A Study on Educational Data Mining

*Smitha Harikumar
*Asst.Professor-MCA Dept, SNGIST, N.Paravoor, Kerala

ABSTRACT

Educational Data Mining 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.Currently there is very much interest in data mining and educational systems, making educational data mining as a new growing research community. This paper surveys the application of data mining to web-based courses 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, different data mining techniques can be applied.

Keywords: *Data mining, Educational Data mining. Exploration, modeling, pattern identification*

1.INTRODUCTION

Data mining is the automatic extraction of implicit and interesting patterns from large data collections. It can be used not only to learn the model for the learning process but also to evaluate and to improve e-learning systems by discovering useful learning information from learning portfolios.

Educational data mining is an interesting research area which extracts constructive, previously unknown patterns from educational database for better perceptive, improved educational presentation and judgment of the student learning process [14]. It is concerned with developing methods for exploring the unique types of data that come from educational environment which include students' results repository. Figure 1 shows the application of data mining in educational systems. C. Romero, S. Ventura introduced a survey from 1995 to 2005 for 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.

EDM can be considered one of the major area of data mining A key areas of EDM includes mining computer logs of student performance, During the past decades, the most important innovations in educational systems are related to the introduction of new technologies as web-based education. This is a form of computer-mining enrollment data of student, predicting student performance based on data, studying learning in order to recommend improvements to current educational practice, detecting undesirable student behaviors ,visualization and analysis of data , feedback facility for supporting instructors and recommendations for students, creating student modeling by clustering students behavior ,social network analysis , developing concept maps constructing courseware ,Planning and scheduling of course etc.

Currently there is very much interest in data mining and educational systems, making educational data mining as a new growing research community. This paper surveys the application of data mining to web-based courses 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, different data mining techniques can be applied.

aided instruction virtually independent of a specific location and any specific hardware platform.[9] It has considerably gained in importance and thousands of web courses have

been deployed in the past few years. But many of the current web-based courses are based on static learning materials, which do not take into account the diversity of students.

In conventional teaching environments, educators are able to obtain feedback on student learning experiences in face-to-face interactions with students, enabling a continual evaluation of their teaching programs. Decision making of classroom processes involves observing a student's behavior, analyzing historical data, and estimating the effectiveness of pedagogical strategies.
However, when students work in electronic environments, this informal monitoring is not possible; educators must look for other ways to attain this information. Organizations, which run distance education sites, collect large volumes of data, automatically generated by web servers and collected in server access logs.

1.1 Web-based learning environments

In the WBE context a "remote teacher" can't see the signs of understanding and confusion on the faces of the students. With this severe lack of feedback it becomes hard to identify , bright students who need to be challenged ,troubled students who need additional attention, as well as the parts of learning material that are too easy, too hard, or confusing.
WBE systems can track every action of the student, but it's almost unfeasible for a human teacher to make any sense of the large volume of data they are collecting. Intelligent class monitoring systems attempt to use Artificial intelligence It help the teacher in this context.
Web based learning environments are able to record most learning behaviors of the students, and are hence able to provide a huge amount of learning profile. Recently, there is a growing interest in the automatic analysis of learner interaction data with web-based learning environments.
In order to provide a more effective learning environment, data mining techniques can be applied. The objective of data mining in e-learning is to improving the learning. This goal is more subjective and more subtle to measure.

This stream of work was pioneered by Hyper Classroom (Oda, et al., 1998) that used fuzzy technology to identify "deadlocked" WBE students. Hyper Classroom is one of the examples in this class,

1.2 Adaptive and intelligent Web-based educational systems (AIWBES)

They provide an alternative to the traditional "just-put-it-on-the-Web" approach in the development of Web-based educational courseware (Brusilovsky & Miller, 2001). AIWBES attempt to be more adaptive by building a model of the goals, preferences and knowledge of each individual student and using this model throughout the interaction with the student in order to adapt to the needs of that student. They also attempt to be more intelligent by incorporating and performing some activities traditionally executed by a human teacher - such as training students. [15]
The first pioneer intelligent and adaptive Web-based educational systems were developed in 1995-1996 .Since then many interesting systems have been developed and reported. An interest to provide distance education over the web has been a strong driving force behind these research efforts.

2. TECHNIQUES

Educational systems have special characteristics that require a different treatment of the mining problem. As a consequence, some specific data mining techniques are needed to address in particular the process of learning. Some traditional techniques can be adapted and some cannot. The application of knowledge extraction techniques to educational systems in order to improve learning can be viewed as a formative evaluation technique. Formative evaluation is the evaluation of an educational program while it is still in development, and with the purpose of continually improving the program. Aiming how students use the system is one way to evaluate the instructional design in a formative manner and it may help the educator to improve the instructional materials.

Data mining techniques can discover useful information that can be used in formative evaluation to assist educators establish a pedagogical basis for decisions when designing or modifying an environment or teaching approach. The application of data mining in educational systems is an iterative cycle of hypothesis formation, testing, and refinement. 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.

Educators are in charge of designing, planning, building and maintaining the educational systems. Students use and interact with them. Starting from all the available information about courses, students, usage and interaction, different data mining techniques can be applied in order to discover useful knowledge that helps to improve the e-learning process. The discovered

2.1. The Cycle of Applying Data Mining in Educational System

Fig. 1. The cycle of applying data mining in educational systems.

2.1.1 Oriented towards educators.

The objective is to get more feedback for instruction, evaluating the structure of the course content and its effectiveness on the learning process, classify learners into groups based on their needs in guidance and monitoring, finding learner's regular as well as irregular patterns, find the most regularly made mistakes, find activities that are more successful for learning, discover information to improve the adaptation and customization of the courses, restructuring the sites for better courseware, organize the contents efficiently and constructing instructional plans, etc.

knowledge can be used not only by educators but also by students.

So, the application of data mining in educational systems can be oriented to different actors with each particular point of view. The objective is to recommend to learners activities, resources and learning tasks that would favour and improve their learning, suggest good learning experiences for the students or simply links to follow, based on the tasks already done by the learner and their successes, and on tasks made by other similar learners, etc.

2.1.2. Oriented towards academics responsible and administrators
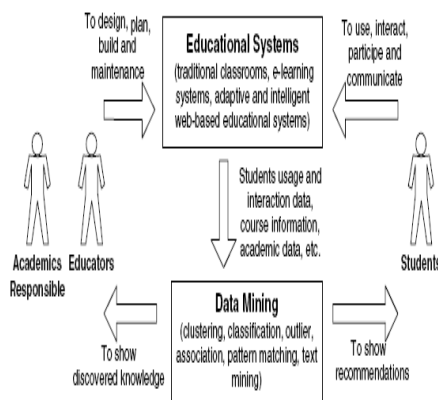
The objective is to find the suitable parameters about how to improve site efficiency and adapt it to the behavior of their users (optimal server size, network traffic distribution, etc.), have measures about how to better organize institutional resources (human and material) and their educational offer, enhance educational programs offer and determine effectiveness of the new computer mediated distance learning approach.

There are many data mining tools that provide mining algorithms, filtering and visualization techniques. Some examples of commercial and academic tool are DBMiner, Clementine, Intelligent Miner, Weka, etc. However these tools are not specifically designed and maintained for pedagogical purposes. Moreover, It is difficult for an educator who does not have an extensive knowledge in data mining to use these tools. In order to solve this problem, some specific educational data mining, statistical and visualization tools have been developed to help educators in analyzing the different aspects of the learning process.

3. APPLICATION OF DATA MINING TECHNIQUES IN EDUCATIONAL SYSTEMS

Data mining can be applied to data coming from two types of educational systems: traditional classroom and distance education. The data sources and objectives for both the type is different. So it is necessary to deal separately with the application of data mining techniques in each type.

3.1. Traditional classroom environments

They are the most widely used educational systems. It is based on face-to-face contact between educators and students organized through lecturers. In conventional classrooms, educators attempt to enhance instructions by monitoring student's learning processes and analyzing their performances by paper records and observation. They can also use information about student attendance, course information, curriculum goals etc. An educational institution has many diverse and varied sources of traditional databases with a student's information, educator's information, class and schedule information, online web pages, course content pages, multimedia databases, etc.

Data mining can help each actor of the learning process. Institutions would like to know which students will enroll in a particular course. An administrator may wish to find out information such as the admission requirements and to predict the class enrollment size for timetabling. Students may wish to know how best to select courses based on prediction of how well they will perform in the courses selected. Instructors may wish to know what learning experiences are most contributive to overall learning outcomes, why is one class out performing the other, similar groups of students, etc.
There are some works about the application of data mining in traditional education. They apply knowledge discovery in the form of statements ''Pattern X holds for data in Range Y'' to university databases. The results were presented to an administrator in order to make strategic decisions about the institutional policies. Another work on the use of data mining to identity and understand whether curriculum revisions can affect students in a university or not. They verify the qualitative impact of revisions and evaluate it using a number of techniques, such as summarization, association, classification.
In a related work, [11]the objective is to select the weak students to attend remedial classes , they use a scoring function that is based on association rules. First, they identify the potential weak students and then select the course that each weak student is recommended to take. Finally, an application in higher education for doing a comprehensive analysis of student characteristics is done.
There are different proposals to use different unsupervised and supervised data mining

algorithms to do clustering and prediction in order to enable educational institutions to better allocate resources and staff, manage student outcomes, and improve the effectiveness of alumni development.

3.2. Distance Education Environment

Distance education or distance learning consists of techniques and methods providing access to educational programs for students who are separated by time and space from lecturers. E-Learning systems lack a closer student–educator relationship (one to one). The distance education includes different subtypes such as paper-based correspondence education, videotape education, computer aided education, multimedia education, internet education or web-based education etc. Currently, the most used is web-based education allowing students to conveniently learn through the Internet. Web-based education is a form of distance education delivered over the Internet. Today, there are a lot of terms used to refer to web-based education such as e-learning, e-training, online instruction, web-based learning, web-based training, web-based instruction, etc. And there are different types of web-based systems: synchronous and asynchronous, collaborative and non-collaborative, closed corpus and open corpus, etc. These web-based education systems can normally record the student's accesses in web logs that provide a raw trace of the learners' navigation on the site.

3.3. Particular web-based courses

Particular web-based courses are specific courseware that uses standard HTML (Hyper Text Markup Language). There are a lot of courses, tutorials, etc. of this type on the Internet, and as another web site, they have the same kinds of data sources. Data mining can be used to know how students use the course, how a pedagogical strategy impacts different types of students, in which order the students study subtopics, what are the pages/topics that students skip, how much time the students spend with a single page etc.

These systems gather large log data of the students' activities and usually have built-in student monitoring features they can record whatever student activities it involves, such as reading, writing, taking tests, performing various tasks in real or virtual environments, even communicating with peers. They normally also

provide a database that stores all the systems information: personal information of the users (profile), academic results, user's interaction data, etc. Although some platforms offer reporting tools, when there are a great number of students, it becomes hard for a tutor to extract useful information.

Data mining can be applied to explore, visualize and analyze data in order to identify useful patterns and to evaluate web activity to get more objective feedback for your instruction and knowing more about how the students learning on the etc.

3.4. Adaptive and intelligent web-based educational systems

AIWES are the result of a joint evolution of intelligent tutoring systems (ITS) and adaptive hypermedia systems (AHS). The data from AIWEBS are semantically richer and can lead to more diagnostic analysis than data from

Adaptive and intelligent web-based educational systems have been seen as a solution to individually richer learning environments. These systems try to offer learners personalized education by building a model of the individual's goals, preferences, and knowledge.

4. Related Works

So many research works are going on in this area. Those works are assumed as literature study of this paper . Some of the works as given in the following table.[6]

Authors	Task	Educational system
Back andWoolf(2000)	Prediction	AIWBE system
Backer et al(2000)	Association and Classification	Traditional Education
Chen et al(2000)	classification	Web-based courses
Ha et al(2000)	Association	Web-based courses
Ma et al(2000)	Association	Traditional Education
Tang et al(2000)	Text mining	AIWBE system
Yu et al(2001)	Association	Web-based courses

traditional web-based education system .The available data come from the domain model ,pedagogical dataset (set of problems with their answer and complexity information), interaction log files (data related with user interaction) and student model (list of the satisfactions and violations constraints). Even if there is a standard model for monitoring, for the purpose of data mining, it is necessary to have a new model of student interaction with augmented information with contextual data. [3]

The student's interaction can be analyzed at a number of different layers of granularity: course, sessions, problems, attempts and constraints. Data mining can be used in order to know the causes of problems in the system, for example, incorrect feedback statements, to adapt the level to the progress of the learner, to suggest personalized learning experiences and activities for the students etc.

5. CONCLUSION

It is a great challenge for Synthesizing the vast amount of research and ideas and condensing them into one with the aim of introducing data mining to the institutional research audience in education. By using well defined algorithms from the disciplines of machine learning and artificial intelligence to decision rules, associations, and likelihood of events, data mining has profound application significance. If it were not for the fast, vast, and real-time pattern identification and event prediction for enhanced business purposes, there would not have been such an exponential growth in dissertations, models, and the considerable amount of investment in data mining in the corporate world. As we have discovered, insights from data sets and variable lists, previously seen as unwieldy and can be obtained with data mining and developed into the foundations for program planning or to resolve operational issues. The power of data mining lies in the fact that it simultaneously enhances output and reduces cost. achieved with no additional cost. Donations may correctly pinpoint the right donors and the right target amount. This saves campaign costs and increases the campaign's effectiveness. The ability to provide intervention to individual students who are seen as likely to drop out or to

transfer also holds value beyond cost and savings. Data mining conducted to predict the likelihood of an applicant's enrollment following their initial application may allow the college to send the right kind of materials to potential students and prepare the right counseling for them. The potential of data mining in education cannot be underestimated.

REFERENCES:

[1]. Brieman, L. "Comment." Statistical Science, 1994, 9(1), 38–42.

[2]De Veaux, R. "Data Mining: What's New, What's Not." Presentation at a Data Mining .
"An Exploratory Approach to Data Mining in Higher Education: A Primer and a Case Study." Paper presented at the AIR Forum, Seattle, Wash., 2000.

[3]Luan, J., and Willette, T. "Data Mining and Knowledge Management." Paper presented at the Research and Planning Group Conference, Lake Arrowhead, Calif., 2001.

[4]SPSS. SPSS Clementine 6.0 User's Guide..

[5] Westphal, C., and Blaxton, T. Data Mining Solutions: Methods and Tools for Solving Real-World Problems. Wiley Computer Publishing, 1998.

*[6]Educational data mining:- A survey , Expert systemswith application; C. Romero *, S. Ventura*

[7]Data mining concepts and techniques-Jiawei &Micheline Kamer.

[8] Peter Brusilovsky, Adaptive Hypermedia, UserModelling and Useradapted interaction, v,11 n,1-2,p,87-110 2001.

[9]R.M.Felder, L.K.Silvermann,Learning and Teaching Styles in Engineering Education 78(7), 674-681,1988.

[10]. Paramythis,S.Loidl_REisinger, Adaptive Learning and e-learning standards, Electronic journal of e-learning, 2(1), 181-194,2004.

[11].C.Romero and S.Ventura, Data mining in e-learning.

[12] F.Esposito, O.Licchelli, G.Semeraro – Discovering student models in e-learning systems, Journalof Universal Computer Science,10(1),47-57,2004.

[13].R.Koper, D.Burgos, DesigningLearning Activities fromcontent based to context based learning services, International journal of advanced Technology for learning 2(3), 2005.

[14]http://sci2s.ugr.es/keel/pdf/keel/articulo/CAE-VersionFinal.pdf
[15]http://citeseerx.ist.psu.edu/viewdoc/download?doi =10.1.1.103.702&rep=rep1&type=pdf

Author's Biography:

Smitha Harikumar has obtained her post graduate degree in Computer Applications and M.Phil in Computer Science from MK university ..She has completed her PhD in computer science and submitted her thesis to Karpagam University, Coimbatore. She is now working as an Asst.Professor in the Faculty of Computer Applications, Sree Narayana Guru Institute of Science and Technology, Paravoor, Kerala.. She is interested in data mining studies and its Applications. She has published 8 different articles in international journals with impact facor and also presented many papers in international and national seminars regarding data mining and warehousing. She has nearly 15 years of experience.