

A Review of Applications of Data Mining in the Field of Education

Hardeep Kaur

Master of Technology, Department of CSE, GNDU RC, Jalandhar, India

Abstract: Data mining is the process of analysis of large amount of data to find out useful patterns and rules. Data mining is the process of taking information from a data set and convert it into an understandable and meaningful structure for further use. There are various techniques of data mining like classification, clustering, association rule mining etc. Each technique has its own importance according to his role. There are various applications of data mining in various fields like education, scientific and engineering, healthcare, business and many more. In this paper we will discuss basics of educational data mining. In this paper we will mainly focus on the applications of data mining in the field of education.

Keywords: Data mining, Data Mining Elements, Educational Data Mining (EDM), Applications in education.

I. INTRODUCTION

A. Data mining

Data mining is a process of taking out useful information and patterns from large amount of data. It is also called knowledge discovery process, knowledge/pattern analysis or knowledge mining from data. [1] The main goal of this analysis process is to take information from a data set and convert it into an understandable and meaningful structure for further use. Data Mining is used for solving problems by analyzing data that is present in the databases. [2]

Data mining is necessary process where various intelligent techniques are applied for extraction of useful patterns. Data mining comprises five major elements:

- Extract, convert, and load transaction data into data warehouse system.
- Storage and then management of this data in a multidimensional database system.
- Provide access of this data to information technology professionals and business analysts.
- Analysis of data using application software.
- Present the data in a useful form, such as a table or graph. [2]

B. Educational Data Mining (EDM)

Educational Data Mining (EDM) is a process which is concerned with developing various techniques or methods for extracting the different types of data that come from educational settings, and use of those methods for better understanding of students. The main area of EDM is analyzing student's performance. Another important field is mining enrolment data. Main uses of EDM [10] include student performance prediction and studying students learning to suggest improvements in current educational practice. EDM can be taken as one of the learning sciences and as a field of data mining. [4][8][10]

C. Data mining applications

There are various applications of data mining in the area of healthcare, telecommunication, market based analysis,

financial services, manufacturing, banking, retail, food industries, road traffic, motor industries, educational field, scientific and engineering and business and many more other.[3]

This paper is organized as follows: In section II we present applications of data mining in education field. In section III we present work related to these applications. In section IV we discuss open issues and future scope of these applications. Conclusion is presented in section V.

II. APPLICATIONS OF DATA MINING IN EDUCATION FIELD

Some applications of data mining in education sector are given below:

D. Analysis and Visualization of Data

It is used to highlight meaningful information and support decision making. In the educational sector, for example, it can be helpful for course administrators and educators for analyzing the usage information and students' activities during course to get a brief idea of a student's learning. Visualization information and statics are the two main methods that have been used for this task. Statistical analysis of educational data can give us information like where students enter and exit, the most important pages students browse, how many number of downloads of e-learning resources, how many number of different type of pages browsed and total amount of time for browsing of these different pages.

It also provides information about reports on monthly and weekly user trends, usage summaries, how much material students will study and the series in which they study topics, patterns of studying activity, timing and sequencing of activities. Visualization uses graphical methods to help people in understanding and analyzing data. There are number of studies related to visualization of different educational data such as patterns of hourly, daily, seasonal and annual user behavior on online forums. [4]

E. Predicting Student Performance

In student performance prediction, we predict the unknown value of a variable that defines the student. In educational sector, the mostly predicted values are student's performance, their marks, knowledge or score. Classification technique is used to combine individual items based upon quantitative traits or based upon training set of previously labeled items. Student's performance prediction is very popular application of DM in education sector. Different techniques and models are applied for prediction of student's performance like decision trees, neural networks, rule based systems, Bayesian networks etc. This analysis is helpful for someone in predicting student's performance i.e. prediction about student's success in a course and prediction about student's final grade on the basis of features taken from logged data.

Several regression techniques are used for prediction of student's marks such as linear regression to predict student's academic performance, stepwise linear regression to predict time spent by a student on a learning page and multiple linear regression for identification of variables that are helpful for predicting success in colleges courses and for prediction of exam results in distance education courses. [5][6][7][10][11]

F. Enrolment Management

Enrolment management is frequently used in higher education to explain well-planned strategies and ways to shape the enrolment of college to meet planned goals. It is an organizational concept and also a systematic set of activities designed to allow educational institutions to exert more influence over student's enrolments. Such practices often include retention programs, marketing, financial aid awarding and admission policies. [7]

G. Grouping Students

In this case groups of students are created according to their customized features, personal characteristics, etc. These clusters/groups of students can be used by the instructor/developer to build a personalized learning system which can promote effective group learning. The DM techniques used in this task are classification and clustering. Different clustering algorithms that are used to group students are hierarchical agglomerative clustering, K-means and model-based clustering.

A clustering algorithm is based on large generalized sequences which help to find groups of students with similar learning characteristics like hierarchical clustering algorithm which are used in intelligent e-learning systems to group students according to their individual learning style preferences discriminating features and external profiling features. [4]

H. Predicting Students' Profiling

Data mining can help management to identify the demographic, geographic and psychographic characteristics of students based on information provided by the students at the time of admission. Neural networking technique can be used to identify different types of students. [7]

I. Planning and scheduling

Planning and scheduling is used to enhance the traditional educational process by planning future courses, course scheduling, planning resource allocation which helps in the admission and counseling processes, developing curriculum, etc. Different DM techniques used for this task are classification, categorization, estimation, and visualization. Decision trees, link analysis and decision forests have been used in course planning to analyze enrollee's course preferences and course completion rates in extension education courses. Educational training courses have been planned through the use of cluster analysis, decision trees, and back-propagation neural networks in order to find the correlation between the course classifications of educational training. Decision trees and Bayesian models have been proposed to help management institutes to explore the probable effects of changes in recruitments, admissions and courses. [4]

J. User Modelling

User modeling encompasses what a learner knows, what the user experience is like, what a learner's behavior and motivation are, and how satisfied users are with online learning. EDM can be applied in modeling user knowledge, user behavior and user experience. [9]

K. Organization of Syllabus

Presently, organization of syllabi is influenced by many factors such as affiliated, competing or collaborating programs of universities, availability of lecturers, expert judgments and experience. This method of organization may not necessarily facilitate students' learning capacity optimally.

Exploration of subjects and their relationships can directly assist in better organization of syllabi and provide insights to existing curricula of educational programs. One of the applications of data mining is to identify related subjects in syllabi of educational programs in a large educational institute. [11]

L. Detecting Cheating in Online Examination

Now a day's exams are conducted online remotely through the Internet and if a fraud occurs then one of the basic problems to solve is to know: who is there? Cheating is not only done by students but the recent scandals in business and journalism show that it has become a common practice. Data mining techniques can propose models which can help organizations to detect and to prevent cheats in online assessments. The models generated use data comprising of different student's personalities, stress situations generated by online assessments, and common practices used by students to cheat to obtain a better grade on these exams. [11]

III. LITERATURE SURVEY

Naeimeh DELAVARI, Somnuk PHON-AMNUAISUK [5] (2008) has discussed various application of data mining in education. They have studied important application i.e. predicting student's performance that helps institutions to

predict about student's final grade and his success in course.

Dr. Varun Kumar, Anupama Chadha [11] (2011) studied various applications of data mining in education like Organization of Syllabus, Detecting Cheating in Online Examination and student performance analysis. They studied various techniques that are used in these applications.

Monika Goyal and Rajan Vohra [4] (2012) have studied various applications of data mining like analysis and visualization of data, grouping of students, student performance analysis and planning and scheduling etc. and these can be implemented using various algorithms of classification and clustering techniques. They discussed that DM techniques helps educational institutions in analysing student's performance, designing course curriculum to group students and to motivate students.

Dr. Mohd Maqsood Ali [7] (2013) has discussed applications like enrolment management, predicting student's profiling and predicting student's performance in the field of education using various techniques. This technique helps in identifying students and their performance.

M.Santhi Swaroop, K.Venkat Raju [3] (2013) have given various application of data mining like data mining applications in healthcare, data mining for market basket analysis, data mining in education system and applications of data mining techniques in CRM any many other applications. They said that this review would be helpful for researchers to concentrate on the issues of data mining.

Nikita Jain, Vishal Srivastava [2] (2013) presented concept of data mining and surveyed that at present data mining is a new and important area of research and ANN itself is a very suitable for solving the problems of data mining because its characteristics of good robustness, self-organizing adaptive, parallel processing, distributed storage and high degree of fault tolerance. They found that commercial, educational and scientific applications are increasingly dependent on these methodologies.

Dina Abdulaziz Alhammadi [6] (2013) presented the implementation of student performance prediction using various algorithms (C4.5, k-means clustering, k star and naïve bayes). After implementation he has concluded that C4.5 performs better than all other algorithms. C4.5 has the highest accuracy than all other algorithms.

Mrs. M.S. Mythili, Dr. A.R.Mohamed Shanavas [10] (2014) presented implementation of student performance analysis using various classification algorithms (J48, random forest and decision tree). After implementation they found that random forest performance is best than that of other algorithms used in the implementation. Random forest algorithm has highest accuracy in student performance analysis.

IV. OPEN ISSUES AND FUTURE SCOPE

Based on the extensive literature survey done in section no. III following issues have been extracted and they are given below. As we know there are various algorithms of classification and clustering techniques that are used in education data mining. From all these algorithms there are some algorithms that have problem of accuracy and processing time. Some algorithms have very high processing time and low accuracy that is a big problem.

With the development of applications based on education there is a stronger need to develop algorithms that have higher accuracy and less processing time. To develop new algorithms there is a lot of scope in this field. A lot of work is being done in this area by researchers.

V. CONCLUSION

As given above various authors have discussed various application of data mining in the field of education like student performance analysis, cheating detection in online examination, planning and scheduling, grouping students and enrolment management etc. Some authors have given the implementation of student performance analysis using various classification and clustering algorithms (J48, decision table, random forest, naïve bayes etc) and given the best algorithm in performance on the basis of accuracy. It is hoped that survey done here in this paper will prove to be helpful to researchers working in the area of educational data mining.

ACKNOWLEDGEMENT

First of all I express my sincerest debt of gratitude to the Almighty God who always supports me in my endeavors. I would like to thank to my teachers for their encouragement and support. Then, I would like to thank to my family and my friends. I am thankful to all those who helped me in one way or the other at every stage of my work.

REFERENCES

- [1] Mrs. Bharati M. Ramageri, "Data mining techniques and applications", Indian Journal of Computer Science and Engineering Vol. 1 No. 4, 2008.
- [2] Nikita Jain, Vishal Srivastava, "Data mining techniques: A survey paper", IJRET: International Journal of Research in Engineering and Technology, Volume: 02 Issue: 11, Nov-2013.
- [3] M.Santhi Swaroop, K.Venkat Raju, "Survey of Advanced Data Mining Application", IJEA Vol.2 issue 2, 2013.
- [4] Monika Goyal and Rajan Vohra, "Applications of Data Mining in Higher Education", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012.
- [5] Naeimeh Delavari, Somnuk Phon-Amnuaisuk, "Data Mining Application in Higher Learning institutions", Informatics in Education, Vol. 7, No. 1, 2008.
- [6] Dina Abdulaziz Alhammadi, "Data Mining in Education- An Experimental Study", International Journal of Computer Applications Volume 62, No.15, January 2013.
- [7] Dr. Mohd Maqsood Ali, "Role of data mining in education sector", International Journal of Computer Science and Mobile Computing Vol. 2, Issue. 4, April 2013.
- [8] Richard A. Huebner, "A survey of educational data mining research", Research in Higher Education Journal, 2012.
- [9] S. Lakshmi Prabha, Dr.A.R.Mohamed Shanavas, "Educational data mining applications", Operations Research and Applications: An International Journal (ORAJ), Vol. 1, No. 1, August 2014.
- [10] Mrs. M.S. Mythili, Dr. A.R.Mohamed Shanavas, "An Analysis of students' performance using classification algorithms", IOSR

Journal of Computer Engineering, Volume 16, Issue 1, January 2014.

- [11] Dr. Varun Kumar, Anupama Chadha, "An Empirical Study of the Applications of Data Mining Techniques in Higher Education", International Journal of Advanced Computer Science and Applications, Vol. 2, No.3, March 2011.

BIOGRAPHY



Hardeep Kaur received B. Tech. degree in Information Technology from Punjab Technical University, Jalandhar, India, in 2013, pursuing M. Tech. (final year) in Computer Science Engineering from Guru Nanak Dev University, Amritsar, India. Her research areas include data mining and its decision tree algorithms.