Pak. j. eng. technol. sci.
Volume 4, No 2, 2014, 104-114
ISSN: 2222-9930 print
ISSN: 2224-2333 online

Data Mining in Educational Settings

Rabab Naqvi
Department of Computer Science and Information Systems
Institute of Business Management
rabab.naqvi@iobm.edu.pk

Abstract

Educational mining is an evolving trend in research capacity that deals with the growth of tools and techniques to discover hidden pattern lying in the data in an educational context. Mining is helping all the educational stakeholders working in different lines of education. It is helping administrators to manage resources effectively and faculty to analyze students' performance, their learning behavior, and to predict their future path. Problem definition, data gathering and preparation, model building and evaluation, and knowledge development are the four steps typically involve in educational mining.

Keywords: Mining, Educational Mining, Pattern Analysis, Learning Behavior.

Introduction:

Educational data mining is quite a new discipline. However for a long time researchers have been doing a lot in analyzing data from educational software, merely now educational data mining has been recognized as a field in its own right (Oliver Scheuer, 2011).

Data mining(Jack Mostow, 2005) is extensively used in educational settings to explore the inimitable kinds of data that comes from educational background in order to understand students well and their learning environment.

Whether data, of educational nature, is taken from computer based learning systems, interactive learning environments or from other learning platforms, it always has numerous ranks of evocativeorder which could be analyze by characteristics of data. Issues of segmentation, context, categorization and timing also act as vital ingredients in the examination of academic data (Mostow, 2004).

Educational mining is used to discover the data in educational platforms to comprehend students and their environmentwhich they study in (Ryan S.J.D. Baker, 2009). At one side where the increase in educational software and large student information repositories (Kenneth R. Koedinger, 2008) are reflecting how students learning takes place. While on the other hand, the internet usage has emerged the new terminology of learning, known as e-learning or web-based learning in which big volumes of data about teacher-student communication are continually generated and accessible far and wide (Félix Castro, 2007).

Due to the special nature of educational data and there problems, mining of such data requires treatment in a dissimilar way. Even though the majority of the mining methods can be implemented directly while rests have to be tailored according to the definite educational problems. In addition particular data mining procedures can be used for explicit educational complications(Cristóbal Romero, 2010).

The users of educational mining involve different groups ofparticipants. Each group possesses different mission, vision and objectives therefore targets the information from different perspective(Hanna, 2004). For example mining methods not only help educators to administer with their classes, understanding their learners learning requirements and analyzing their personal teaching expertise but accommodating learner's perception on the circumstances and giving feedback to them as well(Agathe Merceron, 2005). Learners and instructors were the two groups that were initially deliberated in

educational data mining but later discovered other groups also beside these two(Cristóbal Romero, 2010).

Following are some users/actors of educational mining and some of their objectives for using data mining.

i. Learners/ Students: To suggest activities to learners that could be helpful in improving their learning, to recommend courses, etc.
ii. Educators/ Teachers/ Instructors: This class could use data mining to analyze student's performance, to determine the most effectual learning technique, evaluate learning behavior of learners, etc.
iii. Course Developers/ Educational Researchers: To enhance quality of learning, examine course structures and its impact on the learning process, etc.
iv. Organizations/ Institutions/ Universities/ Colleges/ Schools: To identify whether the objectives are achieved or not, curriculum and course designing, etc.
v. Administrators: To examine all the available resource, building the suitable technique to utilize them more effectively, to evaluate the most effective program, to evaluate teachers, etc.

Methodology:

This particular paper is a discussion paper based on discussing the role of data mining in education. In this paper I have discussed how data mining is useful in analyzing the set of educational activities and how it could help them the different stakeholders. Later part consists of some limitations of educational data mining.

Literature review was searched in different journals, web pages; cite seer websites, and internet. This discussion focuses data mining applications and its impact on the stream of education and how it is benefiting the different actors of education in providing the quality education.

Data mining definition:

Data mining or knowledge discovery is a branch of Computer Science that is extensively used to discover unseen patterns and associations among larger data(Usama Fayyad, 1996). It uses techniques & tools from statistics and artificial intelligence along with database management systems to examine the large volumes of data.From many years, it has been utilized by businesses, researchers, scientists, and governments to filter enormous data like census data, airline and railways passenger records, and sales of supermarket (Jiawei Han, 2006).

According to a research study conducted by (Raymond Kosala, 2000), tasks of data mining are divided into four subtasks:

1. Resource finding: selection of resource to be mine
2. Information selection and pre-processing: selection of information & its pre-processing
3. Generalization: discovering patterns and relationship among data
4. Analysis: interpretation of discovered patterns

A lot of work has been done in the field of data mining and successfully produced several methods, tools, and algorithms to handle big data to solve real-world problems. Mining is a vital part of number of applications such as business intelligence, decision support systems, data warehousing, bioinformatics, and predictive analysis. Key idea of using data mining process is to efficiently handle big data, mine actionable trends, and gain perceptive knowledge (Pritam Gundecha, 2012).

Pattern recognition, regression, clustering, statistics, and machine learning are the most common technique used in majority of mining methods. The primary intention of data mining is to predict and describe, as prediction involves variables from database that are unknown or tends to be the future values and description involves human interpretation of describing the data (Usama Fayyad, 1996).

Mining algorithms are diversified in their methods and intentions (Jiawei Han, 2006). Data searching and data visualization are the other

common features that are offered by data mining to help users to present data in an appropriate and easy way(Agathe Merceron, 2005).

Data mining in multidisciplinary fields:

(Usama Fayyad D. H., 1996)Data mining has its applications in many diverse areas and specific techniques are being developed to suit the nature of different areas.

1. PRODUCTION:The idea of using data mining in manufacturing industry is not new. Technique of converting data into useful information started using in 1990 and now it's been 24 years that the production community has utilizing it for different steps of manufacturing process. It is used in industrial engineering to mine useful insights in preemptive maintenance, designing, revealing faulty lines, production, quality maintenance, and forecasting. Data can be evaluated to uncover the unseenoutlines that regulate the production process and enhancing the quality of products. Data can be collected during the daily operation process and thus not required any implementation of dedicated software's (J. A. Harding, 2006).
2. SCIENTIFIC AND ENGINEERING MINING:Data mining is benefiting a wide range of applications from scientific and engineering fields. Remote sensing, data simulation, structural mechanics, material science, fluid dynamics, and semi-conductor are the few areas implementing mining techniques for scientific and engineering applications (Robert L. Grossman, 2001).
3. MEDICINE:Hospitals are now equipped with devices for collecting and recording huge amount of data in inter and intra hospitals health management system (Nada Lavrac, 2005). Data mining is used in clinical medicine to construct models that are based on patient medical information to forecast the disease and thus support clinical decision making (Bellazzo, 2008).
4. MARKETING: Web has made it easier to collect huge amount of data. Since the choices available to customers are increasing day by day, therefore companies are now focusing their customers to release the marketing pressure. Data mining is helping organizations to understand customer behavior, their needs, buying habits, and preferences (Michael J. Shaw, 2001).

5. ASTRONOMY:

Astronomers analyzes the physical properties of objects depending upon the collected data on that object (Borne, 2009). These mining techniques are the reason why astronomers suggested that there exists other solar systems also in other galaxies. Adding to it they also believe that there are billions of other planetary systems like ours in this universe(Haydock, 2003). Astronomers are able to unveil the fact by using data mining techniques on available data to generate this knowledge.

Data mining in educational settings:

Educational mining is an effective and powerful tool. Implication of data mining in education is a continuous cycle in which the extracted knowledge enters in to the system and not only to guide, expedite and improve the process of learning and teaching but also to take decisions (Cristóbal Romero S. V., 2008).

Educational mining has been classified including learning from educational software, machine based learning, online testing system and the factors influencing student failure or not performing better in the course (Ryan S.J.D. Baker, 2009).

Through analyzing of educational data one could easily observe the learning behavior of students. Attributes like student skills, interest or confidence are helpful in determining the overall progress of student(K.Shanmuga Priya, 2013). For example a research study carried out by (Jin Soung Yoo, 2012), in which they were found, through mining, that most of the college students learn through “me” initiated actions like “me reading”, “me writing”, “me asking”, “me-notes”, “me-understanding” or “me-books”. Learning in groups or with the interaction of teachers was not observed. This kind of study is helpful

in generating ways to create better learning environment for students in which they are comfortable to learn.

Mining is also used to analyze the student's performance in educational settings by evaluating their final examinations or internal assessments conducted. An internal assessment involves quizzes, workshops, attendance or laboratory work. Besides assessments communication and presentation skills are also the key factors of analyzing students' performance (K.Shanmuga Priya, 2013).

Educational mining is also used in planning and scheduling of typical educational process. It helps in planning future courses, planning resource allocation, course development to assist in the admission process, curriculum designing, and etc. Decision trees are used to analyze the course by evaluating no of students completing the course or students preferences in taking the course.

As competition in the education sector has also began, therefore institutes have to adopt policies that results in there high enrollment and achieve there set goals. It is a concept that is used to exert students influence over their programs. This exercise involves student retention programs, marketing strategies, admission policies and financial aid programs. Mining helps institutes to generate useful information in all these areas(Monika Goyal, 2012).

Data mining techniques of clustering and classification are used to split students into groups according to their demography. This could result in designing a custom-made learning group that can endorse affective group studying.

Conclusion:

Predicting students' performance is a very major and significant aspect in educational settings. Personal, social, psychological and other environmental factors are the key factors influencing students learning ability (Brijesh Kumar Bhardwaj, 2011).Educational data mining has become the handiest tools in shaping today's education sector. It is regarded as an upcoming research area.Numbers of techniques are used in educational mining to support educational process applications(Claus Pahl, 2002). These techniques are used in academics

to help understand the students' performance, their leaning ability and behavior, preferences, designing course curriculum, resource allocation and faculty evaluation. Analyzing these factors could help in improving learning environment and better utilizing the available resources.

References:

Agathe Merceron, K. Y. (2005). Educational Data Mining: a Case Study. *International Conference on Artificial Intelligence in Education,* (pp. 1-8). Amsterdam.

Andreas M. Kaplan, M. H. (2010, January-February). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59-68.

Anwar M. A., N. A. (2012, November). Information Mining in Assessment Data of Students' Performance. *International Journal of Engineering Science and Innovative Technology (IJESIT), 1*(2), 207-212.

Bellazzo, R. (2008). Predictive data mining in clinical medicine: Current isseus and guidelines. *International Journal of Medical Informatics, 77*(2), 81-97.

Borne, K. D. (2009). *SCIENTIFIC DATA MINING IN ASTRONOMY.* arXiv preprint arXiv: 0911.

Brijesh Kumar Bhardwaj, S. P. (2011, April). Data Mining: A prediction for performance improvement using classification. *International Journal of Computer Science and Information Security, 9*(4).

Claus Pahl, D. D. (2002). Data Mining Technology for the Evaluation of Web-based Teaching and Learning Systems.

Cristóbal Romero, S. V. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education, 51*(1), 368-384.

Cristóbal Romero, S. V. (2010). Educational Data Mining: A Review of the State-of-the-Art. *Systems, Man, and Cybernetics, Part C: Applications & Reviews, IEEE Transactions, 40*(6), 601-618.

Derek E. Baird, M. F. (2005-2006). NEOMILLENNIAL USER EXPERIENCE DESIGN STRATEGIES: UTILIZING SOCIAL NETWORKING MEDIA TO SUPPORT "ALWAYS ON" LEARNING STYLES. *J. EDUCATIONAL TECHNOLOGY SYSTEMS, 34*(1), 5-32.

Félix Castro, A. V. (2007). Applying Data Mining Techniques to e-Learning Problems. *Studies in Computational Intelligence, Springer*, 183-221.

Hanna, M. (2004). Data Mining in the e-learning Domain. *Campus-Wide Information Systems, 21*(1), 29-34.

Haydock, M. P. (2003). Data Mining in Astronomy. In G. F. Tito A. Ciriani, *Operations Research in Space and Air* (pp. 143-160). Dordrecht: Kluwer Academic Publishers.

J. A. Harding, M. S. (2006). Data Mining in Manufacturing:A Review. *Journal of Manufacturing Science and Engineering, 128*, 969-976.

Jack Mostow, J. B. (2005). An Educational Data Mining Tool to Browse Tutor-Student Interactions: Time Will Tell! *Educational Data Mining, Natioanal Conference on Artificial Intelligence* (pp. 15-22). Maceio: Springer.

Jiawei Han, M. K. (2006). *Data Mining: Concepts and Techniques.* Morgan Kaufmann Publishers.

Jin Soung Yoo, M.-H. C. (2012). Mining Concept Maps to Understand University Students' Learning. *International Educational Data Mining Society*.

K.Shanmuga Priya, A. K. (2013). Improving the Student's Performance Using Educational Data Mining. *International Journal of Adanced Networking & Applications, 4*(4), 1680-1685.

Kenneth R. Koedinger, K. C. (2008). An open repository and analysis tools for fine-grained, longitudinal learner data. *EDM*, 157-166.

Kirsi Silius, T. M. (2010). Students' Motivations for Social Media Enhanced Studying and Learning. *Knowledge Management & E-Learning: An International Journal, 2*(1), 51-67.

Michael J. Shaw, C. S. (2001). Knowledge management and data mining for marketing. *Decision Support Systems, 31*(1), 127-137.

Monika Goyal, R. V. (2012, March). Applications of Data Mining in Higher Education. *I International Journal of Computer Science Issues, 9*(2), 113-120.

Mostow, J. (2004). Some Useful Design Tactics for Mining ITS Data. *ITS2004 workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes* (pp. 20-28). Maceio: Springer.

Nada Lavrac, B. Z. (2005). Data Mining in Medicine. In L. R. Oded Maimon, *Data Mining and Knowledge Discovery Handbook* (pp. 1107-1137). Springer.

Oliver Scheuer, B. M. (2011). Educational Data Mining. *Encyclopedia of the Sciences of Learning, Springer*, 1075-1079.

Pritam Gundecha, H. L. (2012). Mining Social Media: A Brief Introduction. *Tutorials in Operaion Researchs*, 1-17.

Raymond Kosala, H. B. (2000). Web Mining Reserch: A Survey. *ACM Sigkdd Explorations Newsletter, 2*(1), 1-15.

Robert L. Grossman, C. K. (2001). *Data Mining for Scientific and Engineering Applications.* Dordrecht: Kluwer Academic Publishers.

Ryan S.J.D. Baker, K. Y. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Eeducational Data Mining, 1*(1), 3-17.

Usama Fayyad, D. H. (1996). Mining Scientific Data. *Communications ofthe ACM, 39*(11), 51-57.

Usama Fayyad, G. P.-S. (1996). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence, 17*, 37-54.