher, student, administrator, etc.). In order to do so, we can use a general or a specific data mining tool, and we can use a commercial or free data mining tool. The data mining algorithms are applied to create and execute the model that discovers the knowledge and patterns of interest. In order to achieve this goal a data mining tool can be used.

4. **Interpret, evaluate and deploy the results.** The results or model obtained are interpreted and used by the teacher for further actions. The teacher can use the information discovered to make decisions about the students and the Moodle activities of the course in order to improve the students’ learning. The model obtained is interpreted and used by the educator for further analyses. The educator can use the information discovered to make decisions about e-learning system and process.
Where Does Data Mining Fit in e-Learning Processes?

Some researchers have pointed out the close relation between the fields of Artificial Intelligence (AI) and ML – main sources of Data Mining techniques and methods – and education processes. In [8], the author establishes the research opportunities in AI and education on the basis of three models of educational processes: models as scientific tool, are used as a means for understanding and forecasting some aspect of an educational situation; models as component: corresponding to some characteristic of the teaching or learning process and used as a component of an educative artefact; and models as basis for design of educational artefacts: assisting the design of computer tools for education by providing design methodologies and system components, or by constraining the range of tools that might be available to learners.

In studies on how Data Mining techniques could successfully be incorporated to e-learning environments and how they could improve the learning tasks were carried out. Data clustering was suggested as a means to promote group-based collaborative learning and to provide Incremental student diagnosis.

A review of the possibilities of the application of Web Mining (Web usage mining and clustering) techniques to meet some of the current challenges in distance education. The proposed approach could improve the effectiveness and efficiency of distance education in two ways: On the one hand, the discovery of aggregate and individual paths for students could help in the development of effective customized education, providing an indication of how to best organize the educator organization’s courseware. On the other hand, virtual knowledge structure could be identified through Web Mining methods: The discovery of Association Rules could make it
possible for Web-based distance tutors to identify knowledge patterns and reorganize the virtual course based on the patterns discovered. An analysis on how ML techniques – again, a common source for Data Mining techniques – have been used to automate the construction and induction of student models, as well as the background knowledge Applying Data Mining Techniques to e-Learning Problems 187 necessary for student modeling. In this work, the difficulty, appropriateness and potential of applying ML techniques to student modeling was commented.

6 The application of data mining in e-learning systems: can help to resolve different problems using different data mining techniques.

Data mining can be used to resolve classification problems in e-learning. Only a few data mining techniques can be applied to e-learning to resolve classification problem. The techniques are: fuzzy logic methods; artificial neural networks and evolutionary computation, graphs and trees; association rules; multi-agent and systems. Furthermore, application of data mining to resolve clustering problems in e-learning; includes artificial neural network and clustering.

Other data-mining techniques that can be used in e-learning are prediction techniques, visualization techniques, and case-based reasoning.

Data mining application in e-learning is an iterative cycle 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.
In this work we have shown how useful the application of data mining techniques in course management systems can be for online instructors. 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. The classifier shows what the main characteristics of the students in each group are, and it allows new online students to be classified. Finally, the instructors can apply association rule mining to discover if there is any relationship between these characteristics and other attributes. These rules can not only help to classify students, but also to detect the sources of any incongruous values obtained by the students.

7 Conclusions and future research

Educational data mining's an upcoming field related to several well-established areas of research including e-learning, adaptive hypermedia, intelligent tutoring systems, web mining, data mining, etc. The application of data mining in educational systems has specific requirements not present in other domains, mainly the need to take into account pedagogical aspects of the learner and the system.

Although the educational data mining is a very recent research area there is an important number of contributions published in
journals, international congress, specific workshops and some ongoing books that show it is one new promising area. Some of the most promising work line is the use of e-learning recommendation agents. These recommender agents see what a student is doing and recommends actions (activities, shortcuts, contents, etc.)

They think would be beneficial to the student. Recommender agents can also be integrated in evolving e-learning systems in which materials are automatically found on the web and integrated into the system. In this way, they help educators to detect which parts of existing materials from heterogeneous sources as the Internet are the best to use for composing new courses.

Besides recommenders can also be integrated with domain knowledge and ontologies, combining web mining and semantic web in semantic web mining. Semantic web mining is a successful integration of ontological knowledge at every stage of the knowledge discovery process. Educational data mining is a young research area and it is necessary more specialized and oriented work educational domain in order to obtain a similar application success level to other areas, such as medical data mining, mining e-commerce data, etc. We believe that some future researches lines are:

• **Mining tools more easy to use by educators or not expert users in data mining.** Data mining tools are normally designed more for power and flexibility than for simplicity. Most of the current data mining tools are too complex to use for educators and their features go well beyond the scope of what a educator may want to do.
So, these tools must have a more intuitive and easy to use interface, with parameter-free data mining algorithms to simplify the configuration and execution, and with good visualization facilities to make their results meaningful to educators and e-learning designers.

- **Standardization of methods and data.** Current tools for mining data from a specific course may be useful only to its developers. There are no general tools or re-using tools or techniques that can be applied to any educational system. So, a standardization of data, and the preprocessing, discovering and post processing tasks is needed.

- **Integration with the e-learning system.** The data mining tool has to be integrated into the e-learning environment as another author tool. All data mining tasks (preprocessing, data mining and post processing) have to be carried out into a single application. Feedback and results obtained with data mining can be directly applied to the e-learning environment.

- **Specific data mining techniques.** More effective mining tools that integrate educational domain knowledge into data mining techniques. Education-specific mining techniques can help much better to improve the instructional design and pedagogical decisions. Traditional mining algorithms need to be tuned to take into account the educational context.
References


