Review on Application of Data Mining Educational Big Data

Rishiram
M.Tech Scholar
Department of CSE
Vaishnavi Group of Institutions
Bhopal (M.P), India
rishirammaskare8@gmail.com

Sumit Sharma
Professor
Department of CSE,
Vaishnavi Group of Institutions
Bhopal (M.P), India
sumit_sharma782022@yahoo.co.in

Abstract: **—In recent years, research on Educational Data Mining (EDM) has developed rapidly. However, most researches focus on data source issues, and ignore the importance of data preprocessing and data mining algorithms. This paper has studied EDM, with a special focus on educational big data mining algorithms. Firstly, it analyzed the relevant elements of EDM and introduces big data technology based on the requirements of educational data application. Then it introduced the common educational big data mining algorithms and their applications, and finally discussed the development trend of educational big data mining algorithms.**

Keywords: Data mining, Students' learning, Big Data, Educational Big Data.

I. INTRODUCTION

Data plays an extremely important role in the process of education reform and development. Education data covers a wide range, mainly from two sources: the generation of teaching activities and the collection of education management processes. These data are specially collected, classified, sorted, counted and analyzed to become educational big data. In recent years, online education and online courses have continued to increase, and big data in the field of education has also gained a wider application space. It is necessary to pay special attention to how the collected educational data is properly implemented. Processing is very important for their future availability [1]. Analyzing and processing data through data mining techniques has become one of the current research hotspots. Educational Data Mining (EDM) is a technical method that comprehensively uses mathematical statistics, machine learning, and data mining. First of all, the data generated in education are analyzed and processed properly, then the relationship between learners' learning results and variables such as learning content, learning resources and teaching behavior is found by modeling the data, and finally, the future learning trends of learners are predicted to improve and guide educational work [2]. EDM is of great significance to the sustainable development of intelligent education, the reform of teaching mode and the reconstruction of educational evaluation system.

The recent development trend of education data is to allow researchers to accumulate a large amount of unstructured data. Most of the structured education data comes from the collection and arrangement of the education sector for many years. The continuous development of education methods in the current era makes the More and more unstructured education data is emerging [3].

From an application perspective, educational data has the following characteristics.

- Educational data is hierarchical; smart terminal layer, network layer, student layer, classroom layer, teacher layer, school layer, and layer nesting.
- Educational data has time and sequence attributes, and these attributes have specific contexts. Time attributes include the time node of learning, the length of learning, etc. sequence attributes indicate how the ordering of knowledge, exercises, and teaching is performed.

II. APPLICATION OF BIG DATA IN EDUCATION

Big Data is now widely used to describe and define the recent emergence and existence of large data sets. It can be found in many areas. The public, corporate and social sectors consistently receive and produce large amounts of data from a variety of sources and in different formats. Big data and analytics have added value to data in different contexts and have therefore proved to be an extremely useful approach to their potential impact both in industry in the form of business intelligence and analysis [4], and in science with data mining techniques in education to study and learn analysis [5]. Given the limited research on the use of big data and analytics in the context of the education system, we will introduce the reader to the new field of big data about

education that puts big data into education and like data about education. education can be treated in different dimensions and from different dimensions. Perspectives to highlight various stakeholders such as policy makers, academic faculties, assessment specialists, researchers and students in computer science, engineering and computer science courses, and to promote data-driven activities to improve the quality of education.

One of the areas where volume, diversity and speed coexist in data is higher education. Large amounts of education data from various sources and in different formats are collected and generated daily in the higher education ecosystem. Education data varies from data obtained from student use and interaction with learning management systems and platforms (LMS) to learning activities and course information. which define a program such as learning objectives, programs, learning materials and activities, exam results and course assessment. include other types of data related to administrative, educational and quality improvement processes and practices. The limited use of big data on education, as well as the size and nature of such data in the context of higher education, means that specific techniques must be applied to uncover useful new knowledge currently hidden in the data. [6]. More recently, big data and analytics have shown great promise in promoting various interventions in higher education. These measures concern "administrative decision making and the allocation of organizational resources", the prevention of failure of vulnerable pupils through early diagnosis, the development of effective teaching techniques and the revision of the traditional curriculum view. rethinking the various data collected by the LMS, social networks, learning activities and curriculum and regularly created as a network of relationships and connections [8]. In particular, one of the areas identified where big data and analytics are suitably applicable to research and improvement in higher education is the curriculum and its content as an important part of big data on education [7].

Therefore, various work has been carried out towards developing an effective big data analytical approach on educational data. Figure 1 highlights the frequently addressed problems with its methodologies and limitations associated with analytical approach.

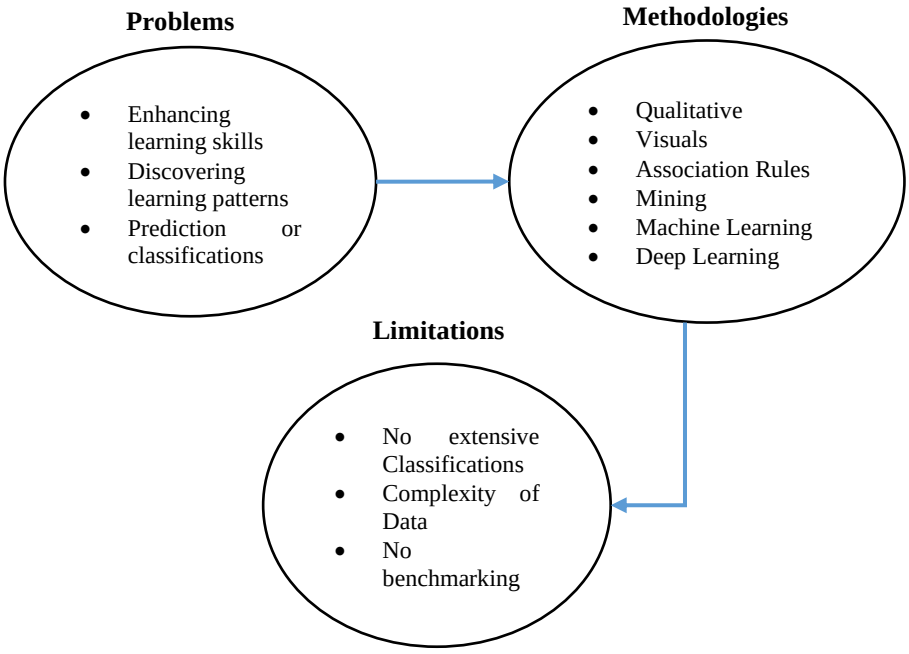

Figure 1: Observations from Existing Educational Big Data

The main focus of educational data mining and application of machine learning are described as below:

1. Predicting Future Behavior of Student: The Modeling can be used to create student Models based on learners' characteristics which may used detailed information such as their motivation and learning behavior.
2. Identify or Improving Domain Models: By applications of data mining or more precise Educational data mining used to find a new or improvements to existing models are possible.
3. Study of Educational Support: The learning system can be used to studying the effects of educational Support.
4. Scientific Knowledge of Learners: To build student models, in the field of Educational Data Mining research and the technology software and Machine Learning can be used.

5. Techniques of Machine Learning and Their Use in EBD: The two main techniques of Machine Learning supervised and unsupervised Learning and for what purpose they are used in educational database.

III. Education Big Data Analysis

Big data analysis and mining is a process of extracting, refining, and analyzing cluttered data from aspects of visualization analysis, data mining algorithms, and predictive analysis [7].

A. Visual analysis

Visual analysis refers to analysis methods that clearly and effectively communicate information by means of graphical means. In the application of educational big data, it is mainly the correlation analysis of massive data, which is the process of correlation analysis of scattered and heterogeneous data to form a complete analysis chart.

B. Data mining algorithm

Data mining algorithms, which is the methods for calculating, processing, and analyzing data by creating a data mining model. It is the theoretical core of the entire education big data analysis. There are many kinds of data mining algorithms. Different algorithms applied to different types of data will show different results. In educational big data analysis and mining, usually the first step is to analyze the original data, and then find the mining algorithm for specific types of patterns and trends according to the application requirements. The analysis results obtained through data mining can define the best parameters to create a mining model in turn, simultaneously apply them to the entire education dataset to extract feasible patterns and detailed statistics.

C. Predictive analysis

Predictive analysis is one of the most important application areas of big data analysis. It achieves the purpose of predicting uncertain events by combining multiple advanced analysis functions. Predictive analysis of educational data often uses functions such as statistical analysis, predictive modeling, data mining, text analysis, optimization, machine learning, deep learning and other methods to help users discover data evolution trends, data patterns, and data relationships presented in the original data. Use these indicators to predict and provide decision-making basis for taking corresponding measures.

IV. Related Work

Yu et al. [7] focuses on mining useful data from the massive online education data, by using transfer learning, relying on Hadoop, to construct Online education data classification framework (OEDCF), and design an algorithm Tr_MAdaBoost. This algorithm surpasses traditional classification algorithms where the required data should be limited to independent and identically distributed data, as the correct classification can be achieved through online instruction with this new algorithm, even if the distribution of the data is different. At the same time, with the help of Hadoop's parallel processing architecture, OEDCF can significantly improve the efficiency of data processing, create favorable conditions for learning analytics, and promote personalized learning. and other activities in the era of Big Data.

Li et al. [8] used of the data such as the achievement information of students in a vocational college and the consumption information of campus one-card. Firstly, the original data is preprocessed so that the processed data can meet the basic requirements of data analysis. Then, the data after preprocessing is discretized with the data discretization technology so that the data can meet the requirements of the Apriori algorithm of association rules. Finally, Apriori algorithm is used to mine the correlation between students' academic performance and campus one-card consumption data.

Santos et al. [9] proposed a computational approach using educational data mining and different supervised learning techniques (Decision Trees, K-nearest Neighbor, Neural Networks, Support Vector Machines, Naive Bayes and Random Forests) to evaluate the behaviour of different prediction models in order to identify the profile of at-risk university students in a Brazilian university environment. The results of this paper indicate that some algorithms can be used as tools for supporting decisions that reduce school dropout.

Nadabi et al. [10] used on this project to predicate student selection was J48, sample logistic and Naïve Bayes. The results of this project show that many students must choose applied mathematics because of their level in mathematics subject.

Ketui et al. [11] focused on classification models for applying in Education Data Mining. The classification models are applied to identify the suitable subject to the science students. The experiment is set to improve the student performance which comparing the performance of five classification models and then predicting the appropriated academic achievement in each major.

Shao et al. [12] proposed an optimized mining algorithm for analyzing students' learning degree based on dynamic data. The algorithm first uses the optimized text classification

technology to match the question texts to the knowledge points automatically, so as to improves the efficiency and quality. Then, it uses the subjective weighting method combined with the expert experience to generate the learning degree matrix of students on knowledge points based on dynamic data of the students' records. Finally, the DBSCAN clustering algorithm is used to cluster the personalized learning characteristics of students according to the learning degree matrix.

V. CONCLUSION

Big data technology is now the forefront of scientific and technological development, and is an indispensable technology in education data mining. This paper first introduced EDM and big data technologies, and then introduced popular data mining algorithms and their applications. Then, according to the characteristics of educational data, combined with big data mining technology, the more practical educational big data mining algorithms and their core ideas are introduced. EDM research is developing rapidly, but the focus is different, and we make an outlook on the research of EDM. At present, there are less research on data preprocessing technology and data mining algorithms in the research system of EDM, this is the most significant problem . Data preprocessing methods is as important as data mining algorithms in EDM research, and in some cases even exceed the latter, which is one of the directions for our future research.

REFERENCES

[1] Zhou Q, Mou C, Yang D. Research progress on educational data mining: A survey. Ruan Jian Xue Bao/Journal of Software, 2015, 26(11): 3026−3042 (in Chinese). http://www.jos.org.cn/1000-9825/4887.html.

[2] Sun Xuejuan. Application of data mining technology in teaching information [J]. Information communication. 2019.277-278.

[3] Amjed Abu Saa.Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques. Technology, Knowledge and Learning, 2019, Vol.24 (4), pp.567-598.

[4] Tapani Toivonen. Augmented intelligence in educational data mining. Smart Learning Environments, 2019, Vol.6 (1), pp.1-25.

[5] Shah J. Miah.Editorial note: Learning management systems and big data technologies for higher education. Education and Information Technologies: The Official Journal of the IFIP Technical Committee on Education, 2020, Vol.25 (1), pp.725-730.

[6] L. Ji, X. Zhang and L. Zhang, "Research on the Algorithm of Education Data Mining Based on Big Data," 2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI), Xinxiang, China, 2020, pp. 344-350.

[7] L. Yu, X. Wu and Y. Yang, "An Online Education Data Classification Model Based on Tr_MAdaBoost Algorithm," in Chinese Journal of Electronics, Vol. 28, no. 1, pp. 21-28, 1 2019.

[8] Z. Li, "New Employee Student Repast Big Data Analysis Research Application," 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Vientiane, Laos, 2020, pp. 583-586.

[9] K. J. de O. Santos, A. G. Menezes, A. B. de Carvalho and C. A. E. Montesco, "Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout," 2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT), Macei??, Brazil, 2019, pp. 207-208.

[10] S. S. Al-Nadabi and C. Jayakumari, "Predict the selection of mathematics subject for 11th grade students using Data Mining technique," 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), Muscat, Oman, 2019, pp. 1-4.

[11] N. Ketui, W. Wisomka and K. Homjun, "Using Classification Data Mining Techniques for Students Performance Prediction," 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON), Nan, Thailand, 2019, pp. 359-363.

[12] Z. Shao, H. Sun, X. Wang and Z. Sun, "An Optimized Mining Algorithm for Analyzing Students' Learning Degree Based on Dynamic Data," in IEEE Access, Vol. 8, pp. 113543-113556, 2020.