

Research Article

Analysis of Students' Behavior in English Online Education Based on Data Mining

Chunxia Wang 

Baotou Medical College, Baotou 014040, China

Correspondence should be addressed to Chunxia Wang; 102002066@btmc.edu.cn

Received 16 October 2021; Accepted 10 November 2021; Published 23 November 2021

Academic Editor: Mukhtaj Khan

Copyright © 2021 Chunxia Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the formation of global economic integration for better exchange and cooperation with nations around the world, mastering English is extremely essential. In the context of today's big era with a variety of English learning methods, it is required that data mining be applied to online English education. Owing to the continuous application of data mining techniques and the improvement of the online learning system, its application in education is also more and more prevalent. In the face of a large amount of learning data and student behavior data, the traditional methods have the problems of low processing efficiency, more memory requirements, and large prediction error. Therefore, this paper proposes a student behavior analysis method of online English education based on data mining. The student behavior data is collected, and an online English education learning behavior model is established. The data mining model is built to filter the obtained behavior data through data preparation, data statistics, and analysis. Furthermore, the apriori algorithm is used to mine association rules and calculate the similarity of data followed by the application of a fuzzy neural network to mine the behavior data of English online education students. The experimental results show that this method has high data processing efficiency, takes up less space, and produces a low prediction error.

1. Introduction

With the continuous development of information communication technology, people have entered a new information age. The popularization of computers and other electronic devices and the rapid promotion of computer networks have made great changes in people's life, learning, and thinking. The widespread application of computers and the Internet in the field of education has greatly changed the traditional educational methods, skills, and ideas and developed rapidly in various forms [1]. The Internet has become the central information acquisition platform. How to make use of the Internet learning platform for the improvement of the teaching process is the crucial problem to be solved. Using Internet and data mining techniques, the development of an education-related model can provide active support for platform decision-making [2].

The analysis of E-learning behavior in online education is of great importance to formulate and develop a distance education platform with a better learning experience and carry out more efficient and accurate E-learning evaluation

and guidance in learning the English language [3]. In this context, relevant scholars in the educational field and researchers in other fields have conducted in-depth research on students' online learning behavior [4]. Among them, Zhang and Wei [5] proposed a user learning behavior analysis method of the "National Library Open Course" based on big data. The method can realize the systematic combination of association algorithm and result in visualization by using association data mining method and visualization techniques, to optimize students' learning process model. A user learning behavior analysis system was established based on big data, and a nonoverlapping grid optimization mapping was designed to realize the visualization of three kinds of behavior elements: event correlation, event ranking, and network social networking. The user, session, and event datasets in the data resource of "National Library Open Course" were analyzed and tested. The results showed that the system realizes the effective association clustering of user learning behavior data and obtains a good visualization effect. Wu and Tian [6] presented a learning behavior feature analysis method based on the learning

result prediction framework. Based on literature research and interviews with teachers, the learning results prediction framework was divided into four dimensions: student-student interaction, student-teacher interaction, student-content interaction, and student-system interaction, including 10 characteristic variables. Through correlation analysis and information gain (rate) analysis, eight effective characteristic variables were screened to form the final feature set. The results show that the research summarizes and reflects on the research results from the aspects of effective learning behavior indicators and effective learning behavior characteristics, which can provide research support for learning analysis and evaluation in the hybrid learning environment. Jiang [7] used the student academic and behavior data to investigate the data characteristic patterns and correlations and applied these patterns to provide directions for teaching activities and teaching management and enhance the quality of teaching management. The global search process of the genetic algorithm is employed to form a GABP hybrid prediction model and resolve the local minimum problem of the BP neural network algorithm. Results showed that the accuracy of the model is improved significantly. Sun and Jiang [8] summarized the development and application of several data mining techniques in the English online learning system and put forwarded some challenges in the application of data mining in education. Kularbphetong [9] developed a model for the analysis of student behavior using E-Learning based on a data mining technique. The student dataset included 5392 personal records. The model was created using machine learning techniques such as decision trees and Bayesian networks. The result showed that the Bayesian networks technique exhibited higher performance as compared to the decision tree. Monica et al. [10] employed targeted data mining to improve the development of the English online platform and its application. The data source was investigated and the preprocessing steps were analyzed. Moreover, the advantages and disadvantages of different data mining algorithms were also examined. In addition, some scholars have proposed an online learning behavior analysis method based on big data K-means clustering algorithm and developed students' online learning behavior model and behavior analysis platform [11–13].

The above traditional methods have achieved remarkable research results in the analysis of students' learning behavior, which provides a certain basis for students' learning behavior prediction and students' learning problem-solving. However, in practical application, the above methods still have some problems to be solved, such as low processing efficiency, more memory requirement, and large prediction error. This study proposes a method of analyzing students' behavior in English online education based on data mining. The major contribution of this research is as follows:

- (i) An online English education learning behavior model is employed to collect student behavior data.
- (ii) A data mining model is established to filter the acquired behavior data via data preparation, data statistics, and analysis

- (iii) The association rules are mined through the apriori algorithm, the similarity of the data is calculated, and the fuzzy neural network is applied to realize the mining of student behavior data in English online education
- (iv) The model presented higher processing efficiency, with the minimum memory requirement and low prediction error

The rest of the paper is organized as follows: Section 2 provides an overview of the learning behavior in online English education. In Section 3, the proposed data mining technique is explained. Section 4 is about the results, and finally, the conclusion is given in Section 5.

2. Design of Student Behavior Model for Online English Education

2.1. The Concept of Learning Behavior in Online English Education. With the rapid development of distance education, E-learning has become more and more common. At present, there is no clear definition of E-learning behavior. The generally accepted concept is a way for people to achieve their learning effect through the Internet [14]. Some scholars believe that E-learning behavior is also a way of online autonomous learning behavior, distance learning behavior, or online learning. Although E-learning behavior has various names, these names have something in common; that is, learners use computers or other multimedia devices to actively obtain relevant knowledge and information by using various distance education learning platforms [15]. Compared with the traditional curriculum and face-to-face education, E-learning behavior is more flexible, and the learning content is more diversified and accepted by more people.

Online English education learning behavior shows diversified dimensions according to learners' different learning needs, mainly in interpersonal communication, information indexing, information processing, problem-solving, resource sharing, and so on [16]. In addition, English online education learning behavior can also be classified vertically, including low, medium, and advanced, and each category also has different content. According to their actual needs, learners can select appropriate content, control the speed of learning, adopt reasonable learning strategies, and obtain specified learning resources. They can also explore and communicate the learning content with the help of various convenient software to realize network collaborative learning.

2.2. Online English Education Learning Behavior Model. The architecture of the proposed English online education learning behavior model is shown in Figure 1. The model describes the influence and degree of learners, learning resources, learning tools, and learning groups on learning behavior intention and actual learning behavior of English online education.

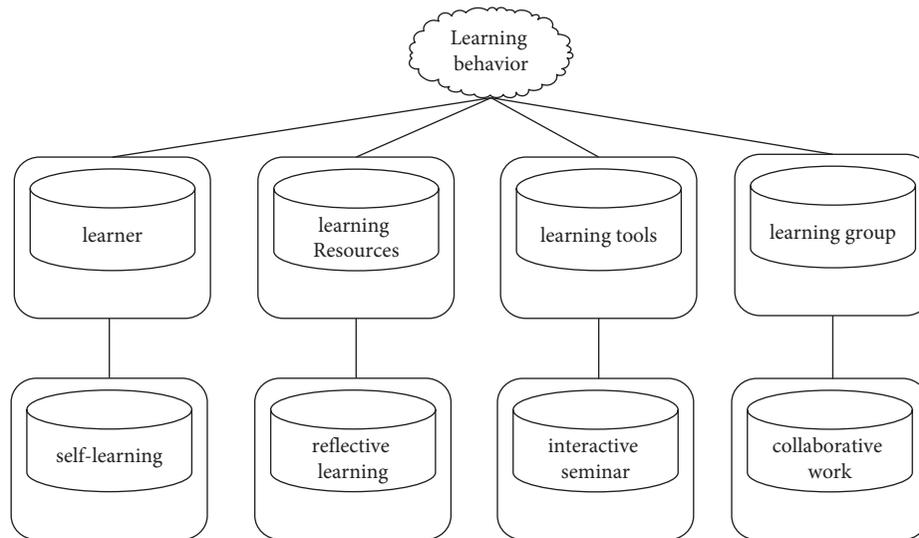


FIGURE 1: The learning behavior model of English online education.

During online English education and learning, the student's actual learning behavior is affected by learners, learning objects, learning groups, and learning tools. The joint action of these four factors determines the specific E-learning behavior. This study divides students' English online education learning behavior into four themes: autonomous learning, reflective learning, interactive discussion, and collaborative work. Each theme is analyzed and studied from the dimensions of participation, social interaction, and cognition. This proposed data model of students' English online education and learning behavior is defined as Who-Do-What, which is described as follows.

- (i) Who (behavior subject): In this study, i.e., student identity is a student English online education and learning platform account composed of strings or numbers, which is unique.
- (ii) Do (behavior activity): It is composed of the operation and operation time of students accessing the e-learning platform, and this part of the data is saved in the platform database.
- (iii) What (behavior object): It is the object that students operate when learning behavior occurs. The learning behavior database system completely records the operation object, including object type, object name, or object identification. The learning behavior model of English online education is the basis of data analysis and mining.

3. Analysis of Students' Behavior in Online English Education

Data mining is a technique for processing information. It can obtain the specific information that people acquire from a large amount of data. This information is generally hidden in a large amount of data and has certain value and significance for practical application [16]. To find the hidden rules and association of different information data in

massive data, it is necessary to mine and analyze them. In this study, data mining techniques are employed to the analysis of students' behavior in online English education to obtain data about students' learning behavior and realize the analysis of behavioral characteristics.

3.1. Behavior Data Collection. Before the data mining of students' behavior in online English education, students' behavior data are collected based on the learning behavior model of online English education. Data collection is divided into different types by different dimensions. According to different collection strategies, the collection methods are divided into weblog-based collection methods and web service-based collection methods. For online learning platform users, it is required to collect learners' personal information. This part of information can be collected when learners register their accounts. Only when learners become platform users can their learning behavior data be further stored. The data collection framework of students' behavior in online English education is shown in Figure 2.

It is depicted that the source of behavioral data of English online education students includes mobile terminal, Web terminal, client terminal, and other terminals. For learner data, it can be divided into two parts: personal information and behavioral data. Data collection methods can be collected through Web services, data burying points, and Weblog records. Weblog records can record the user's access host, authorized users, request date, request type, requested resources, and other information. Log records can be stored in files or can set corresponding fields in the database to store the information in the database. Compared with Weblog records, the data collection method and data burying point collection method based on Web service are more flexible.

Data embedding is to embed codes in relevant parts of the system to record the behavior of learners. It can be performed simultaneously at the front and back ends of the system. Web service can realize data recording through code, which is mainly located at the back end of the system. For behavioral

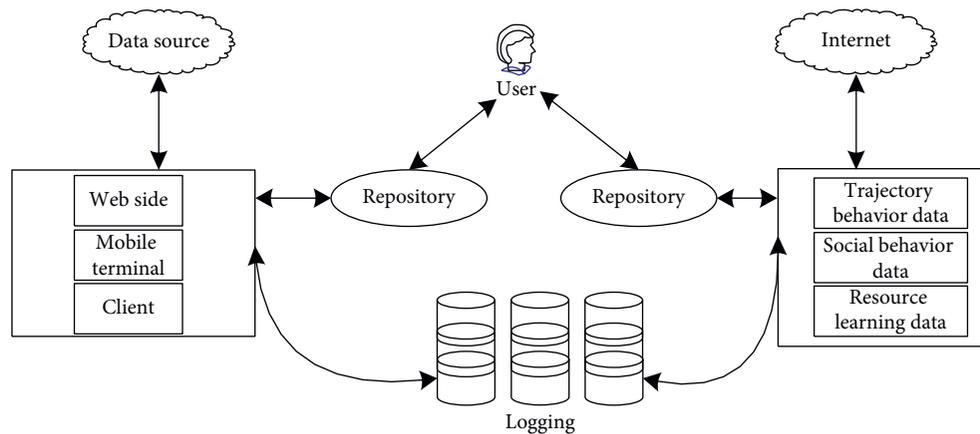


FIGURE 2: A framework for collecting student behavior data in English online education.

data, not all the subtle data of learners can be collected by log records; data burying points and Web services can be more flexible to collect learner behavioral data than log records. System designers can collect target data in any functional module of the system according to the collection requirements. When the learner's behavior data is collected, it needs to be stored in the database. For massive data, a nonrelational database can be selected considering the performance of data read and write, and the massive data can be preprocessed and stored according to requirements, to solve the problem of large memory requirements in the data processing.

3.2. Data Mining Model of Student Behavior in Online English Education. Data mining techniques can be used to collect and analyze student behavior data in online English online education. Using behavior theory and data mining techniques, the proposed data mining process is depicted in Figure 3.

From left to right, it can be divided into data preparation, data statistics and analysis, and result output. The data is prepared to present the data sources and databases that the learner may generate. Data statistics and analysis employ data mining methods to analyze the characteristics of each learning behavior and predict learners' learning styles. The result output part reflects the relationship between each member that affects the effect of online learning and the relationship between online learning behaviors. Although the model presents linear results from left to right, it is a cyclical process. The learner and the network learning platform are both the data provider and the beneficiary of the data, so the data is constantly updated iteratively. The following section explains each of these processes.

3.2.1. Data Preparation. Data preparation is the process of collecting related to learners and forming a network learning behavior database. On the network learning server, there is a large amount of diverse data, which can be found by classifying it according to the description of the object [17]. The data mainly contains two types of information, namely, learner information and information associated with learners. Learner information refers to the attributes of the online learner or the data generated by the learner, such as

the learner's personal information and the data information published on the personal learning environment, learning management system, social network service, and student information system. The information associated with learners in online learning mainly includes contextual semantic information, information about courses learned, and additional information related to courses learned. These two types of information are extracted, processed, and analyzed to establish a database of network learning behavior characteristics. The database refers to a database that stores specific online learning behaviors and expresses certain characteristics, such as online learners' course learning behaviors and login and logout behaviors.

3.2.2. Data Statistics and Analysis. The data statistics and analysis techniques are used to extract learners' various learning behavior rules and characteristics from the network learning behavior characteristic database, and to predict learning styles. It mainly includes the following:

(1) *Statistical Analysis of the Characteristics and Laws of Learning Behavior.* The main online learning behaviors are the interaction behavior between the learner and the learning platform (environment), the interaction behavior between the learner and the learning resources (content), and the interaction behavior between the learner and the learner/teacher (interpersonal). The interactive behavior between the learner and the learning platform includes login, online time, and system feedback. The interactive behaviors of learners and learning resources include browsing resources, retrieving resources, resource interaction frequency, online notes, and resource comments. The interactive behaviors between the learner and the learner/teacher include posting, following, replying, and teacher answering questions.

(2) *Predict Learning Style.* In network teaching, it is essential to consider the evaluation method to judge learners' learning styles. For this purpose most scholars have adopted two methods; one is cooperation, and the other is automatic recognition. Compared with the

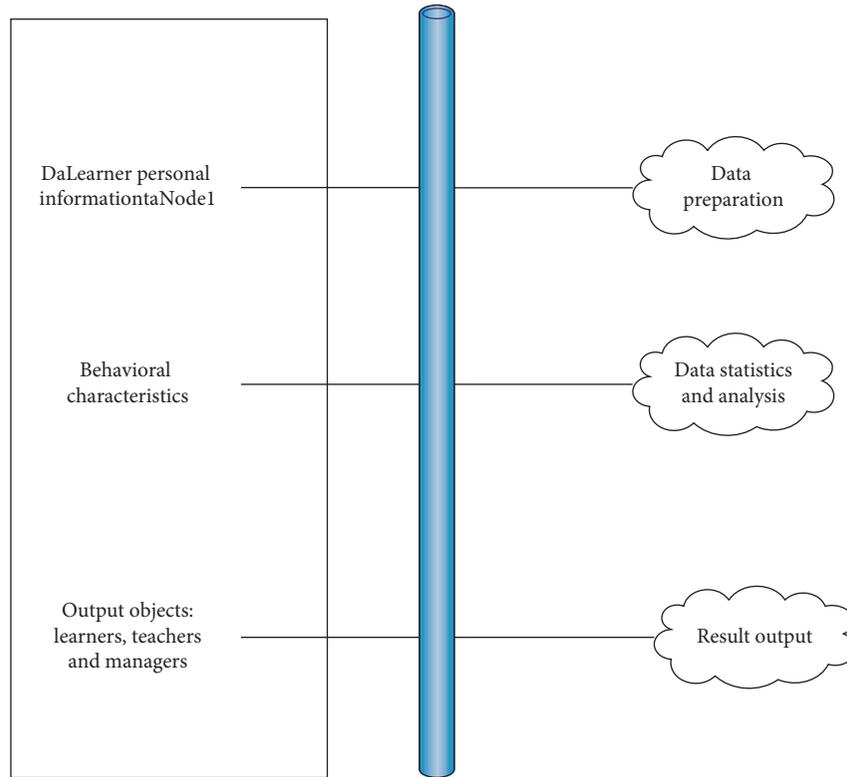


FIGURE 3: Data mining model of student behavior in English online education.

collaborative method, the automatic recognition method is more accurate and convenient. On the one hand, it can save learners' time to fill in the questionnaire. On the other hand, the behavior data is more objective and real. At the same time, the use of automatic recognition methods can dynamically acquire learners' learning styles according to changes in behavioral data and provide some learning support services. However, it must be admitted that the current behavioral data that can be recorded by the online learning platform is limited. Therefore, when conducting related research, it is necessary to screen the acquired behavioral data to a certain extent.

3.2.3. Result Output. The third part of the model is the visualized result output, and the final targets are learners, teachers/education managers, and developers. For learners, visually presenting the results is helpful to conduct self-evaluation and adjust the learning plan according to learning needs. Likewise, the data providers are also the biggest beneficiary of the model. For teachers or education managers, the visual analysis results enable them to understand learners' learning dynamics in time and make timely adjustments to their teaching model design. By presenting the rules of interaction between learners, learning content, and course resources, it is helpful for platform developers to improve the construction of learning resources and determine, which resource functions can be improved to effectively improve students' learning efficiency.

3.3. Mining Association Rules. Association rules are the implication of the forms $X-Y$, in which X and Y are, respectively, called forerunners and successors of association rules; it reflects the relevance and interdependency of one thing to others. If there is a relationship between two or more things, then one thing can be inferred from the other things by association rules. Association rule mining is to find the knowledge of potentially useful information among data from huge data resources. Association rules refer to the interdependence between transaction objects. If there is a certain internal relationship between multiple transactions, then a single transaction can be inferred from other transactions. The purpose is to find the potential between different transactions from the transaction set [18].

In this study, the collected data and corresponding databases are selected to provide the required data for association rule mining. The apriori algorithm is used to scan the database one by one, new candidate sets are generated through the set support, and frequent item sets are found, and association rules are generated according to the set minimum confidence.

Let $X = \{x_1, x_2, \dots, x_n\}$ be a set consisting of n different transactions, where x_i is a certain item in the set $i = 1, 2, \dots, m$. Suppose that we set a transaction data set S , and a transaction h is a subset of X , and the transaction set S is composed of different transactions together forms the association rule transaction database. Assume that a, b and are item sets of type S , where $a \subset X, b \subset X, a \subset b \neq \emptyset$. Based on the frequency of a, b appearing at the same time (support)

and the strength of both a and b (confidence), the establishment of the association rule $a \Rightarrow b$ can be determined. a is called the premise of the association rule, and b is called the conclusion.

The apriori algorithm scans transaction set S multiple times to find all frequent item sets and calculate the frequency (i.e., support) of all individual items S during the first scan to generate candidate item set K_1 , and then, a new candidate item set K_2 is generated through W_1 self-connection, and after scanning S , a frequent item W_2 set is generated after counting each candidate item, and then backward until no new frequent item set can be found. In addition, candidate item sets generate frequent item sets through pruning; that is, item sets whose support of candidate item sets are lower than the minimum support threshold are deleted.

The idea of pruning is that any item set of a frequent itemset must also be a frequent itemset [19]. The new candidate item set K_j is generated through the self-connection of the frequent itemset W_{k-1} , and the premise of self-connection is that the previous $k - 1$ items are the same.

3.4. Similarity Calculation of Student Behavior Data in English Online Education. To realize the accurate mining of students' behavior data in online English education, it is essential to calculate the similarity of the data. In the learning process of students in online English education, the process of learner A selecting the feature vector of the tag set is to analyze the user document according to the adjacent nodes, and the feature data is expressed as

$$A(\partial) = \sum_{i=1}^n c_i e_i + \phi_i(M), \quad (1)$$

where c_i represents the Weblog information, e_i shows the page link information, and ϕ_i is the learning record information; M represents the total amount of behavior data. In learner A learning behavior mining, the semantic stage matrix expression of learning behavior is computed using

$$U_A = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ u_{n1} & u_{n2} & \dots & u_{nN} \end{bmatrix}, \quad (2)$$

where N represents the data dimension.

Because there will be certain similarities between the semantics of student behavior data, the collaborative filtering of synonyms and ambiguous words improves the accuracy of the similarity calculation results and uses cosine similarity to filter synonyms. The cosine similarity is computed as

$$\delta_i(k) = \frac{\text{sim}|y(k) - x_i(k)|}{|y(k) - x_i(k)|^2 + \text{sim}K_u}, \quad (3)$$

where $y(k)$ represents the data static click stream, $x_i(k)$ is the data dynamic click stream, and $\text{sim}K_u$ represents the similarity of the learning behavior data.

The Pearson correlation coefficient method [20] is used to further analyze the relevance of student behavior data in online English education. It is computed as

$$P_i = \frac{1}{N} \sum_{l=1}^N \delta_i(k) \times \bar{q}_i. \quad (4)$$

In equation (4) \bar{q}_i represents the average value of data similarity. Under the constraint of Pearson's correlation coefficient, the calculation criteria for similarity of student behavior data in online English education are as follows:

$$\text{sim}E_i = \frac{1}{N} \exp \left[\frac{(y(k) - x_i(k))^2}{2C_i} \right], \quad (5)$$

where C_i represents the learner's preference similarity.

By analyzing the relevance of student behavior in online English education, the processing of behavioral data is realized, which has a positive effect on reducing the time of data mining and improving the efficiency of data mining.

3.5. Implementation of Student Behavior Data Mining in English Online Education. Based on the student behavior data collection results of online English education of data similarity calculation results, the fuzzy neural network [20] is designed to mine the behavior data of online English education students to explore the hidden information. Based on fuzzy neural network, the steps of student behavior data mining in online English education are divided into the following five stages:

(1) *Selection of Student Behavior Data.* The collected student behavior data is stored in the database to facilitate the realization of data query analysis and visualization functions. In the visualization stage, the selection tools such as rectangles or circles are used to complete the sample data extraction of the database.

(2) *Preprocessing of Student Behavior Data.* To further process the extracted sample data, the completeness and consistency of the data are evaluated. The noisy data is processed, and statistical methods are used to supplement the missing data. The large changes in the data within a period are normalized using

$$F_j^t = \frac{f_j}{\sum_{j=1}^n \alpha_j \times \beta_j}, \quad (6)$$

where f_j represents data processing behavior, a_j is the interaction between learners, B_j shows the interaction between learners and learning resources, and t represents the learner's online time. Through the normalization process, the larger data can prevent the smaller data from having too much inhibitory effect.

(3) *Training Dataset Architecture.* The normalized data is stored in the database. The selected data feature set is set as $V = (v_1, v_2, \dots, v_n)$; the attribute value range of the i data feature is v_i , and if the number of data classification categories is G , then its category set is $G = (g_1, g_2, \dots, g_m)$. In

summary, equation (7) is used to describe the training data set:

$$\sigma_{it} = \max \{ \sigma_1, \sigma_2, \dots, \sigma_m \}^2, \quad (7)$$

where $1 \leq m \leq M$, $1 \leq k \leq m$

The given data is divided into training data groups L_1 and test data group L_2 s. If the number of data contained in the training data group is l_n , then L_a is the input variable element relative to the training data group, and then the best output of the training data group is represented by L_b .

(4) *Fuzzy Neural Network Construction and Training.* Assuming that equation (8) is a fuzzy neural network structure with multiple inputs and single outputs, that is, the input-output relationship expression, and the number of fuzzy rules contained in the association is μ :

$$f_{ij}^2(\mu) = \frac{d_{ij} - A_i}{d_{ij} - A_j}, \quad (8)$$

where A_i and A_j , respectively, represent input and output, d_{ij} represent the input and output variable space, and the expression is

$$\omega^2 = \sqrt{\frac{1}{N} - \left(\frac{R}{2L}\right)^2}. \quad (9)$$

The parameters and structure of the fuzzy neural network are adaptively adjusted, and the fuzzy neural network is trained by learning the preprocessed training data to achieve nonlinear mathematical mapping. After fuzzing the input variables, data mining is implemented using a parameter-free clustering algorithm based on Laplacian centrality and density peaks to obtain the hidden information in the student behavior data of English online education [14].

(5) *Data Mining Verification.* In this phase, the validity of the data is verified. When the error minimization standard is met, the fuzzy neural network training is completed; otherwise, the training dataset is adjusted, and iterative training is carried out until the minimum error standard is met. The minimum error standard is computed as

$$E_\zeta = \sqrt{e_k + \Delta\zeta}, \quad (10)$$

where e_k represents the standard deviation, and $\Delta\zeta$ is the minimum error threshold.

To sum up, by mining the student behavior data of online English education, the analysis of student learning behavior is realized, and the data foundation is provided for student learning summary and teacher teaching research.

4. Simulation Experiment

To comprehensively verify the effectiveness and application value of the proposed student behavior analysis method of online English education based on data mining, simulation experiment analysis is carried out. In the experiment, the user learning behavior analysis method of "National Library Open Course" based on big data and the learning behavior feature analysis method based on learning result prediction

framework are used as the comparison methods. Different methods are compared using statistical analysis, and the result data are analyzed using MATLABR2015 software.

4.1. *Experimental Data Collection.* The data used in this experiment were collected from the NPELS platform, which is a platform for English online learning for every freshman and sophomore in a university. Therefore, the platform contains not only the learning behavior information of current students, but also the learning behavior information of other juniors, senior, or graduated students. To facilitate this study, a class that was currently in use and had frequent interaction was selected as the research object on the NPELS platform. In this class, there were 19 girls and 50 boys, a total of 69 students. The teaching teacher was a young teacher who was very interested in online teaching. The English course was taught in a combination of online and offline.

This study collected the 4-month behavioral data left by the class on the NPELS platform from 2019 to the first semester of 2020. There are two ways to collect data: asynchronous collection and synchronous collection. The so-called asynchronous collection is mainly to collect students' static data, such as name, age, gender, student number, class, and other information. The data will not change due to changes in the learning process but can be carried out before teaching activities. Compared with asynchronous acquisition, synchronous acquisition is the dynamic information of students, which changes with the change of learning behavior. This kind of information includes the behavior information of students' e-learning, such as the time of logging in to the learning platform, the time of staying, the link address clicked, the frequency of e-learning, the number of discussions, the learning content browsed, the type of learning resources frequently visited, the completed tests and test time, the obtained scores, and the retrieved keywords. Information such as the time to complete the homework and the number of homework redoes is the main information to be collected in this study, and it is also an important index to quantify the learning behavior.

Based on the above data, a comparative experiment is carried out. In the experiment, the data processing efficiency occupied space, and the prediction error of students' learning behavior is taken as experimental indicators to analyze the application effects of different methods.

4.2. Result Analysis

4.2.1. *Data Processing Efficiency.* To better compare the method proposed in this study with the user learning behavior analysis method of "National Library Open Course" based on big data [5] and the learning behavior feature analysis method based on learning result prediction framework [6], the three methods are coded by Java language on MyEclipse, and Mushroom in UCI machine learning data source is selected as the data set of association rule mining. Among them, the Mushroom data set contains a total of 8121 transaction records and 23 items. The test environment was Intel(R)

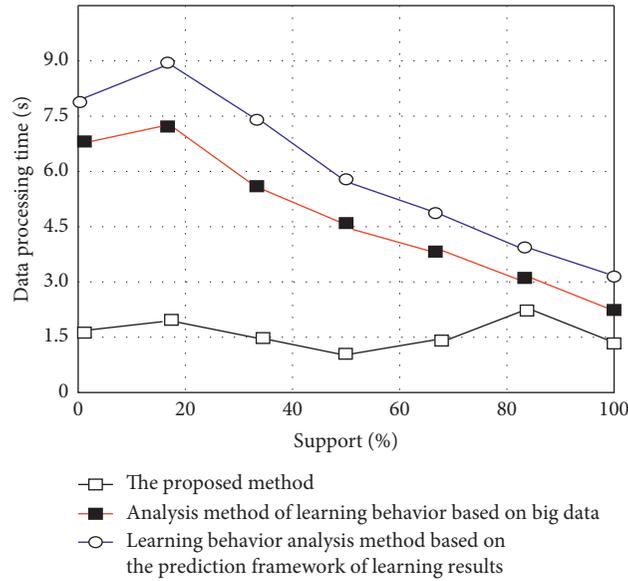


FIGURE 4: Comparison of data processing efficiency under different support levels.

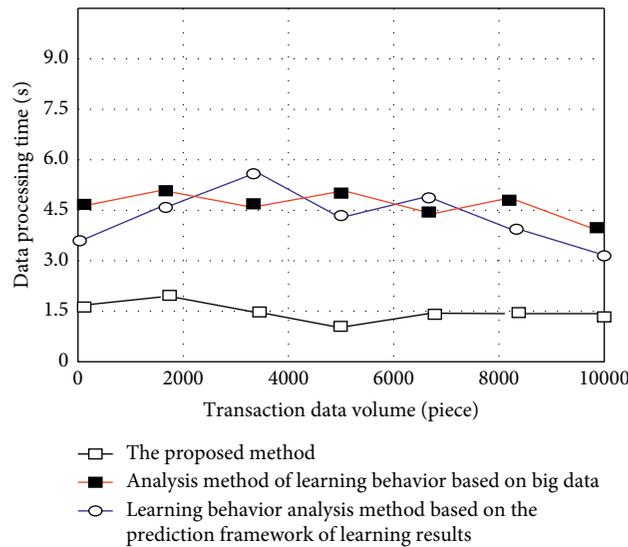


FIGURE 5: Comparison of data processing efficiency under different transaction data volumes.

Core(TM) i7-4790 CPU @ 3.60GHZ with memory 16.0 GB, and the operating system was Windows 7.

Initially, the three methods were run under the conditions of different support degrees, and the running time results are shown in Figure 4.

When the degree of support is small, the processing time of the user learning behavior analysis method of the “National Library Open Course” based on big data and the learning behavior feature analysis method based on the learning result prediction framework is longer and is greater than 0.6 s. As the degree increases, the data processing time gradually decreases. The data processing time of the method in this paper is always less than 3.0 s, indicating that it is not easily affected by the support degree, and the data processing efficiency is higher.

To get the operation of the method under different transaction data volumes, the minimum support is 30%, and 2000, 4000, 6000, 8000, and 10000 data records are, respectively, taken for the calculation. The running time results of different methods are shown in Figure 5.

When the minimum support is the same, the data processing time required by this method is always less than 2.0 s with the increase of transaction data; The data processing time of the “National Library Open Course” user learning behavior analysis method based on big data and the learning behavior feature analysis method based on the learning result prediction framework fluctuates, and the processing time is always higher than that of the proposed method. It shows that the data processing efficiency of the proposed method is better than that of the traditional

TABLE 1: Comparison of memory consumption (Mbit).

Transaction data volume/piece	Reference [5]	Reference [6]	Proposed method
1000	3.257	2.366	0.341
2000	6.946	4.257	0.465
3000	12.131	8.132	0.520
4000	15.249	8.963	0.577
5000	15.965	9.614	0.580
6000	17.253	11.258	0.618
7000	19.369	12.374	0.693
8000	19.457	12.637	0.702
9000	19.904	13.458	0.7258
10000	20.143	14.002	0.7634
11000	21.034	14.527	0.7921
12000	21.439	15.021	0.8029
13000	21.852	15.362	0.8154

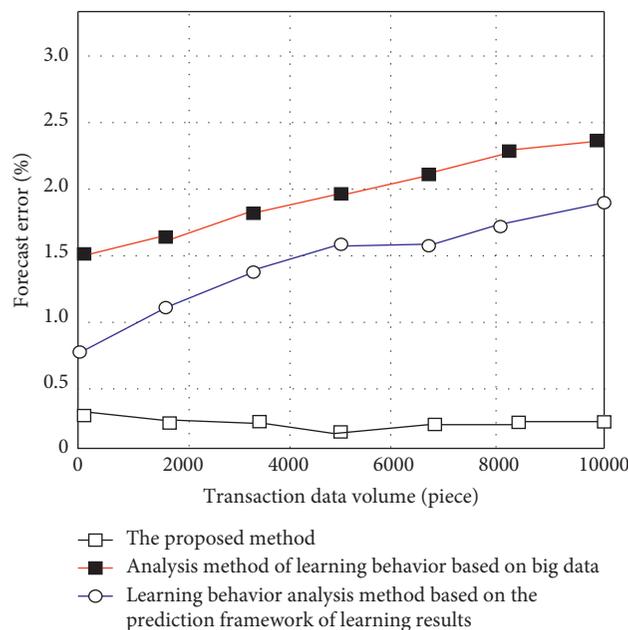


FIGURE 6: Comparison of prediction errors of students' learning behavior.

method under the conditions of different support and different transaction data volume. This confirms that the proposed method has significantly improved the data processing efficiency and shortened the processing time.

4.2.2. Occupied Space. The comparative results of the memory consumption of different methods are shown in Table 1.

As can be seen from the data in Table 1, the amount of transaction data is directly proportional to the memory overhead. The user learning behavior analysis method of "National Library Open Course" based on big data and the learning behavior feature analysis method based on learning result prediction framework have large memory overhead, which leads to insufficient computer memory. In contrast, the memory overhead of the proposed method is small, and the memory requirement is far lower than the two traditional methods. The minimum memory consumption of the proposed method is only 0.341 Mbit, which is 2.916 Mbit and 2.025 Mbit lower than that of the other two methods,

indicating that this method is not easy to be affected by the scale of large data operation and has great advantages in data mining.

4.2.3. Prediction Error of Students' Learning Behavior. Figure 6 shows the prediction error of students' learning behavior for the three methods.

It can be seen that there is an obvious gap between the prediction error of students' learning behavior of the proposed method and the other two methods. This is because the proposed method analyzes the similarity between data and excavates association rules in the analysis of students' learning behavior, to improve the accuracy of students' learning behavior analysis and reduce the prediction error.

5. Conclusion

In the face of a large number of learning data and student behavior data, the traditional methods have the problems of

low processing efficiency, occupying a large amount of space, and large prediction error of students' learning behavior. Therefore, this study proposed a student behavior analysis method for online English education using data mining techniques. The student's behavior data was collected and filtered using the processes of data preparation, and data analysis. To mine the association rules, the apriori algorithm was employed, and the similarity among the data items was computed. The fuzzy neural network was used to mine the behavior data of online English education students. The experimental results show that this method has high data processing efficiency, requires less memory space, and produces a low prediction. In the actual operation process, due to the need to continuously generate candidate and frequent term matrices, the memory consumption is increased. How to improve the efficiency of the algorithm in time and space at the same time needs to be improved and tried in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by 2022 Inner Mongolia Autonomous Region University Research Project (Project Title: Analysis of Student Behavior in English Online Education under the Big Data Field; Project No.: NJSY22064).

References

- [1] R. M. Baker, M. E. Leonard, and B. H. Milosavljevic, "The sudden switch to online teaching of an upper-level experimental physical chemistry course: challenges and solutions," *Journal of Chemical Education*, vol. 97, no. 9, pp. 3097–3101, 2020.
- [2] K. A. Jeffery and C. F. Bauer, "Students' responses to emergency remote online teaching reveal critical factors for all teaching," *Journal of Chemical Education*, vol. 97, no. 9, pp. 2472–2485, 2020.
- [3] Y. L. Bao and Y. Sun, "Research on the application path of artificial intelligence in distance education," *Adult Education*, vol. 40, no. 11, pp. 13–16, 2020.
- [4] G. S. Zhang and X. N. Liu, "Application of digital resource-sharing system for network distance education," *Modern Electronics Technique*, vol. 43, no. 8, pp. 29–31, 2020.
- [5] H. Zhang and D. W. Wei, "Research on user learning behavior of the national library open course based on big data," *Journal of the National Library of China*, vol. 28, no. 03, pp. 63–74, 2019.
- [6] F. T. Wu and H. Tian, "Mining meaningful features of learning behavior: research on prediction framework of learning outcomes," *Education Research*, vol. 25, no. 6, pp. 75–82, 2019.
- [7] X. Jiang, "Online english teaching course score analysis based on decision tree mining algorithm," *Complexity*, vol. 2021, Article ID 5577167, 10 pages, 2021.
- [8] Y. Sun and Y. Jiang, "Application of data mining in english online learning platform," *Journal of Physics: Conference Series*, vol. 1992, no. 2, Article ID 022118, 2021.
- [9] K. Kularbphettong, "Analysis of students' behavior based on educational data mining," *Applied computational Intelligence and mathematical methods*, vol. 662, pp. 1–6, 2018.
- [10] J. Amutha Monica, J. Hilda Malar, and S. Mercy Gnana Gandhi, "Modeling analysis with advanced technology in English OnlinE learning platform," *Journal of Critical Reviews*, vol. 7, no. 12, 2020.
- [11] S. Lai, B. Sun, F. Wu, and R. Xiao, "Automatic personality identification using students' online learning behavior," *IEEE Transactions on Learning Technologies*, vol. 13, no. 1, pp. 26–37, 2020.
- [12] J. Tinoco, M. D. Grant, D. Dias, T. Miranda, and A. G. Simon, "Piezometric level prediction based on data mining techniques," *Neural Computing and Applications*, vol. 32, no. 1, pp. 4009–4024, 2020.
- [13] L. Dey and A. Mukhopadhyay, "Biclustering-based association rule mining approach for predicting cancer-associated protein interactions," *IET Systems Biology*, vol. 13, no. 5, pp. 234–242, 2019.
- [14] L. Bustio-Martínez, A. Muñoz-Briseño, R. Cumplido, R. Hernández-León, and C. Feregrino-Urbe, "A novel multi-core algorithm for frequent itemsets mining in data streams," *Pattern Recognition Letters*, vol. 125, no. 6, pp. 241–248, 2019.
- [15] E. Saccenti, M. H. W. B. Hendriks, and A. K. Smilde, "Corruption of the Pearson correlation coefficient by measurement error and its estimation, bias, and correction under different error models," *Scientific Reports*, vol. 10, no. 1, p. 438, 2020.
- [16] Y. D. Koo, H. Seon Jo, M. G. Na, K. H. Yoo, and C.-H. Kim, "Prediction of the internal states of a nuclear power plant containment in LOCAs using rule-dropout deep fuzzy neural networks," *Annals of Nuclear Energy*, vol. 156, no. 3, Article ID 108180, 2021.
- [17] M.-J. Hsu, Y.-H. Chien, W.-Y. Wang, and C.-C. Hsu, "A convolutional fuzzy neural network architecture for object classification with small training database," *International Journal of Fuzzy Systems*, vol. 22, no. 1, pp. 1–10, 2020.
- [18] A. A. Abusnaina, R. Abdullah, and A. Kattan, "Supervised training of spiking neural network by adapting the E-MWO algorithm for pattern classification," *Neural Processing Letters*, vol. 49, no. 2, pp. 661–682, 2019.
- [19] B. R. He, "Simulation of time series data mining algorithm based on multi-objective decision," *Computer Simulation*, vol. 36, no. 11, pp. 243–246, 2019.
- [20] C. Romero, S. Ventura, and E. García, "Data mining in course management systems: moodle case study and tutorial," *Computers & Education*, vol. 51, no. 1, pp. 368–384, 2008.