

# *Research on the Application of Data Mining in Educational Big Data*

Qianqian Wang<sup>a</sup>, Weizhen Zhang<sup>b</sup>

Henan Vocational University of Science and Technology, Zhoukou, 466000, China

<sup>a</sup>wqq131111016@163.com, <sup>b</sup>wizen\_zhang@163.com

**Keywords:** Education Big Data; Data Mining; Student Behavior Analysis

**Abstract:** The rise of educational big data has provided ample prospects for the application of data mining technology in the field of education. This article begins with the characteristics of educational big data, exploring the current status and challenges of data mining technology in education. Through analysis of practical applications of data mining in areas such as student behavior analysis, teaching quality evaluation, and personalized learning, the article expounds how data mining technology can support educational decision-making, improve teaching effectiveness and learning experience. Additionally, it discusses data privacy protection and ethical issues, as well as the future development direction of data mining in the field of educational big data, providing new ideas for realizing intelligent and personalized education.

## 1. Introduction

With the rapid development of information technology, the field of education has gradually entered the era of big data, which covers information such as students' learning behavior, teachers' teaching activities, and educational management, with characteristics such as diversity, dynamism, and real-time. How to use these massive data to improve education quality and optimize teaching processes has become a hot topic in current research. As an important tool for extracting valuable information from big data, data mining technology has shown great potential in the field of education. This article will delve into the application of data mining technology in educational big data, analyzing practical cases and effects in student behavior analysis, teaching quality evaluation, and personalized learning.

## 2. Characteristics and challenges of big data in education

### 2.1 Diversity and heterogeneity of educational big data

Education big data has significant diversity and heterogeneity characteristics, education data sources, including students' personal information, learning behavior records, teaching activity data, education management data, etc., these data types, format, both structured data such as student performance, attendance records, etc., also have unstructured data such as teaching video, study notes, discuss BBS content, etc. In addition, different education institutions and platform collected

data standards and specifications, leading to data heterogeneity, data diversity and heterogeneity to data integration and analysis brought challenges, need to adopt appropriate data preprocessing and integration technology to different sources and standardization of data in different formats in order to facilitate the subsequent data mining and analysis work[1].

## **2.2 Dynamic and real-time nature of educational big data**

Education data as the teaching activities and learning process and constantly generated and update, embodies the dynamic education process, such as students 'learning behavior data, teachers' teaching interaction data, the use of education resources data are changing over time of dynamic data, these data reflects the various participants in the education process of the real-time state and behavior characteristics. Education decision-making and interventions often need to be based on real-time data analysis and implementation, data mining technology can quickly process dynamic data, provide real-time analysis results and feedback, the efficiency and scalability of data mining algorithm put forward higher requirements, need to use incremental learning, online learning technology, realize real-time data processing and analysis.

## **2.3 Challenges of big data analysis in education**

Education big data analysis is faced with many challenges. First of all, the privacy protection and security of education data have attracted much attention. Education data contains a large number of students' personal privacy information, so how to ensure the security and privacy protection of data in the process of data mining and analysis is an important challenge. The quality of education data problems also affects the effect of data mining, education data noise, missing, inconsistent quality problems, need to data cleaning and pretreatment to improve the accuracy and integrity of the data, the final education big data analysis also need to consider ethics and fairness, avoid prejudice and discrimination from the results of data mining, to ensure the fair distribution and use of education resources.

# **3. Application of data mining in student behavior analysis**

## **3.1 Learning behavior mode mining**

Data mining technology can be applied to the mining and analysis of students learning behavior pattern, by collecting the students 'learning behavior data such as learning time, learning resources access records, homework completion, classroom interaction, etc., using data mining algorithm can find the students' learning behavior patterns and rules, such as using cluster analysis can be divided into different learning behavior groups, identify positive learning attitude, high learning efficiency and negative learning attitude, students with low learning efficiency. Through association rules mining can find the correlation between students 'learning behavior, such as learning time and learning effect, the use of learning resources and learning performance, etc., these learning behavior pattern mining helps to understand the students' learning characteristics and needs, provide the basis for developing personalized teaching strategies and interventions.

## **3.2 Learning risk early warning and intervention**

Data mining technology in students learning risk early warning and intervention has important application, through the analysis of students' learning behavior data combined with academic performance, attendance information can build learning risk prediction model, the model can timely

find learning difficulties, learning progress lag, learning interest decline learning risk signal, early warning students may face learning problems. Prediction model can be based on classification, regression supervision learning algorithm, using historical data training, when the new student data input, the model can predict the students' learning risk and risk level and possible causes, teachers can be based on the early warning results timely targeted interventions such as individual tutoring, learning resources recommendation, learning plan adjustment, etc., to help students overcome learning difficulties, improve the learning effect.

### **3.3 Evaluation and feedback of learning effect**

Data mining technology can also be applied to the student learning effect evaluation and feedback, through the analysis of students' learning behavior data, homework completion, classroom performance and other multidimensional data, using education data mining algorithm can establish a learning effect evaluation model, the model considering the indicators, gives the students' comprehensive learning effect score and identify students in different knowledge points and skills. The evaluation results can be presented to students and teachers as feedback information, and students can adjust their learning strategies according to the feedback results to fill the gaps; teachers can optimize the contents and methods of teaching according to the feedback results according to their aptitude, and the learning effect supported by data mining helps to realize the precise and personalized teaching and improve the teaching quality and learning experience.

## **4. Application of data mining in teaching quality evaluation**

### **4.1 Teaching process data mining**

Data mining technology can be applied to the teaching process data mining and analysis to evaluate the teaching quality, teaching process data including teachers' teaching behavior data, students' classroom participation data, teaching interaction data, etc., by collecting and analyzing these data can reveal the characteristics and rules of the teaching process, such as using sequence pattern mining can find teachers' teaching behavior pattern, identify the effective teaching strategies and methods, through the social network analysis can depict the interaction between teachers and students, evaluate the effect of teaching interaction. We can also apply text mining technology analysis in the process of teaching text data such as teaching discussion record, student feedback, mining the theme and emotional tendency, evaluate the pertinence of teaching content and the acceptance of students, teaching process data mining provides data support for teaching quality assessment, help to fully evaluate the teaching process, found that teaching advantages and disadvantages, provide the basis for teaching improvement.

### **4.2 Recommendation and optimization of teaching resources**

Data mining technology is widely used in teaching resources recommendation and optimization, through the analysis of students' learning behavior data, learning preferences combined with metadata teaching resources can establish personalized teaching resources recommendation model, the model using collaborative filtering, content recommended technology, according to the students' learning characteristics and demand recommend high matching degree of teaching resources such as teaching materials, courses, exercises, etc., personalized teaching resources recommendation can improve students' learning interest and efficiency, implementation according to their aptitude. Data mining can also be applied to the optimization of teaching resources. High quality teaching resources and teaching resources to be improved are identified by analyzing the usage of teaching

resources and students' feedback and evaluation with high frequency and positive evaluation can be revised and updated to provide support for the improvement of teaching quality.

### **4.3 Evaluation and improvement of teaching effect**

Data mining techniques play an important role in teaching effectiveness evaluation and improvement, as data mining methods can provide a more comprehensive and objective evaluation perspective. By collecting and analyzing multiple dimensions of data, such as students' learning behavior data, homework completion status, and classroom performance, applying educational data mining algorithms can establish an evaluation model for teaching effectiveness. This model comprehensively considers various indicators to quantitatively evaluate the teaching effectiveness of teachers, identify strengths and weaknesses of teaching, and the evaluation results can be presented as feedback information to teachers and teaching managers. Teachers can use the evaluation results to reflect on their teaching practices, improve teaching methods and strategies. Teaching managers can use the evaluation results to optimize course settings, resource allocation, and other measures to promote teaching quality improvement[2].

## **5. Application of data mining in personalized learning**

### **5.1 Learner feature analysis and modeling**

Data mining technology can be applied to learners characteristic analysis and modeling to lay the foundation for realizing personalized learning, through collecting the personal information of learners, learning behavior data, learning preferences, etc., using data mining algorithm can fully depict the characteristics of learners, such as using clustering analysis can divide learners into different groups, such as learning style partial visual, learning style auditory type, reveals the learners in learning differences, through the associated rules mining can discover the correlation between learners knowledge, identify the strengths and weaknesses of learners. We can also apply deep learning technology to build learners knowledge model, cognitive model, etc., dynamic tracking the learning status and ability level, learners characteristic analysis and modeling provides data basis for personalized learning, help to understand the individual differences of learners, provide support for subsequent personalized learning services.

### **5.2 Personalized learning path recommendation**

Data mining technology based on the results of learners characteristics analysis and modeling can be further applied to personalized learning path, personalized learning path according to the personal characteristics and needs, for its customized learning content, learning activities and learning schedule, through the mining learners learning behavior pattern, knowledge mastery data combined with the dependence of course knowledge points and difficulty information can build personalized learning path recommendation model. The model comprehensively considering the learning characteristics of learners, the current knowledge level and learning objectives, using sequence pattern mining, figure mining technology generation match the personalized needs of learning path, including learning content selection, the arrangement of learning activities and learning schedule control, aims to optimize the learners learning experience and learning effect, personalized learning path recommended can help learners efficient and targeted learning, improve learning efficiency and learning motivation.

### **5.3 Generation of adaptive learning resources**

Data mining technology can also be applied to the adaptive generation of learning resources to support personalized learning, adaptive learning resources refers to the personal characteristics and demand of learners dynamic generated personalized learning materials and learning activities, through the analysis of learners' learning behavior data, learning style, knowledge, etc., the use of natural language processing, knowledge graph technology can realize automatic generation and personalized adjustment of learning resources. For example, visual learning materials or interactive learning activities; learning tasks and exercises. The generation of adaptive learning resources considers the individual differences of learners, provide learning content that match the needs of learners 'ability and needs, and promote learners' understanding and master. The adaptive learning resource generation can realize personalized and dynamic adjustment of teaching content, provide tailored learning experience for learners, and improve the learning effect and learning satisfaction[3].

## **6. Data privacy protection and ethical issues**

### **6.1 Education the necessity of big data privacy protection**

Education big data contains a large number of students 'personal privacy information, such as students' basic information, learning behavior records, performance data, these data once leaked or abuse may cause serious impact on students' personal privacy and future development, so in the education big data collection, storage, analysis and application process must attach great importance to data privacy protection. First we need to develop strict data privacy protection policy and specification clear data collection, use and sharing scope, to ensure that the data compliance use, followed by technical and management measures such as data encryption, access control, data desensitization to prevent data leakage and unauthorized access, in addition to strengthen the privacy awareness of data users education, improve the sense of responsibility and moral level, only on the premise of ensuring students' privacy security to give full play to the value of education big data, promote the fairness and quality of education[4].

### **6.2 Data desensitization and security access control**

Data desensitization and security access control is an important technical means to protect education big data privacy, data desensitization refers to the reserved data analysis and mining value at the same time of sensitive data conversion or replacement, no longer sensitive or personal identity characteristics, common data desensitization methods including data encryption, data confusion, data deletion, through data desensitization can protect students' privacy to some extent, at the same time can support data analysis and mining work. Security access control refers to the education big data access and use of strict authority management and control, through identity authentication, access authorization, data encryption transmission measures to ensure that only after authorized users to access and use the corresponding data, at the same time to data access and operation monitoring and audit, timely detection and handling of abnormal behavior, data desensitization and security access control application can from two aspects of data source and access link, comprehensive security education privacy of big data[5].

### **6.3 Ethical considerations in the data mining process**

During the process of education big data mining, it is crucial to consider ethical issues. Although

data mining can provide valuable insights for educational decision-making and practice, it can also bring ethical risks such as violating student privacy, data misuse, and outcome discrimination. Therefore, it is necessary to follow ethical principles during the data mining process to ensure the reasonable and legal use of data. Firstly, it is important to respect students' rights to information and choice by informing them of the purposes, methods, and scope of data collection and usage and obtaining their explicit consent. Secondly, it is necessary to ensure fairness and non-discrimination in the data mining process, avoiding biases and discrimination based on sensitive attributes such as race or gender. And ethical review of the data mining results, we need to evaluate the influence on students' individual and groups, prevent improper decisions and action, finally to strengthen the data mining practitioners of ethics education and regulation, improve the ethical consciousness and sense of responsibility, only on the basis of the ethical principles of education data mining work can really play the value of the data, promote the healthy development of education.

## 7. Conclusion

Data mining technology has achieved preliminary results in the application of education big data, providing powerful support for educational decision-making, promoting intelligent and personalized development of education, however, the characteristics and challenges of education big data also pose new requirements for data mining technology. In the future, it is necessary to further explore the deep integration of data mining technology and the education field, strengthen research on data privacy protection and ethical issues, and promote the application of data mining in education big data to maturity, providing powerful support for the sustainable development of education.

## References

- [1] Wang Qian. *Exploration on the application of educational big data mining based on deep learning technology [J]. Digital Communication World, 2023, (11): 112-114.*
- [2] Meng Weidong. *The application of data mining in Big Data in Education [J]. Public Relations World, 2021, (20): 139-140.*
- [3] Wang Weilin. *Big data analysis and decision-making assistance platform for basic education [D]. Shandong University, 2021.*
- [4] Zheng Yafeng, Zhao Yaning, Bai Xue, et al. *Review of big data visualization studies in education [J]. Computer Science and Exploration, 2021, 15 (03): 403-422.*
- [5] Le Jieyu, Luo Chaoyang, Ding Jingshu, et al. *Research on the privacy protection mechanism and technology of big data in education [J]. Big Data, 2020, 6 (06): 52-63.*