

Data Mining Techniques and Their Implications on Educational Data Base (U-DISE)

Sheeraz Ahmad Peerzada¹, Tahir Ahmad Shiekh², Dr.Jitendra Seethalani³,

¹⁻³Department of Computer Science,
Sri Satya Sai University of Technology and Medical Sciences,
Shore, MP, India-466001.

ABSTRACT: These days data mining is an emerging trend, which can be used in any field to discover hidden Knowledge. The Department of School Education in India collect the information from students ,teachers as well as from institutions and store it in a National data base called unified district information system for education (U-DISE).It is very hard and time consuming to analyze the data and find the hidden knowledge or information manually from U-DISE. Adaptation of various techniques of data mining on mining the U-DISE data and discover the knowledge that can help in enhancing overall performance of students. In this paper we are going to present some data mining techniques and their implications on overall performance of students.

KEYWORDS-Data Mining, Educational Data Base (U-DISE), Educational Data Mining, Data Mining Techniques.

1 INTRODUCTION

Data Mining is referred to the process of extracting hidden and useful information from large databases. With the advent of Information Communication Technologies the various fields had lead the large volumes of data storage in various formats.

The data stored in Educational databases required proper extraction methods for extracting knowledge from these large repositories for better decision making. Knowledge discovery in data bases often called as data mining. From last 10 years in India all Schools either government, private or public store information up to 12th class regarding students, teachers and facilities in a data base called Unified District Information System for Education (U-DISE).

Researchers have realized that this data can be useful to educators, government, private and public organizations. Thus data mining techniques that enable the extraction of hidden potentially useful information from the Unified District Information System for Education (U-DISE) are used.

Some of the popular data mining techniques such as clustering, classification and association rules of data mining have been applied to education data with promising results. The main reason behind the rapid development of Educational data mining (EDM) research in recent years is due to availability enormous amount of educational data mostly present in normalized form in U-DISE data base.

Department of Education is collecting information on yearly basis in India, and last reference date is 30th of September every year. The said data is stored in a nationalized Data Base called Unified District Information System for Education (U-DISE).

Educational Data Mining is yielding good results on this U-DISE data. U-DISE software is operational all over India. The Central Government has launched U-DISE for collection statistical data, every school must fill information on U-DISE data capture format yearly-DISE data is used for decision making and annual works plan formation (AWP & B).The data is also used for various activities of SSA & RMSA .The main objectives of this paper is to use data mining techniques in unified district information system for education (U-DISE) for overall performance of students.

2.LITERATURE REVIEW

Data mining techniques that have been used to predict students' performance and also focuses on how the prediction algorithm can be used to identify the most important attributes in a student's data[1]. The data set consists of 203 records and 57 attributes. A classification based on an association rules algorithm is used to build a classifier to help evaluate the student's performance in the programming course [2].

Chew Li Sa et.al, "Student Performance Analysis System 2013 In this paper [3] they proposed a system named Student Performance Analysis System (SPAS) to keep track of student's result in a particular university. The proposed system offers student performance prediction through the rules generated via data mining technique. The data mining technique used in this project is classification, which classifies the students based on their grades.

Sheeraz Ahmad Peerzada, Zubair Ahmad and Dr. Jitendra Seethlani published a paper on IJARSE "Role of Data Mining on Educational Data Bases" where they suggested for each indicator different Data Mining algorithms may be applied and the algorithm with optimal results may be used for improving overall performance of students and look in to weak areas of students as well as institutions.[4].

Rahi Bansal, Akansha Mishra and Dr. Shailendra Narayana, "Mining of Educational Data for Analyzing students overall performance" they give application areas for Educational data mining[5].

Remero & Ventura [6] has reviewed various articles on EDM and after reviewing recommended that futuristic research on EDM.

2.1 Data Mining

Simply stated, data mining refers to the extraction of knowledge from large amounts of data. Data mining is the entire process of applying computer based methodology including new techniques for knowledge discovery from data.

2.2 Knowledge discovery

Concrete information discovered from known data or data we may not know, but supported by recorded facts.

2.3 Knowledge Prediction

Uses known data to forecast future trends, events etc. Prediction is very hard, especially when it is about the future.

3 Educational Data Mining (EDM): is a sub-domain of Data Mining which deals with data from academic databases which is used to develop various techniques and to recognize patterns that are unique [7]. The obtained knowledge can then be used to offer suggestions to the academic planners in higher education institutes to enhance their decision making process, to improve students' academic performance, to decrease failure rates, to understand students' behavior in a better way, to assist instructors, to improve teaching, and to construct regression models and decision trees to predict over all student performance on the bases of U-DISE data in terms of their results, enrolment, facilities etc.

3.1 GOALS FOR EDUCATIONAL DATA MINING IN EDUCATIONAL FIELD

Baker and Yacef[7] describes the following four goals of EDM:

3.1.1 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.

3.1.2 Discovering or improving domain models - Through the various methods and applications of EDM, discovery of new and improvements to existing models are possible.

3.1.3 Studying the effects of educational support - It can be achieved through learning systems.

3.1.4 Advancing scientific knowledge about learning and learners - By building and incorporating student models, the field of EDM research and the technology and software used [8].

4 TECHNIQUES OF DATA MINING AND THEIR USE IN EDUCATIONAL DATA BASE (U-DISE)

The various techniques of data mining and for what purpose they are used in educational database called unified district information system for Education (U-DISE) are given below.

4.1 Classification

Classification is the most commonly used data mining technique which employs on builds a model using a training dataset with known classes of data. That model is used to classify new, unknown data in to those classes. The classification techniques are as K-Nearest Neighbor and decision trees.

4.2 Association Rule

Association and correlation is usually to find frequent item set finding among large data sets .Association rules are used to show the relationship between data items. Purchasing one product when another product is purchased.

4.3 Clustering

Clustering can be said as identification of similar classes of objects. Clustering is similar to classification in that data are grouped unlike classification, the groups are not predefined, they are discovered. Grouping accomplished by finding similarities between data according to charistics found in the data. It has two types K-Means clustering and Nural Networks.

4.4 Predication

Regression techniques can be adopted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict.

5APPLICATIONS OF DATA MINING ON EDUCATIONAL DATABASE (U-DISE)

There are many application areas of data mining like customer analytics, business analysis, Agriculture, banking sector, Security Applications, Educational data mining, Mass surveillance, Privacy preserving techniques etc. The main concerned area is about data mining applications in educational systems. Educational Data Mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational institutions and stored in a educational database called U-DISE, and using the methods to better understand students, and the settings which they learn in. A key area of EDM is mining student's performance. Another key area is mining enrollment data. Key uses of EDM include predicting student performance and studying learning in order to recommend improvements to current educational practice and the facilities in these educational institutions.

The main applications of EDM are listed as follows:

5.1Analysis and Visualization of U-DISE Data

It is used to highlight useful information and support decision making process. In the educational environment, for example, it can help educators and course administrators to analyze the students' course activities and usage information stored in U-DISE to get a general view of a student's learning. Statistics and visualization information are the two main techniques that have been most widely used for analysis and visualization. Statistics is a mathematical science concerning the collection, analysis, interpretation or explanation, and presentation of U-DISE data. It is relatively easy to get basic descriptive statistics from statistical software, such as SPSS. Statistical analysis of educational data (logsfiles/databases) can tell us things such as where students enter and exit. Statistical analysis is also very useful to obtain reports assessing how many minutes student worked, number of problems here solved and his correct percentage along with our prediction about his score and performance level.

Student tracking data to analyze student's attendance, results, teacher working hours and student working hour activities.

5.2 Predicting Student Performance on the bases of U-DISE data

In this case, we estimate the unknown value of a variable that describes the student. In education, the values normally predicted are student's performance, their knowledge, score, working hours or results. This value can be numerical/continuous (regression task) or categorical/discrete (classification task). Regression analysis is used to find relation between a dependent variable and one or more independent variables. Classification is used to group individual items based upon quantitative characteristics inherent in the items or on training set of previously labeled items. Prediction of a student's performance is the most popular applications of DM in education. Different techniques and models are applied like neural networks, Bayesian networks, rule based systems, regression, and correlation analysis to analyze educational data. This analysis helps us to predict student's performance i.e. to predict about his success in a course and to predict about his final grade based on features extracted from logged data. Several regression techniques are used to predict student's marks like linear regression for predicting student's academic performance, stepwise linear regression for predicting time to be spent on a learning page, multiple linear regressions for identifying variables that could predict success of students.

5.3 Grouping Students on the basis of different indicators in U-DISE

In this case groups of students are created according to their customized features, personal characteristics, and U-DISE indicator resemblance. 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.

5.4 Enrollment Management

This term is frequently used in education to describe well-planned strategies and tactics to shape the enrollment of an institution and meet established goals. Enrollment management is an organizational concept and a systematic set of activities designed to enable educational institutions to exert more influence over their student enrollments. Such practices often include marketing, admission policies, retention programs, and financial aid awarding. Strategies and tactics are informed by collection, analysis, and use of data to project successful outcomes.

Activities that produce measurable improvements in yields are continued and/or expanded, while those activities that do not provide measurable results are discontinued or restructured.

"Enrollment management" has increased in recent years only in private institutions while as in government the enrollment is decreasing day by day. The government needs to change its policies in order to improve education. These offices serve to provide direction and coordination of efforts of multiple offices such as admissions, financial aid, registration, and other student services. Often these offices are part of an enrollment management division. Some of the typical aims of enrollment management include improving yields at inquiry, application, and enrollment stages.

6. CONCLUSION

The increased use of Technology in Education leads arrival of big data, the data from students, teachers as well as from educational institutions. This paper discusses about Data mining, educational data mining, its techniques and applications to be used in educational data base called unified district information system for education (U-DISE).

This study will help Government, private and public organizations to improve overall performance of students.

Acknowledgements: we are thankful to our research guide Dr.Jitendra Seethlani for his suggestions, considerations and his guidelines which he provided us day to day. His responsive advices help us to work better and quality improvement.

7. REFERENCES

- [1] Amirah Mohamed Shahiria, Wahidah Husaina, Nur'aini Abdul Rashida,"A Review on Predicting Student's Performance using Data Mining Techniques", Science Direct,pp. 414 – 422, 2015.
- [2]Asmaa Elbadrawy, Agoritsa Polyzou, Zhiyun Ren, Mackenzie Sweeney, George Karypis, Huzefa Rangwala,"Predicting Student Performance Using Personalized Analytics", IEEE ,pp. 61-69, April 2016.
- [3] Chew Li Sa; Bt Abang Ibrahim, D.H.; Dahlia Hossain, E.; bin Hossin, M., "Student performance analysis system (SPAS)," in *Information and Communication Technology for The Muslim World (ICT4M), 2014 The 5th International Conference* on , vol., no., pp.1-6, 17-18 Nov. 2014.
- [4] Sheeraz ,Zubair Dr. Jitendra Seethlani published a paper on "Role of Data Mining on Educational Data Bases", IJARSE Journal 2017.
- [5] Rashi Bansal, Akansha Mishra and Dr. Shalindra Narayana "Mining of Educational data for Analysing Students Overall Performance "In IEEE Journal 2017.
- [6] Romero, C. and Ventura, S "Educational Data Mining: A Survey from 1995 to 2005", Expert system with Applications (33), 2007
- [7] <http://wikipedia.org>.
- [8] Baker, R.S.; Yacef, K (2009). "The state of educational data mining in 2009: A review and future visions". JEDM-Journal of Educational Data Mining (1): 2017.