

APPLYING DATA MINING TECHNIQUE IN EDUCATION A REVIEW

Suganya S¹, Dr.V.Narayani²

¹Research Scholar, Bharathiar University, Coimbatore.

²Director i/c- MCA, Karpagam College of Engineering, Coimbatore

ABSTRACT

Applying data mining (DM) in education is an emerging interdisciplinary research field also known as educational data mining (EDM). It is concerned with developing methods for exploring the unique types of data that come from educational environments. Its goal is to better understand how students learn and identify the settings in which they learn to improve educational outcomes and to gain insights into and explain educational phenomena. Educational information systems can store a huge amount of potential data from multiple sources coming in different formats and at different granularity levels. Each particular educational problem has a specific objective with special characteristics that require a different treatment of the mining problem. The issues mean that traditional DM techniques cannot be applied directly to these types of data and problems. As a consequence, the knowledge discovery process has to be adapted and some specific DM techniques are needed. This paper introduces and reviews key milestones and the current state of affairs in the field of EDM, together with specific applications, tools, and future insights.

Keywords: *Eductaion, Mining techniques, Learning Analytics, Applications, current state*

I. INTRODUCTION

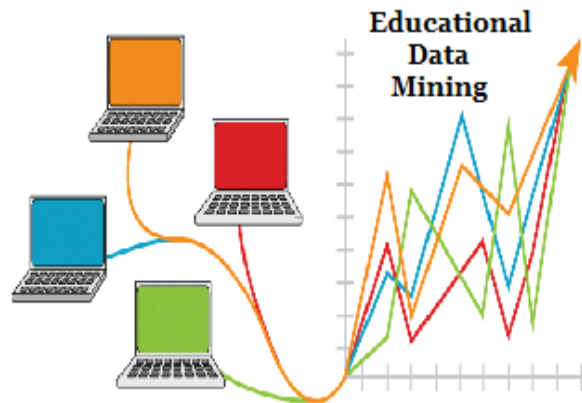
One of the primary goals of any educational system is to equip students with the knowledge and skills needed to transition into successful careers within a specified period. How effectively global educational systems meet this goal is a major determinant of both economic and social progress. Data mining provides many techniques for data analysis. The large amount of data currently in student databases exceeds the human ability to analyze and extract the most useful information without help from automated analysis techniques. Knowledge discovery (KD) is the process of nontrivial extraction of implicit, unknown, and potentially useful information from a large database. Data mining has been used in KD to discover patterns with respect to a users needs. The pattern definition is an expression in language that describes a subset of data. An example of a KD pattern definition appears in [1]

Educational data mining methods often differ from methods from the broader data mining literature, in explicitly exploiting the multiple levels of meaningful hierarchy in educational data. Methods from the psychometrics literature are often integrated with methods from the machine learning and data mining literatures to achieve this goal.

For example, in mining data about how students choose to use educational software, it may be worthwhile to simultaneously consider data at the keystroke level, answer level, session level, student level, classroom level, and school level. Issues of time, sequence, and context also play important roles in the study of educational data.

II.GOALS OF EDUCATIONAL DATA MINING

Predicting student's future learning behavior-With the use of student modeling, this goal can be achieved by creating student models that incorporate the learner's characteristics, including detailed information such as their knowledge, behaviors and motivation to learn. Discovering or improving domain models. Through the various methods and applications of EDM, discovery of new and improvements to existing models is possible.



Studying the effects of educational support-It can be achieved through learning systems. Advancing scientific knowledge about learning and learners-By building and incorporating student models, the field of EDM research and the technology and software used.

III.EDUCATIONAL DATA MINING AND APPLICATIONS

The main goal of EDM is to extract useful knowledge from educational data including student records, student usage data, intelligent , and LMS systems. The extracted knowledge can improve the process of teaching and learning in the educational system .It can also lead to the development of new teaching processes. Similar ideas have been applied successfully in different domains of knowledge. For example, e-commerce systems and basket analysis are popular applications in data mining [2]. Romero & Ventura[4] reviewed the EDM articles and suggested that future EDM research focus on the following aspects: - integrate EDM tools with e-learning systems , standardize data and models , make EDM tools easier for educators and non-expert users , customize traditional mining algorithms for educational context.

IV.CHALLENGES OF EDUCATIONAL DATA MINING

The research trends on EDM since the year 1998 to 2015 and found that maximum research focuses were on academic objectives. The other issues are:

- 1) Educational data is incremental in nature: Due to the exponential growth of data, the maintaining the data in data warehouse is difficult. To monitor the operational data sources, infer the student interest, intentions and its impact in a particular institution is the main issue. Another issue is the alignment and translation of the incremental educational data. It should focus on appropriating time, context and its sequence. Optimal utilization of computing and human resources is another issue of incremental.
- 2) Lack of Data Interoperability: Scalable Data management has become critical considering wide range of storage locations, data platform heterogeneity and a plethora of social networking sites. E.g: Metadata Schema Registry is a tool to enhance Meta data interoperability. So there is a need to design a model to classify/ cluster the data or find relationships. Examples of clustering applications are grouped students based on their learning and interaction patterns used in and grouping users for purposes of recommending actions and resources to similar users. It is possible to introduce Neuro Fuzzy mining technique to remove the gap of data interoperability.
- 3) Possibility of Uncertainty: Due to the presence of uncertain errors, no model can predict hundred percent accurate results in terms of student modeling or overall academic planning.
- 4) Research Expertise Relation between Student-Teacher. In most of the higher Educational institutions (e.g. Engineering Institutions) final year students have a compulsory project work which area research work based on their area of interest. Generally Supervisors are assigned as per availability and area of expertise in the respective department. But still it is not possible to assign all the students –supervisor with similar area of interest hence the result of the project is not applicable to real scenarios. There is need to find the relation between areas of interest, students' interest, applicability of the project/research and mining cross faculty interest. It will be beneficial to introduce using Association Mining to optimize this issue [3].

V. PRESENT STATE OF ART

The educational data needs proper preprocessed before being fed as input to the model. This is because the model is trained first in order to be put to use. Usually the data is divided into two sets , training set and test set. The training set is used to train the model and the test data is then used to check the accuracy of the resultant model. A large number of experiments and models have been developed and tested on educational data in order to extract knowledge about the quality of learning and related parameters. The models are being further researched upon to increase the success rate of prediction of the target parameters in turn helping the students in acquiring knowledge in an easier way, to aid faculty in understanding the progress of their students. [5] The paper provide a prediction of applying data mining technique to identify whether students' online learning experiences can be assessed based on their log files but It is limited to the available data in online database while factors such as students' position in the collaborative group and Structure of the collaborative tasks is not considered. The benefits of educational analytics include aiding in predicting the success rate of students, factors which results in students opting for a transfer to other universities, generating hierarchical list of time tables with an increasing order of difficulty based on the prior knowledge of the student.

Various data mining approaches which include classification algorithms, regression algorithms, and association rule mining algorithms, clustering algorithms are being used to extract knowledge from the

educational data. Many data analytics tools like SPSS Clementine 12.0, Rapid Miner, Weka data mining tool, R data mining software tool etc. are widely being employed for mining educational data.

VI. DATA MINING MODELS

A few of the many model functions being incorporated in KDD include:

Classification: It consists of predicting a certain outcome based on a given input[10]. Mapping or classifying data into one of several predefined classes.[6] For example, a bank may establish classes based on debt to income ratio. The classification algorithm determines within which of the two classes an applicant falls and generates a loan decision based on the result. Classification is the most commonly used data mining technique which contains a set of pre classified samples to create a model which can classify the large set of data. This technique helps in deriving important information about data and metadata (data about data).

Regression: A regression algorithm estimates the value of the target as a function of the predictors for each case in the build data. It is "a learning function which maps a data item to a real-valued prediction variable." [7] Comparing a particular instance of an electric bill to a predetermined norm for that same time period and observing deviations from that norm is an example of regression analysis. These relationships between predictors and target are summarized in a model, which can then be applied to a different data set in which the target values are unknown. Regression models are tested by computing various statistics that measure the difference between the predicted values and the expected values.

Clustering: Clustering is the grouping of a particular set of objects based on their characteristics, aggregating them according to their similarities. Regarding to data mining, this methodology partitions the data implementing a specific join algorithm, most suitable for the desired information analysis. It "maps a data item into one of several categorical classes (or clusters) in which the classes must be determined from the data, unlike classification in which the classes are predefined. Clusters are defined by finding natural groupings of data items based on similarity metrics or probability density models." [8] An example of this technique would be grouping patients based on symptoms exhibited. The clusters need not be mutually exclusive.

Summarization: Summarization is a key data mining concept which involves methods for finding a compact description for a dataset. A simple example would be tabulating the mean and standard deviations for all features. Such summarization techniques are often applied for exploratory data analysis, data visualization and automated report generation where the actual data is huge and cannot be handled manually. Generating a concise description of the data. Routine examples of these techniques include the mean and standard deviation of specific data elements within the dataset. **Dependency modeling:** developing a model that shows a how variables are interrelated. An example would be a model showing that electrical usage is highly correlated with the ambient temperature.

VII. CONCLUSION

EDM is both a learning science, as well as a rich application area for data mining, due to the growing availability of educational data. EDM contributes to the study of how students learn, and the settings in

which they learn. It enables data-driven decision making for improving the current educational practice and learning material. The future direction of this research is to provide a standardized data collection method, based on identified metrics, parameters and indicators, and build an Educational Data Warehouse. Applying highlighted Data Mining tools and methods, Education Providers and Managers will be able to predict students' academic performances. This will assist in adequate placement in tertiary education courses, enable proper guidance of students in achieving their academic goals and objectives, and also maximize utility in the overall learning experience for both the learners and the learning providers and managers. The work can be further extended out by designing the student model analysing records of students extracurricular skills and provide a suggestions on communication and technical skill development by which students can be build in professional aspect of talents.

REFERENCES

- [1] S.-T. Wu, "Knowledge discovery using pattern taxonomy model in text mining," 2007.
- [2] N. S. Raghavan, "Data mining in e-commerce: A survey," *Sadhana*, vol. 30, no. 2-3, pp. 275–289, 2005
- [3] R.Jindal, M.D Borah, A Survey on Educational Data Mining and Research trends, *International Journal of Database Management System (IJDMS)*, 5(3), 2013, 53–73.
- [4] Baker, R. S., A. T. Corbett, K. R. Koedinger, and A.Z. Wagner. 2004. "Off-Task Behavior in the Cognitive Tutor Classroom: When Students 'Game the System'." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. New York, NY: Association for Computing Machinery, 383–390
- [5] *International Journal of Scientific Engineering and Applied Science (IJSEAS)* –Volume-2, Issue-6, June 2016 ISSN: 2395-3470 www.ijseas.com
- [6] Hand, D. J. 1981 *Discrimination and Classification*. Chichester, U.K.: John Wiley and Sons
- [7] Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P; "From Data Mining to Knowledge Discovery: An overview" in *Advances in Knowledge discovery and Data Mining*. Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P; Uthurusamy, R. MIT Press. Cambridge, Mass.. 1996 pp. 1-36
- [8] Fayyad, U.; Piatetsky-Shapiro, G; Smyth, P.; "The KKD Process for Extracting Useful Knowledge from Volumes of Data" in *Communications of the ACM*, November 1996/Vol 39, No.11 pp.27-34
- [9] [http://www.intechopen.com/books/theory and applications for advanced text mning.](http://www.intechopen.com/books/theory_and_applications_for_advanced_text_mining)
- [10] https://courses.cs.washington.edu/courses/csep521/07wi/iprj/leonardo_fabricio.pdf