Research on Large Data Mining for Online Education of Mobile Terminal Based on Block Chain Technology

Zhitao Yu(✉) and Wei Zou

Jiangxi Software Vocational and Technical University, Nanchang 330041, China
yuzhitao1210@163.com

Abstract. In order to manage the massive online instructional resources effectively and realize the goal of quickly mining massive instructional resources, this paper proposes a research method of online instructional large data mining based on block chain technology. Based on the block chain technology, this paper constructs the recognition model of online education of mobile terminal, optimizes the management system of big data of online education of mobile terminal. Experimental results show that the block chain-based mobile terminal online education large data mining method has high practicability in the practical application, and fully meets the research requirements.

Keywords: Blockchain technology · Mobile terminal · Online education · Data mining

1 Introduction

With the development of the Internet, people begin to spread and learn content quickly and receive online education through information technology and Internet technology. Currently, building a high level of interaction is an important way to promote online education [1]. At present, the interactive analysis system of online education does not involve the analysis of unstructured data, which is the key factor that restricts the development of online education. Take this as a guide to improve online education services and achieve a two-way balance between online education services and learners' needs [2].

Reference [3] proposes the application and practice of Apriori based data mining algorithm. First, it analyzes the problems existing in mathematics teaching activities, and then analyzes the construction of information-based teaching model with Apriori algorithm in big data as the main idea, data mining is completed. This method realizes the fast mining of educational resources, but the mining accuracy is low. Reference [4] proposes the research of terminal online education data mining technology based on model driven. Using association analysis data conversion method to convert data, and then using model driven crowd behavior modeling method, the task flow of terminal

W. Fu and L. Yun (Eds.): ADHIP 2022, LNICST 468, pp. 286–299, 2023.
https://doi.org/10.1007/978-3-031-28787-9_22

online education data mining is designed After completing the above work, the key technology of model driven data mining is optimized through screening, selecting data subsets, coding, setting thresholds, and evolution steps to achieve efficient data mining for terminal online education, but the mining recall rate is low.

To this end, the research on mobile terminal online education big data mining based on blockchain technology is proposed. According to the characteristics of educational analytics and educational data mining, a scoring model of curriculum review is constructed to objectively evaluate students' learning ability. Process teaching data through blockchain technology and extract important educational interaction information. The knowledge-based information recommendation algorithm completes data mining and improves the accuracy of data mining.

2 Mobile Terminal Online Education Big Data Mining

2.1 Identification of Online Education Characteristics of Mobile Terminals

According to the different research perspectives and emphases, the research of educational data analysis for online learning platform mainly includes two research fields: educational data mining and educational analysis. These two areas have a lot in common, but also have their own key concerns. Educational data mining focuses more on analytical methods for online educational data, while educational analysis focuses more on analysis of learning patterns based on online data to help the development of pedagogical research. Through modeling and analyzing various data variables and their relationships, educational data mining hopes to explore the factors that affect learning goals, learning interests and learning effects. The main goal of educational data mining is to improve learning software, online learning sites and learning models, while educational analysis focuses more on instructing teachers and students to improve teaching and learning methods [5]. Both of them make use of pedagogy theory and data mining analysis technology synthetically. There are many common points in technical means, but the emphasis is different from the final goal. They are both the new cross research fields of pedagogy and computer science. The characteristics of educational analytics and educational data mining are shown in Table 1.

The user's comments on the course can reflect the user's preference for the course to some extent. Through collecting the comments information of users and courses, using a text classification model to score course comments, the preference matrix of users is constructed, and on this basis, the recommendation algorithm is used to recommend personalized courses to users. This paper first describes the comment scoring model, then introduces the recommendation algorithm, and finally explains the effect of the algorithm on the dataset. The value range of course review features is shown below.

$$y_i = \begin{cases} 1, \text{ if } P_i > \Theta \\ 0, \text{ other} \end{cases} \tag{1}$$

In formula (1), P_i is the possible probability of positive and negative evaluation, Θ represents the threshold value of the classification model, which is 0.45 during the experiment. Based on the definition of the problem, a specific neural network model is

Table 1. Characteristics of educational analysis and educational data mining

	Educational analysis	Educational data mining
Research means	Manual analysis is the main method and automatic mining method is the auxiliary tool	Automatic mining method is the main method, and manual analysis is the auxiliary tool
Research method	Pay more attention to the law analysis as a whole	Decompose the system and study the influence of various factors and the relationship between factors
Origin	Educational output prediction and system analysis in "intelligent course" on Semantic Web	Student modeling in learning software, such as course output prediction and other problems
Educational improvement	Focus on guiding educators and learners to improve	Focus on the automatic improvement of learning software and learning model
Main technology	Social network analysis, emotion analysis, influence analysis, discourse analysis, learner effect prediction, etc.	Classification, clustering, Bayesian model, relationship mining, pattern mining, visualization, etc.

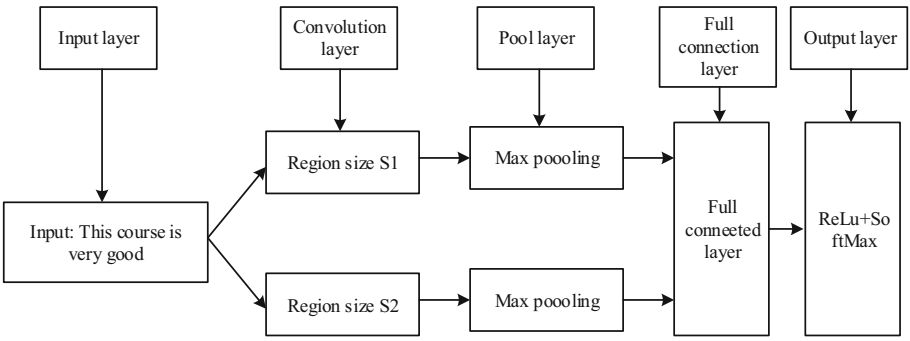

Fig. 1. Grading model for course reviews

designed to grade the curriculum review. The deep learning model for grading course reviews is shown in Fig. 1.

The field of educational data mining focuses on the research of online educational data mining and analysis methods, including prediction, structure mining, relationship mining, model discovery and so on. Prediction, refers to the mining of online education data to get a model about a variable, so as to predict the future trend of this variable, such as data trend prediction. At present, the common forecasting methods include classification, regression and potential knowledge assessment. In particular, the potential knowledge assessment, as an evaluation method of students' knowledge mastery, can

evaluate students' knowledge mastery and ability more objectively. It has been widely used in online education and even traditional education.

2.2 Mobile Terminal Online Education Management Model

The forms of data online education are very complex and diverse, which can be divided into structured data and unstructured data according to the different forms of storage, explicit data and implicit data according to the different ways of obtaining, and video information, user information, user behavior information, course information, teaching information and forum information according to the different sources. The various types of data are intricate, so the management requirements are extremely high. Centered on the needs of learners, learn from the learning support service model, integrate the human and material resources needed by students, information, advice and advisory support, emotional support and other services. And try to integrate the individualized learning function into the learning process support service module provided by the network teaching platform, on the basis of the traditional service function, combine the decentralized and single individualized learning service with the learning process support service, design a continuous and multi-angle individualized learning support service system for distance education, provide individualized course selection guidance, course learning process counseling and consulting service, multi-dimensional evaluation and other services for students, create personalized network learning environment for students, and solve all kinds of difficult problems encountered in the learning process. The system mainly includes four modules: management service, information and consultation service, resource service and learning process support service. The system of distance education personalized learning support service is shown in Fig. 2.

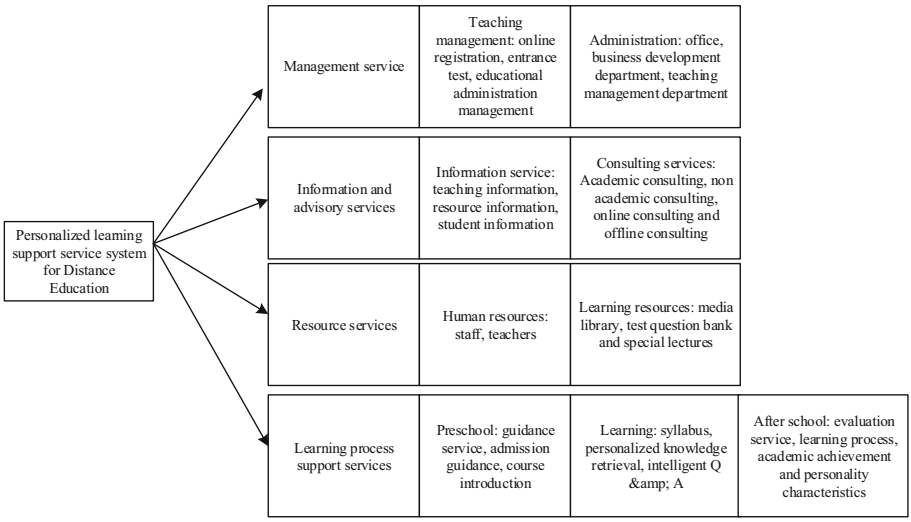

Fig. 2. Distance education personalized learning support service system

Large data mining generally includes data clean-up, transformation, integration, selection, mining, assessment mode and knowledge, etc. In order to ensure the effective integration and utilization of data and facilitate the development of online education platform, the relevant personnel need to mine or extract effective knowledge from a large amount of original data. However, due to the complexity and diversity of the sources of large data generated by school education platforms and the different storage forms, the connection, clean-up and integration of data are very complicated and cumbersome. In the process of data mining, the relevant personnel shall not only comprehensively consider the relationship between multiple characteristics of data, but also carry out independent analysis of data. In addition, due to the differences in analysis methods and tools for different data, it is necessary to have a sound and good data model for processing large data [6]. With the increasing homogeneity of online education, the interactive mobile terminal platform of online education based on big data technology can provide differentiated education services, and online education will be transformed into personalized services. The users of the interactive platform include teachers and learners, and the services provided for them mainly include online teaching content, teaching management, exchange and interaction and learning management. The interactive platform model of online education based on block chain technology is shown in Fig. 3.

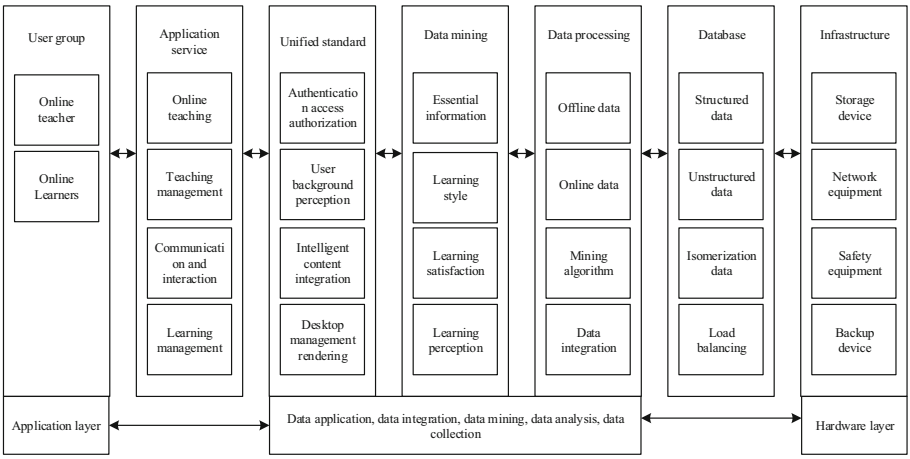

Fig. 3. Online education interaction platform model with blockchain technology

The applications of online education system supported by block chain technology mainly include student oriented mobile terminal function and teacher oriented mobile terminal function. The function of the mobile terminal for students from knowledge map, courses, videos, teaching materials to the forum, all to improve the learning effect of students, so as to improve the service level of the platform and consolidate the development of the platform. The application development of block chain technology is to summarize and summarize the knowledge points of students' learning and teachers' teaching from the angle of students' learning and teachers' teaching, so as to construct

the knowledge framework and the knowledge modules of each course for the convenience of learning and teaching [7]. In addition, they still strengthen the connection between each knowledge dot, facilitate the student to study systematically. The application of personalized course service in online education platform is developed to provide better service for users, which is set up according to the characteristics of human beings. On online education platform, students can selectively watch videos according to their interests and learn what they want to learn, which is different from the overall learning offline and fully respects the personalized development of students. Blockchain technology can make course recommendations based on the test results and learning choices of students during the learning process, in addition to providing similar courses. The caption location of knowledge points is based on the main learning form of the online education platform. The application developed from video teaching is mainly based on caption skipping function for clips of interest to students in the video, and attached with professional vocabulary links and knowledge atlases, which can help students learn in an all-round way [8]. Mobile terminal teaching data training is mainly used to consolidate students' learning with knowledge atlas and teaching data base. Firstly, we set up a teaching data base according to different knowledge points, and pick out relevant knowledge points according to the students' answers, then pick up teaching data randomly from teaching data base to facilitate students' consolidation practice. The service application layer is the window for the external interaction of resources and the bridge for users to use resources, which directly affects the user experience. Therefore, the application service layer requests the reconstruction of information resources according to the user's needs to provide users with personalized service resources. Users do not need to know the integration process of background data resources, and the data resources processing layer of the platform completes completely. In view of the teacher, the platform feedback learner's analysis, especially the learning style and preference, tracking the whole learning process, and the students' behavior on the platform. Using SPSS Clementine 12.0 as a mining tool, C5.0 decision tree algorithm, through data collection, data preprocessing, data classification and rule generation steps, to achieve personalized path recommendation rule building and application. The process of teaching data processing based on blockchain technology is shown in Fig. 4.

According to the four-in-one learning mode of learners, constituting learning, answering questions, testing and evaluation and interaction, enjoy autonomous learning, personalized real-time notes, review and evaluation of targeted courses, and online interactive learning services in multiple ways; and formulate reasonable learning progress and personalized learning programs according to the results of backstage data mining. The preprocessed data are related, classified, clustered and biased, from which valuable interactive information of online education is found. By using the network analysis method of block chain technology, the integrated online educational interactive data are formed by integrating the data nodes and links, and the links reflecting the mutual relations are integrated to form a data network. The commonness of the data and the overall characteristics of the network are shown. Through visual analysis, ranking, classification, clustering and link prediction, we can find a large number of common patterns of educational interaction and extract the value of data relations. Analysis mining involves four steps: Map, sort, partition, and Reduce. The function of each step is independent,

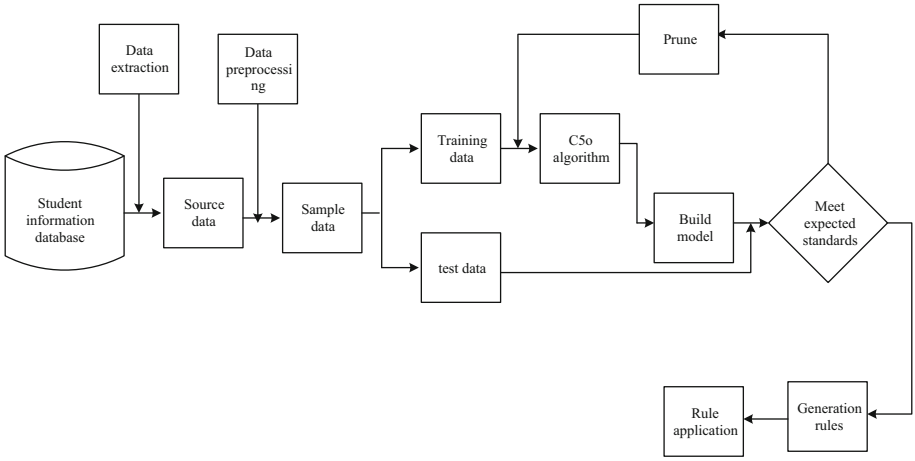

Fig. 4. Teaching data processing flow based on blockchain technology

can not be called each other, between the steps through the "key-value pair connection, the output of the previous step is the next step input." The result of data mining of online educational interaction is unknown and potential, and it can't be used as decision-making basis directly. Only by combining with education quality evaluation system, can it be transformed into the content needed by the platform. Through data filtering, analysis and integration, multi-resource classification results are established, and decisions are made according to users' different needs to facilitate users to access and use services. The purpose of the integrated data is to prepare for the integration of users, to analyze the similarity of users' information resources, to classify the similar users, to provide the allocation of similar information resources according to the basic information, learning style, learning satisfaction and learning perception of online learners, and to realize the customization, personalization and precision services of users. Finally, the analysis results of online education interactive data are presented to the corresponding users simply and clearly, from abstract data to visual structure mapping, allowing users of online education platform to directly observe the relationship between education factors, deducing and reasoning in a very short time, making accurate decisions, and ensuring that the data results provided must be comprehensive, timely and sustainable. At the data processing level, the experience assignment is mainly based on the teacher's experience, but as the user data accumulates, "the algorithm model based on Big Data and Ebbinghaus memory curve will play a role, so that each user's education plan is different."

2.3 Realization of Big Data Online Mining in Education

The process of educational data mining is actually the process of mining knowledge from a large number of data. At the same time, some people call knowledge discovery as data mining, but they are not the same as data mining 3. The main process of data mining includes data collection, data preprocessing, feature extraction, feature selection, data mining and model evaluation. The data mining process is shown in Fig. 5.

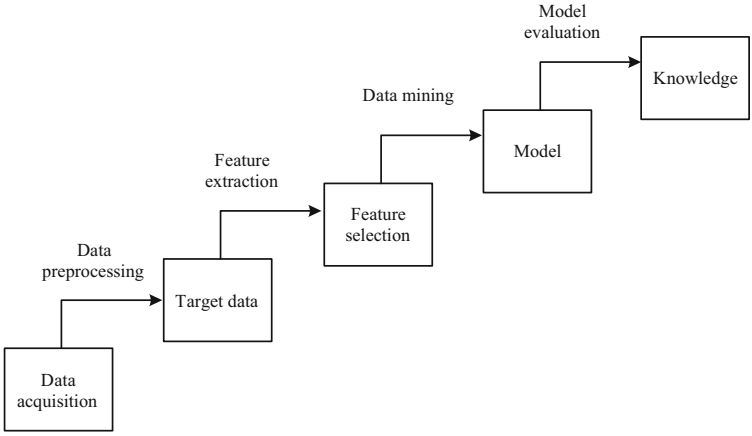

Fig. 5. Data mining procedure

Whether the process of data mining is successful or not depends on the quality of data to a great extent. Therefore, data preprocessing technology plays an important role in model prediction and generalization. In fact, the original data often appear missing value, noise data, incomplete data and abnormal data. These "dirty data" will affect the efficiency of data mining process is not conducive to model training, and may even lead to the deviation of experimental results. Therefore, preprocessing the data before mining, including data cleaning, data integration, data reduction and data transformation, makes the process of searching knowledge and discovering information value more meaningful. Data cleaning. In the information age, every minute and every second of data is generated quickly through the Internet. Dirty data may exist in the data obtained from different ways. In order to achieve high quality data, it is necessary to clean these data. Data cleaning is to solve the abnormal situation that may be encountered in the original data, such as data inconsistency, missing, outliers and so on. In addition, data cleansing is a necessary step in data mining analysis and the most important work in data preprocessing. Different data cleansing tasks aim at different types of errors. Data integration. In order to make the process of data mining more effective and utilize data from multiple sources as much as possible, data with different attributes, dimensions and structures can be integrated together to store and manage them in a unified way. This is why data integration plays a key role. Machine learning is driving the automation of data integration, reducing the cost of data integration in general and improving experimental accuracy. Sometimes the dimensions of different features in the preprocessed data may be inconsistent, and the differences between values may be very large, and the results of data analysis may be affected if not processed. In the various datasets managed, there are also significant differences between the eigenvalues, such as a maximum of 10000 and a minimum of 0.0001. The most common way to generalize attribute values is as follows:

$$\text{value}_1 = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{2}$$

$$\text{value}_2 = \frac{x - x_{\text{mean}}}{x_{\text{stand}}} \tag{3}$$

In formulas (2) and (3), the formula shall be the smallest and largest normalized function and the z-score normalized or naturalized function in the interval. x represents the pre-processed eigenvalues, value_1 and value_2 represents the reduced eigenvalues, $x_{\min}$ and $x_{\max}$ represent the pre-processed minimum and maximum eigenvalues, respectively. In addition, x_{mean} represents the mean of the eigenvector and x_{stand} represents the standard deviation. You can see that as the feature dimension is set to increase from 5 to 100, the performance of the model improves up to 75% in accuracy and over 60% in recall. It is also found that with the increase of the number of hidden units, the curves of the two indexes show a gentle trend. The reason for the low performance in the lower feature dimension may be that the hidden layer with less training times is not accurate enough to represent the students' behavior effectively. The more hidden layer feature dimensions you set, the more likely you are to get a fuller picture of your students. However, the accuracy of the model is more stable than that of the 50-dimensional model, and some repetitive features may appear in the training process. Therefore, it is helpful to set up a suitable feature dimension for students' behavior model, and it also helps educators to understand students' behavior and explore the information contained in data. The principle of knowledge-based information recommendation algorithm is shown in Fig. 6.

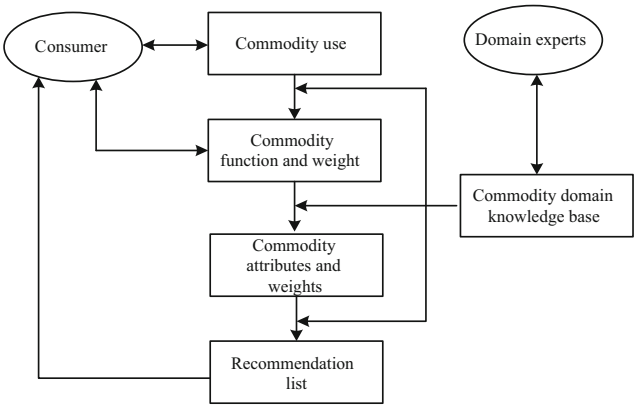

Fig. 6. Schematic diagram of knowledge based information recommendation algorithm

Based on the students' historical information search records, generate the students' historical score matrix (recorded as the A matrix) and then generate the teaching data-knowledge point marker matrix (recorded as the B matrix) based on the knowledge point information involved in the teaching materials marked by experts. These two matrices serve as the original input information to the system. Using the information of Matrix A and Matrix B, combined with the related models of cognitive diagnosis field, the comprehensive cognitive level information of each student was calculated. Similarly, the information of A and B matrices is used to compute the information of each student's

knowledge level. Using the information of the students' comprehensive cognition level and the information of the knowledge points, and combining the traditional collaborative filtering method, we can predict the possible correct rate of each teaching material to be recommended by students. If the final recommended list of teaching materials is Q, then we immediately intersect with List R for each knowledge point less than μ. If the intersection is not empty, we add the intersection to List Q. If the intersection is empty, we iterate through all the predictive accuracy of the relevant teaching materials for this knowledge point, and take the teaching materials closest to [beta 1, beta 2] to add to the list Q. In this way, not only ensure the integrity of the knowledge training, but also ensure the accuracy of the teaching materials within a reasonable range, not because of too simple or too difficult to lead to poor experience, teaching materials recommendation list generation process is shown in Fig. 7.

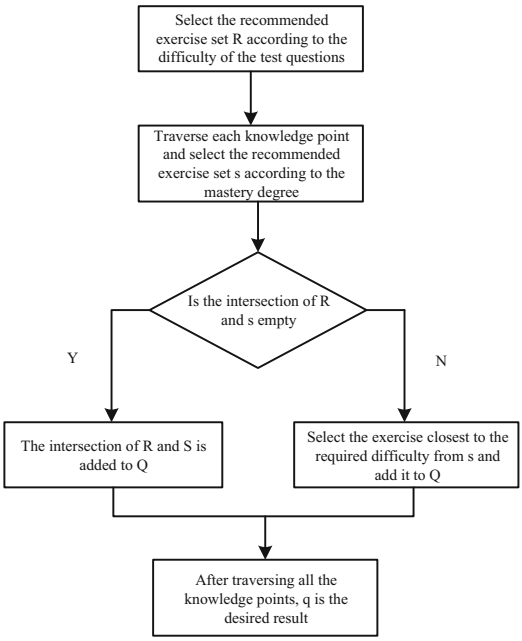

Fig. 7. Generation of the recommended list of teaching materials

There are some defects in the application of data mining technology in online education platform. Although the online education platform has changed the previous teaching model and provided new learning space for students, it is still lack of systematicness and systematicness. In order to improve the application effect of large data mining technology in online education platform, the author thinks, first of all, we should analyze and process the basic operation of online education fundamentally, classify and analyze the data produced by online education, establish corresponding big data model for each kind of data, which is convenient for analysis and processing, reduces the time of data mining and enhances the value of data. Secondly, the application of large data mining technology should be further built, such as teaching database should be updated with

the times, so as to better platform development. Finally, we should strengthen the cooperation with government departments in order to obtain the corresponding information technology support and financial support. The information technology involved in large data mining technology is developing continuously.

3 Analysis of Experimental Results

Use the Python and the no libraries along with Nvidia Tesla K8o GPU. Here are some implementation details: The proposed two-stage classifier model has a 50D embedding layer for campus card devices, and the dimensions of each GRU are set to 50 hidden units. Optimized with Adam 6, the initial learning rate is 0001, and the minimum batch size is fixed at 128. To prevent RNN from over-fitting, two dropout layers are applied in HRNN based on attention: one is that the dropout between GRU layer and GRU layer is set to 25%, the other is 50% between GRU layer and bilinear similarity layer. The BPTT is also truncated to 19 time steps and the batch size is set to 512. In the most advanced method, 6, the number of epoches is set to 30. SVM is used to set c to 0.8, gamma to 0.1, formula 21 is the minimum and maximum normalized function in the interval, and the list of recommended teaching materials is generated by setting the upper and lower limit of recommended difficulty. The algorithm needs to give the mean value of difficulty, beta, and then the algorithm will select [beta 0.1, beta 0.1] difficulty range of teaching data reasonable algorithm, the real problem should be set with the difficulty of the line. Here, we use the SR index in Formula 5.2 to verify whether the recommended difficulty of teaching materials meets the set situation. In this experiment, cognitive modulator was set at 0.35. The experimental results were based on the Exercise dataset, after 100 experiments, take the average value of the results for comparative experiments. And the recommended difficulty of the teaching materials matched the experimental results for example, Fig. 8.

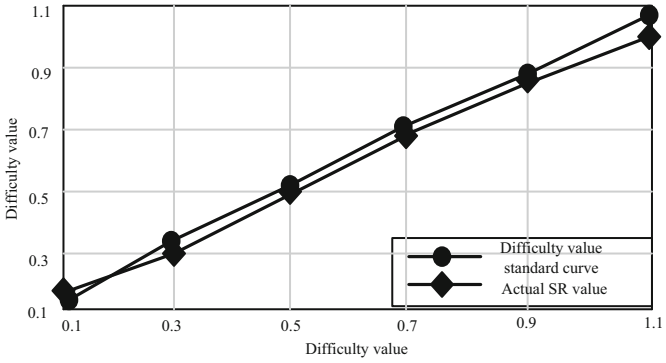

Fig. 8. Comparison chart of recommended teaching data with experimental results

The difficulty value of experiment setting is between 0.1 and 1.1, the straight line in Fig. 8 shows that the expected difficulty value matches the actual difficulty value

exactly. Another curve shows the actual results of the experiment. From the coincidence of the two lines, we can see that the actual difficulty of the teaching materials is basically consistent with the difficulty we set in the algorithm. In the case of high or low difficulty, the error will be greater. This is because at this point in our algorithm according to the knowledge coverage and do some of this treatment has a greater impact, and this affects the real reflection of the difficulty of teaching materials. This also proves that the algorithm in this paper can truly according to the intention of topic selection, select the teaching materials suitable for students' difficulty. Recall rate and accuracy rate are commonly used to evaluate the classification performance. The recall rate and the accuracy rate are the evaluation indexes in the classification task reference information retrieval task. In information retrieval, precision rate and recall rate are usually used to measure the quality of retrieved information. Relevant documents are generally called positive and unrelated documents are called negative examples. In the whole process of information retrieval, there are generally four kinds of results: TP, TN, FP and FN. TP refers to that the search engine retrieves the relevant documents correctly; TN refers to that the irrelevant documents are not retrieved correctly and the irrelevant documents are filtered correctly; FP refers to that the irrelevant documents are retrieved incorrectly and the irrelevant documents are deemed as the relevant documents; FN refers to that the relevant documents are not retrieved incorrectly and the irrelevant documents are not retrieved. The four results in the information retrieval process are shown in Table 2.

Table 2. Four results from the information retrieval process

	Correlation (positive class)	Irrelevant (negative class)
Retrieved	True Positive (TP)	False Positive (FP)
Not retrieved	False Negative (FN)	True Negative (TN)

In order to verify the performance of this method in students' educational data mining task, the following benchmark methods are compared. The experimental results for different data length mining accuracy and recall are shown in Fig. 9.

According to Fig. 9, the highest accuracy and recall of the proposed method can reach 77% and 84%, respectively; The highest accuracy and recall of the comparison method are 63% and 64%, respectively, which are lower than the proposed method. The curves of the two indices show a gentle trend as the number of hidden units increases. When the performance of feature dimension is low, the reason may be that the hidden layer with less training times is not accurate enough to represent the students' behavior effectively. The more hidden layer feature dimensions you set, the more comprehensive you may be about your students. Therefore, it is helpful to establish the model of students' behavior, and it is helpful for educators to understand students' behavior and explore the information contained in data.

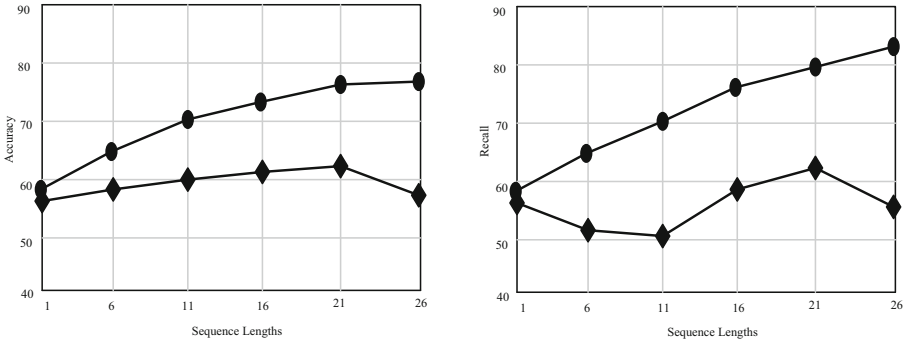

Fig. 9. Experimental results of accuracy and recall of different data length mining

4 Conclusion

In order to accurately mine the required teaching resources from massive data, the research on mobile terminal online education big data mining based on blockchain technology is proposed. Through the characteristics of education data, combined with blockchain technology, education data resources are extracted, and effective management of education resources is realized. The emergence of blockchain has changed our understanding of data. People can find the world hidden behind a large number of data through big data. It also brings new ideas for education and brings opportunities for personalized learning. At the same time, how to guarantee the quality and effect of personalized learning also faces challenges. On the one hand, because of the characteristics of massive data, data processing is different from the traditional data processing, how to effectively process these data is facing challenges. On the other hand, although the education big data theory and technology have a wide application prospect and great development potential in the education field, the application of big data needs the technical achievements of many interdisciplinary fields, and some big data education applications are still in the research and exploration stage. With the progress of network technology and the increase of data volume, in the future work, data mining technology needs to be constantly updated to meet higher-level needs.

References

1. Jin, Y., Chen, H.: Value orientations, realistic limitations and effective strategies of big data mining for education. Theory Pract. Educ. **41**(19), 3–8 (2021)
2. Zhang, T., Zhang, S.: Research on design and computing of learner model of educational big data mining. E-educ. Res. **41**(09), 61–67 (2020)
3. Xu, J., Zhang, G.: Application and practice of data mining algorithms based on apriori. Comput. Technol. Develop. **30**(4), 206–210 (2020)
4. Liu, W., Li, X.: Research on terminal online education data mining technology based on model driving. Modern Electronics Technique **43**(16), 112–114+118 (2020)
5. Lian, J., Fang, S., Zhou, Y.: Model predictive control of the fuel cell cathode system based on state quantity estimation. Comput. Simulat. **37**(07), 119–122 (2020)

6. Wang, D., Liu, H., Qiu, M.: Analysis method and application verification on teacher behavior data in smart classroom. China Educ. Technol. **05**, 120–127 (2020)
7. Wang, C., Cheng, J., Sang, X., et al.: Data privacy-preserving for blockchain: State of the art and trends. J. Comput. Res. Develop. **58**(10), 2099–2119 (2021)
8. Liu, F., Yang, J., Li, Z., et al.: A secure multi-party computation protocol for universal data privacy protection based on blockchain. J. Comput. Res. Develop. **58**(02), 281–290 (2021)