Scholars Journal of Arts, Humanities and Social Sciences **ISSN 2347-5374 (Online)**
Sch. J. Arts Humanit. Soc. Sci. 2014; 2(3B):440-444 **ISSN 2347-9493 (Print)**

(An International Publisher for Academic and Scientific Resources) DOI: 10.36347/sjahss.2014.v02i03.014

Enhancing Education Quality Using Educational Data

Sedigheh Abbasnasab Sardareh[1*], Mohd Rashid Mohd Saad[1], Abdul Jalil Othman[1], RosalamChe Me[2]
[1]Faculty of Education, University of Malaya, Kuala Lumpur, Malaysia
[2]Department of Design, Politecnico di Milano, Italy

***Corresponding Author:**
Sedigheh Abbasnasab Sardareh
Email: abbasnasab@gmail.com

Abstract: Educational data mining (EDM) is an interdisciplinary research area which is concerned with developing methods and techniques to explore big educational data. Due to the persistent growth and increasing availability of educational data, EDM facilitates data-driven decision making for enhancing learning and instruction. In the first place, this analytical study provides an introduction to EDM. Then, the researchers will look at various application areas of data mining in education, and lastly major challenges in mining big educational data are presented. The information provided in this paper enables educators to understand how big data helps them in improving the teaching and learning processes.
Keywords: Learning, Instruction, Educational data mining, Educational System, Dataset

INTRODUCTION

Educational data is growing in a consistent manner. Nowadays, educational institutions such as universities, schools, and learning development centres store massive amounts of data both in structured and unstructured formats. In fact, velocity of various data types in different formats creates new challenges for researchers and educators.

The concept of big data refers to voluminous, enormous quantities of data both in digital and physical formats that can be stored in miscellaneous depositories such as records of students' tests or examinations as well as bookkeeping records [1]. These records are increasing rapidly and significant decisions are being made based on such big data [2]. In 2001, Doug Laney defined big data using three characteristics; volume, velocity, and variety [1, 3].

Volume refers to increase in the amount of data. Enormous quantities of data are being produced from different data repositories. Big data can comprise data gathered from different devices networked together over internet which is called Internet of Things (IoT). Hugein formation available on social network sites such as Facebook and twitter are also considered as big data [4].

Velocity is attributed to the rapid increase in different data types (structured & unstructured) over time. As a result of globalization there is an upward frequency of data capture and decision making based on that data is of utmost importance [1]. The velocity of social network use is increasing day by day. Decisions made based on such big data would essentially impact the next data obtained and analysed.

Variety means using several data types to explore a situation or an event. Internet users produce numerous set of structured, semi-structured and unstructured data. Very large number of devices and sensors on the IoT engender not only large volumes of data but also various types of data [5]. These three dimensions are illustrated in Fig. 1.

Fig. 1: The 3Vs that define big data

Up till now, many have been written on the fruitfulness of big data in different disciplines such as business [2], however, little if any, attention has been paid to the effectiveness of applying big data in education section. This is what we aim to find in this study.

One particular field which is significant for extracting big educational data is educational data

mining (EDM). EDM tries to reveal new patterns in data, and build new models or algorithms [1]. This analytical study provides an introduction to the concept of EDM, describes different application areas of data mining in education, and identifies issues and challenges extracting knowledge from large datasets using different data mining techniques.

Educational Data Mining

Educational data mining which is known as EDM, has been defined by several researchers. [6, p. 3] state that "EDM is both a learning science, as well as a rich application area for data mining, due to the growing availability of educational data. It enables data-driven decision making for improving the current educational practice and learning material". Similarly, [7, 8] define EDM as the application of data mining (DM) techniques to specific type of dataset that come from educational environments to address important educational questions.

EDM makes use of data mining algorithms over various educational data types with the aim of analysing these data and find solutions to educational issues [10, 7]. In other words, EDM is mainly concerned with developing methods and techniques to explore the relationship between different data types gathered from educational settings to gain better understanding of students and their learning context. Fig. 2 demonstrates the application of data mining in educational system.

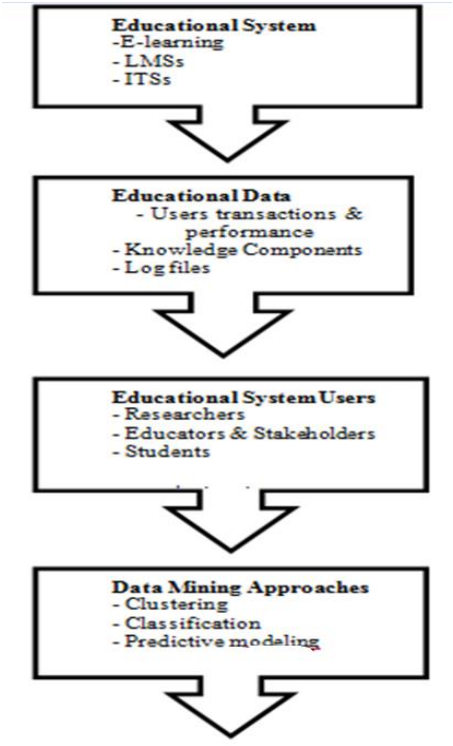

Fig. 1: Application of Data Mining in Education System

Rapid increase in databases of students' learning information and instrumental educational software has generated hugerepository of data on student learning [10]. Large amount of data on teacher-student interactions are also produced in e-learning and web-based education contexts [8]. All this type of information makes a precious mine of educational data available for use. EDM seeks to explore and analyse these untouched data repositories to provide a better appreciation of students' learning abilities and to develop computational approaches that merge theory and data for the purpose of transforming instruction to learners' needs [7].

To persistently improve the education programs and enhance learning effectiveness, data mining approaches can be applied to the design of education programs and it is considered as formatively evaluating the program in its development phase. Observing the way students perform in the education system is a common way to assess instructional system design which creates various data types such as log files, and teacher-student transactions and help stakeholders to improve instructional materials [9]. Information obtained from applying data mining approaches on education programs should enable instructional designers to come up with educational foundation applicable for designing and improving instructive approaches in varied contexts.

As shown in Fig. 2, data mining application to instructional design system is a cyclic procedure of developing a hypothesis, testing it and making improvements. Knowledge extracted during this procedure should contribute to learning improvement as a whole. In other words, during this procedure the aim is not only to transform data into knowledge but also to use this mined data for decision making purposes [11, 12]. According to figure 2, instructional system designers and educators plan educational systems. Then users such as learners make use of the system. Data mining techniques can be applied on the information gained from courses, learners' usage as well as their interactions. The extracted knowledge helps not only instructional system developers and education stakeholders but also the final users of the system-learners [9].In the next part, the researchers discuss different application areas of data mining in education in more detail.

APPLICATION AREAS OF DATA MINING IN EDUCATION

It has been [7] proposed that EDM objectives can be classified based on the viewpoint of the end user (learners, educators, researchers, and stakeholders) and the problem to be solved:

- Learners: To enhance self-reflection among students, give them feedback on their work, satisfy their needs and improve their learning performance.
- Educators: To gain a better understanding of their students' learning processes and reflect on their own instruction and enhance their teaching performance.
- Researchers: To design and develop data mining algorithms to be used to analyse different educational datasets and recommend the most useful one for specific educational problem/task.
- Stakeholders: To find the best way to organize institutional resources and their educational offer.

This perspective distinctly demonstrates the end user benefits from EDM applications in different ways, yet, it is difficult to categorize EDM application goals based on merely these users particularly when a goal relates to more than one user. Therefore, based on the work of Romero C *et al*. [9] that centered on the relevant research goal of EDM applications, the following EDM goals can be recognized:

- Student modeling. User modeling in the educational domain incorporates such detailed information as students' characteristics or states such as knowledge, skills, motivation, satisfaction, meta-cognition, attitudes, experiences and learning progress, or certain types of problems that negatively impact their learning outcomes (making too many errors, misusing or under-using help, gaming the system, inefficiently exploring learning resources, etc.), affect, learning styles, and preferences. The common objective here is to create or improve a student model from usage information.
- Predicting students' learning performance and learning outcomes. The objective is to predict a student's final grades or other types of learning outcomes (such as retention in a degree program or future ability to learn), based on data from course activities.
- Generating recommendations. The objective is to recommend to students which content (or tasks or links) is the most appropriate for them at the current time.
- Analysing learners' behaviour. This takes on several forms: Applying educational data mining to answer questions in any of the three areas previously discussed (student models, Prediction, Generating recommendation). It is also used to group student according to their profile, and for adaptation and personalization purposes.
- Communicating to stakeholders. The objective is to help course administrators and educators in analysing students' activities and usage information in courses.
- Domain structure analysis. The objective is to determine domain structure and improving domain models that characterize the content to be learned and optimal instructional sequences, using the ability to predict the student's performance as a quality measure of a domain structure model. Performance on tests or within a learning environment is utilized for this goal.
- Maintaining and improving courses. It is related to the two previous goals. The objective here is to determine how to improve courses (contents, activities, links, etc.), using information (in particular) about student usage and learning.
- Studying the effects of different kinds of pedagogical support that can be provided by learning software.
- Advancing scientific knowledge about learning and learners through building, discovering or improving models of the student, the domain and the pedagogical support.

The above mentioned goals aim to improve different aspects of educational system in general. Nowadays, different data sources such as administrative databases that store students' enrolment information; fine grained data stored on LMSs; and students' performance data over different learning tasks and educational games stored on ITSs, are available on different data shops like PSLC Data Shop (The world's largest repository of learning interaction data), EdStats (A World Bank Initiative), Data.gov (official website of United States Government on Educational data sets), and UNISTATS Dataset (website provides comparable sets of information about full or part time undergraduate courses and is designed to meet the information needs of prospective students). To accomplish EDM goals, this data can be analyzed using the methods below [7]:

- Predictive modeling: The aim is to create a model that can deduce one aspect of the data (predicted variable) from a combination of other aspects (predictor variables).
- Clustering: Splitting the dataset into categories by putting subjects with similar characteristics in the same group.
- Relationship mining: Aims at finding the relationship between different variables in a dataset.
- Distillation of data for human judgement: Using approaches such as summarization, visualization, and interactive interfaces to depict useful information in datasets and

support decision making based on such information.

- Discovery with models: applying validated models of a particular phenomenon (e.g., prediction or clustering) as component in further analysis such as prediction or relationship mining.
- Outlier detection: To detect data points which are considerably different than the rest of data.
- Social network analysis: SNA or structural analysis, aims at studying relationships between individuals, instead of individual attributes or properties.
- Process mining: To extract process related knowledge from event logs recorded by an information system to have a clear visual representation of the whole process.
- Text mining: focuses on finding and extracting useful or interesting patterns, models, directions, trends, or rules from unstructured text documents such as HTML files, chat messages and emails.
- Knowledge tracing: method for estimating student mastery of skills that has been used in effective cognitive tutor systems.

As mentioned, various number of are used to analyse data produced in educational systems. The choice of an appropriate technique depends on the nature of the learning context, data types, and research objectives. In what follows the authors discuss some challenges of mining educational data.

PROBLEMS FACED

Rampancy of online learning systems offers researchers and educators the opportunity to construct personalized e-learning systems but the problem is that these systems are being used by learners with different learning styles and abilities. When every individual student works with the system, large amounts of log files would be generated [13]. To mine such data, it is important to find out and analyzethe right dataset to get rational results from it. System designers also need to be able to categorize individual learning styles to create an accurate personalized e-learning environment.

Another issue would be developing particular data mining algorithms that suit learning domain [9]. Still traditional algorithms are used to mine educational data and such algorithms may not fit educational datasets, therefore, Specific algorithms that work with educational datasets are needed.

Moreover, due to the fact that big educational data are usually put in storage in different sources and locations, and also because of rapid growth in volumes of big data, it would be difficult to move data across various locations and sources and load it into one main memory [3]. Much more effort is needed to eliminate the above mentioned barriers.

CONCLUDING REMARKS

This paper provided an introduction to the evolving field of EDM. The importance of mining big educational data was discussed and various application areas of data mining in education were presented throughout the paper. Challenges of extracting knowledge from big educational data were addressed as well. Recently, there has been an increased interest in working with big data using data mining techniques in different sectors and education sector is no exception. Assessing and making students' learning and assessment activities visible enhance self-reflection among them. Teachers gain insight into students' performance and use this information to adapt their instruction to meet learner needs. In the e-learning context, data mining techniques help educators see the effectiveness of their intervention and provide constant feedback for improvement. The importance of mining big educational data obtained from large datasets applying various data mining techniques worth further research.

REFERENCES

1. Sagiroglu S, Sinanc, D; Big data: A review. Proceedings of International Conference on Collaboration Technologies and Systems (CTS), 2013: 42-47.
2. Peña-Ayala A; Educational data mining: A survey and a data mining-based analysis of recent works. Expert Systems with Applications, 2014; 41(4): 1432-1462.
3. Trelles O, Prins P, Snir M, Jansen RC; Big data, but are we ready? Nature reviews Genetics, 2011; 12: 224-224.
4. Centola D; The spread of behavior in an online social network experiment. Science, 2010; 329:1194-1197.
5. LaValle S, Lesser E, Shockley R, Hopkins MS, Kruschwitz N; Big data, analytics and the path from insights to value. MIT Sloan Management Review, 2011; 52: 21-31.
6. Calders T, Pechenizkiy M; Introduction to the special section on educational data mining. ACM SIGKDD Explor, 2011; 13(2): 3–6.
7. Romero C, Ventura S; Educational data mining: a review of the state of the art. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., 2010; 40(6): 601–618.
8. Romero C, Ventura S; Data mining in education. Wiley Interdisc. Rev.: Data Min. Knowl. Discovery, 2013; 3(1):12–27.
9. Romero C, Ventura S, Pechenizkiy M, Baker RS; Handbook of educational data mining.CRC press,Boca Raton, 2011.
10. Baker R, Yacef K; The state of educational data mining in 2009: A review and future

visions. Journal of Educational Data Mining, 2009; 1: 3-17.
11. Han J, Kamber M, Pei J; Data mining: concepts and techniques. 2nd edition, The Morgan kaufmann Series in Data Management Systems, Elsevier Science Philadelphia,2006.
12. Liao SH, Chu PH, Hsiao PY; Data mining techniques and applications–A decade review from 2000 to 2011. Expert Systems with Applications, 2012; 39: 11303-11311.
13. Chen MS, Park JS, Yu PS; Data mining for path traversal patterns in a web environment. In Distributed Computing Systems, Proceedings of the 16th International Conference on, 1996: 385-392.