Aiming at the problems of low accuracy and large limitations of the current personalized course recommendation method in the educational big data environment, a personalized course recommendation method based on learner interest mining in the educational big data environment is proposed. First, a corresponding online course recommendation model framework is proposed by adopting GRU, which can effectively solve the problems of gradient disappearance and gradient explosion in the process of training the RNN neural network. Then, by introducing an auto-regressive language model, XLNet (Generalized Autoregressive Pretraining for Language Understanding), the information missing problem under the Mask mechanism in the BERT model is effectively optimized, and bidirectional prediction is achieved. Finally, by introducing a temporal attention mechanism into the model, enough attention is assigned to highlight local important information on key information, which improves the quality of hidden layer feature extraction, and a high-accuracy personalized course recommendation based on learner interest mining is realized. The proposed algorithm is compared with the other three collaborative filtering algorithms and the RNN algorithm through simulation experiments. The results show that the precision, recall, and $F_1$-measure of the proposed algorithm in the personalized course recommendation results for different types of courses under the condition of the same database are all optimal. The largest values were 92.1%, 89.3%, and 90.7%, respectively. The overall performance is better than other comparison algorithms. This method can improve the accuracy and optimization limitations of personalized courses and can fully tap the interests of learners. It is of great significance for learners to choose personalized courses in the current educational big data environment.

1. Introduction

Since entering the twenty-first century, people’s production and lives are changing with each passing day under the influence of the Internet. In terms of education, learners’ learning methods have also undergone great changes. The “education informatization” and “Internet + education” are just bred under the new trend of the Internet, and are also the only way for future educational development [1-3]. Education under the Internet environment helps to promote the development of personalized education and promotes the reform of the education system and education mode. The personalized education concept that varies from person-to-person breaks the traditional education and teaching methods. At the same time, various educational institutions also take this opportunity to build online learning platforms, constantly enrich high-quality educational resources, and provide students with more convenient learning experiences and high-quality online learning courses. Supporting education modernization through education informatization, unremittingly helping the innovation and development of education, forming a new education service system, and creating a new mode of integrated development of online and offline education [4-6].

The widespread sharing of a large number of high-quality curriculum resources under the internet environment provides convenience for learners. Learners can arrange learning according to their own time to meet their personalized learning needs [7, 8]. However, online learning has changed people’s learning styles, which also shows some disadvantages. There are three specific problems. First, at present, there are many kinds of courses on the online
learning platform. Different courses are often classified by simple labels and course names, and the unstructured text information in the course description is not fully utilized, resulting in an unclear classification of courses [9, 10].

Second, in view of the rapid development and expansion of the online learning platform, the learning resources on the platform are gradually accumulated. In order to increase the activity of the platform, some online platforms simply pursue the number of courses on the platform and do not do good supervision on the quality of resources on the platform, resulting in the poor quality of some course resources on the platform. Because these inferior resources are not filtered, it has a great impact on the learning effect of the learners [11, 12]. Third, for the learners of the course, the online learning platform cannot provide targeted learning guidance and personalized recommendation to the learners based on the user’s learning style preference and the similarity between the course content and the prerequisites of the course. As a result, users often lose their direction when faced with many online courses and cannot quickly find which courses they need. It ultimately reduces the user’s learning experience and learning efficiency [13–15]. Therefore, in view of the above problems, this paper proposes a personalized course recommendation method based on learner interest mining in the educational big data environment to solve the problem of low accuracy and limitations of the personalized course recommendation method in the current educational big data environment.

2. Related Works

Aiming at the problems existing in the current network education field, it is an important work in the field of intelligent education to study how to fully mine and explore the valuable data of the online education platform and find the relationship between learners and learning resources. On this basis, we accurately recommend the required courses for learners by using multi-source heterogeneous learning behavior data [16, 17]. Reference [18] calculated the importance of external attribute tolerance and internal attribute quality value on the course and built the LDA user interest model on this basis to calculate the user’s preference for the topic and realize the recommendation of personalized learning resources. However, this method does not actually divide the user’s access sequence into different interest segment sequences according to time, so the recommendation accuracy is low. Reference [19] developed an ontology-based hybrid filtering system framework for the recommendation and selection of higher education courses in universities, that is, ontology-based personalized course recommendation. This method is used for personalized course recommendations according to users’ personal needs. However, this method is slow in computation and weak in generalization. Reference [20] designed a personalized online education platform based on a collaborative filtering algorithm by applying the recommendation algorithm in the recommendation system to the online education platform. This method is based on the hybrid programming mode of cross-platform compatible HTML5 and a high-performance framework. But this method does not give a new personalized recommendation algorithm. It is inefficient for a large-scale online learning system. Reference [21] proposed a deep learning method of recommending MOOC (massive open online courses) to students based on the multiaattention mechanism of learning record attention, word-level review attention, sentence-level review attention, and course description attention. This method integrates multiple data sources, takes students’ learning behavior as the basic basis, and realizes personalized course recommendations. However, the computational efficiency of this algorithm will decrease significantly with the increase in data volume, and it cannot be well applied to the case of sparse data. Reference [22] studied and analyzed the English course recommendation technology by combining the bee colony algorithm and the neural network algorithm. Through the deep learning model, the document vector was used to train the acquired text, and the collaborative filtering method was used to realize the recommendation of user courses. However, this method has limitations when it is used in large-scale E-learning systems due to the complexity of computing requirements. Based on the recommendation standard of traditional MOOCs, reference [23] constructed the ontology model of learning participants for the matching process of the personalized recommendation system introduced by MOOC. This method comprehensively considers the knowledge level, ability, and learning speed of learners. However, this method is difficult to obtain the prior distribution, and it is difficult to characterize the high-dimensional semantics of users. Reference [24] analyzed the research status of robust recommendation technology based on the text vector model and support vector machine and constructed the corresponding sustainable economic learning curriculum recommendation model. However, the recommendation accuracy of this method is low and needs further improvement.

Based on the above analysis, a personalized course recommendation method based on learner interest mining in the education big data environment is proposed to solve the problems of low accuracy and large limitations of the personalized course recommendation method in the current education big data environment. The basic ideas are as follows: (1) using GRU to solve the problems of gradient disappearance and gradient explosion in the process of RNN training. (2) Based on the autoregressive language model XLNet, the bidirectional prediction is realized by learning the sequence feature information of different sorting. (3) Time attention mechanism is used to calculate the probability weight of the word vector at different times through the probability weight distribution so that the important words get more attention. Compared with the traditional personalized course recommendation method, the innovation points of the proposed method are

(1) The GRU-coded module can reduce parameters while obtaining the equivalent result value and eliminate the gradient disappearance and explosion problems in the training process.
(2) Using the autoregressive language model XLNet, the problem of missing information under the Mask mechanism in the BERT model is effectively optimized.

(3) The temporal attention mechanism is used to allocate sufficient weight to improve the quality of feature extraction of the hidden layer.

3. Personalized Course Recommendation Method Based on Learner Interest Mining in the Education Big Data Environment

3.1. Model Framework (XATGRU). A recurrent neural network (RNN) is a kind of a time recurrent network, which can be regarded as the result of the same neural network structure circulating on the time axis many times. Compared with other deep neural networks, RNN is better at processing sequence data because of its structural characteristics. Theoretically, RNN can process any length of time series data, but in practical application, it is found that gradient disappearance and gradient explosion will occur in the process of RNN training. This is because the traditional RNN model tends to update in the right direction according to the weights at the end of the sequence. Small GRU parameters reduce the risk of overfitting, and the GRU solves the problems of gradient disappearance and gradient explosion in the process of RNN training neural network and can retain the information from a long time ago. The network structure of GRU is generally similar to that of RNN, but the structure of the hidden layer is more complex. The online course recommendation model framework based on GRU is divided into input, processing, and output sections according to functions, as shown in Figure 1 below.

The input part is mainly to convert the records that the user initially learned into the data format needed for GRU network computing, that is, the vector representation of each user’s learning course. The processing part mainly processes the input data through the GRU network and then obtains the output result. It is necessary to determine the structure of the GRU network, including the total number of layers, the step length of time, and the connection settings between layers. This paper takes the number of courses as the number of eigenvalues, which defines the dimensions of input data and output data, namely the number of neurons in the input layer and output layer. The length of the user’s learning sequence determines the time step required for each calculation. The maximum time step is defined as the maximum value of the user’s learning sequence. At the same time, the length of the sequence should be specified when reading each user’s learning sequence. Thus, the structure of the entire GRU network model is clear. The Softmax layer maps the value of the output vector of the GRU processing layer to the (0, 1) interval, and the output part can take the last dimension of the Softmax layer processing result to determine the final recommended course vector. Because the role of the softmax layer is to convert the output results of the neural network, the output results are expressed in the form of probability.

3.2. XLNet Pre-Training Model. Unsupervised learning models are divided into Auto-Regressive (AR) language model and Auto-Encoding (AE) language model. Different from the traditional AR language model, the AE language model represented by BERT realizes bidirectional prediction. XLNet realizes bidirectional prediction based on the AR language model. Its core idea is to rearrange the input sequence through the Attention Mask matrix in Transformer. At the same time, it does not change the original word order, and effectively optimizes the information missing problem under the Mask mechanism in the BERT model. Because the mask mechanism in the pretraining stage mainly predicts the words out of the mask by masking some words. The Mask mechanism of XLNet is shown in Figure 2.

In Figure 2, the light-colored circle indicates that the model can take its position information into account, and the dark-colored circle indicates that the model cannot take the position information into account. Taking the input vector \( x = (x_1, x_2, x_3, x_4, x_5) \) as an example, a rearrangement combination of \( x \) is represented by \( \tilde{x} = (x_3, x_2, x_5, x_4, x_1) \). As for the vector \( \tilde{x} \), since \( x_3 \) is located at the first position of the sequence, other word information cannot be used, and only the previous implicit state
information can be used. \( x_3 \) is located at the third position in the sequence, and the first three position information can be used.

Given that the sequence length is \( A \), the total number of sorting methods is \( n = A! \). The model can learn various contexts through \( n \) various sorting methods. In practical application, XLNet randomly takes samples of partial permutation in \( n \). The formula of the full permutation model is shown in the following formula:

\[
\max(\alpha)S_{w,W} \left[ \sum_{a=1}^{A} \lg P(a | x_{w,a}) \right],
\]

where \( S \) represents the sequence set, \( w \sim W_A \) represents all possible text arrangements, \( x_{w,a} \) represents the current word. \( X_{w,a-1} \) represents the previous words of \( a - 1 \). \( P \) represents the probability that the prediction result is the current word. \( \alpha \) represents a parameter.

The core of XLNet is Transformer-XL, which introduces the idea of relative position encoding and recurrent mechanism on the basis of transformer structure. The transformer specifies that the input sequence is a fixed length sequence in the training. After the long sequence is segmented in the training, the model cannot make use of the links between the segments, which will cause the problem of missing information. Transformer-XL inserts implicit state information between segments. The prediction of the current segment can use the information of the previous segment through implicit state information, so the model can learn more long-term semantic information. The information transmission mode of the recurrent mechanism between the two segments is shown in Figure 3.

In Figure 3, the red dotted line represents the memory information. The cache information from Segment 1 can be used in Segment 2 training. XLNet realizes the transfer of historical information through this mechanism.

The Transformer encodes the absolute position into a vector in the form of a sine function. The upper layer can learn the relationship between the relative positions of two words through this vector. The calculation formula is shown in the following formula:

\[
\begin{align*}
    e_{t+1} &= f(e_t, L_{t+1} + U_L), \\
    e_t &= f(e_{t-1}, L_t + U_L),
\end{align*}
\]

where \( e_t \) represents vector encoding at time \( t \). \( L \) represents the position encoding of the current segment text vector, \( U_L \) represents the position code, which is the same in different segments. The model cannot accurately determine the specific position of each segment through vectors. The absolute position code is the same for the same position encoding of each segment, while Transformer-XL can use the historical information of different segments. Considering that different segments and words with the same position code have different information contributions to the current segment, Transformer-XL uses the idea of relative position encoding, which calculates the relative distance according to the current position and the position to be used when calculating attention.

Taking the Transformer-XL framework as the core, XLNet can obtain more accurate word vector representation by introducing the recurrent mechanism and relative position encoding. XLNet considers bidirectional semantic information and mining long-term historical information.

### 3.3. Data Normalization.

The neural network usually needs to normalize the input data before calculation to limit the data to a certain range, which ensures that the model can converge quickly and have the same metric for data characteristics. Here, One-Hot encoding is used to normalize the input data. The one-hot encoding adopts binary vector form,
3.4. GRU. GRU is a variant of RNN and has fewer parameters than LSTM. The basic structure of GRU is shown in Figure 5.

The data update formula of the basic unit in GRU is shown in the following formula:

\[
\begin{align*}
    g(t) &= \sigma [ω(g)x_t + U(g)h_{t-1}], \\
    c(t) &= \sigma [ω(c)x_t + U(c)h_{t-1}], \\
    h_{t} &= \tanh[c(t) ◦ h_{t-1} + ax_t], \\
    h_t &= [1 - g(t)] ◦ h_{t-1} + g(t) ◦ h_t,
\end{align*}
\]

(3)

where \(g(t)\) is the update unit module, which is responsible for determining how much \(h_{t-1}\) pass to \(h_t\). If \(g(t) \approx 1\), \(h_{t-1}\) will almost be directly copied to \(h_t\). On the contrary, if \(g(t) \approx 0\), it will not be directly passed to \(h_t\). The reset gate \(c(t)\) determines how much of the previous memory module information will flow to the current \(h_t\). The symbol ◦ represents the operation of dot product. Compared with LSTM encoding, GRU encoding modules not only have fewer parameters but can also obtain equivalent result values. The bidirectional GRU module can not only use the past information but also combine the future word information.

3.5. Attention Mechanism. The attention mechanism is outstanding in speech recognition, machine translation, part of speech tagging, and other serialized data. The attention mechanism can be used alone or as a layer of other hybrid models. It can be placed after the text vector input layer or after the training data of other network models. Through automatic weighting transformation of the data, connecting two different parts to make the whole system perform better and highlight keywords. The attention mechanism is like the principle that the human brain observes something, such as people observing a painting in order to describe the content of some paintings. They will first observe the words in the title of the picture, and then they will observe the part of the picture that expresses the theme purposefully according to their judgment. When describing this painting, people often describe the most relevant content of this painting first, and then describe other aspects. The attention mechanism is a mechanism that highlights local important information by allocating sufficient attention to key information. It can generally be divided into two types: temporal attention mechanism and spatial Attention mechanism. The temporal attention mechanism is mainly used here. The attention mechanism is a kind of attention resource allocation mechanism similar to the human brain. It calculates the probability weights of word vectors at different times through probability weights, so that some words can get more attention and finally improve the quality of feature extraction of the hidden layer. The basic structure of the attention mechanism is shown in Figure 6.

4. Experiments and Analysis

4.1. Experimental Environment and Dataset. The relevant parameters of the simulation experiment environment are shown in Table 1.

The experimental dataset comes from the actual operation data of an online teaching website, with 14370 users and 816 courses, it is mainly related to some courses related to computer subjects. The data mainly includes the user’s learning records and scoring records. From May 2018 to July 2021, a total of 157825 records were recorded. Among them, the training set accounts for 80%, a total of 126260 records, and the test set account for 20%, a total of 31565 records.

4.2. Evaluation Index. The performance of the model is measured by the results of the model extraction and the actual results. The evaluation indexes include precision (\(P\)), recall (\(R\)), and F1-measure (\(F1\)). The calculation methods of different evaluation indexes are shown in the following formulas (4)–(6).
where $ST$ is the number of knowledge entities and relationships correctly identified by the model. $S$ is the number of knowledge entities and relationships identified by the model. $SG$ is the number of all labeled knowledge entities and relationships.

4.3. Model Training. In order to verify the effect of our model, the comparison model is the classical Pipeline model. The experimental results on the dataset are shown in Table 2.

The overall $F1$ of the two models is shown in Figure 7.

From the above experimental results, it can be seen that for the task of entity recognition and relationship extraction, the proposed personalized course recommendation model XATGRU based on learner interest mining in the education big data environment has improved in precision, recall, and $F1$-Measure compared with the Pipeline model.

4.4. Experimental Comparison and Analysis. In the following, the personalized course recommendation method proposed in this paper is compared with the collaborative filtering algorithm in reference [20, 21, 23]. The indexes of recommendation results of different methods under the same dataset are shown in Table 3.

The following is a comparative analysis of the personalized course recommendation method proposed in this paper and the RNN algorithm. The indexes of recommendation results of different methods under the same dataset are shown in Figure 8.

It can be seen from Table 3 and Figure 8 that when the same database is used, compared with the collaborative filtering algorithm in reference [20, 21, 23] and the traditional RNN algorithm, the precision, recall, and $F1$-Measure of the proposed algorithm for personalized recommendation results of different types of courses are optimal, and the maximum values are 92.1%, 89.3%, and 90.7%, respectively. This is because the introduction of GRU solves the problems of gradient disappearance and gradient explosion in the training process. The XLNet model based on autoregressive language is used for bidirection prediction. The missing information caused by the Mask mechanism in the BERT model is effectively optimized and greatly improves the accuracy of personalized course recommendations.
In view of the low accuracy and large limitations of personalized course recommendation methods in the current education big data environment, a personalized course recommendation method based on learner interest mining in the education big data environment is proposed. The proposed method is verified by simulation experiments. The results show that the network structure of GRU is more complex, but it can effectively solve the gradient disappearance and gradient explosion problems in the training process of RNN, and the number of parameters is small, which can reduce the risk of overfitting. This neural network can improve the accuracy of personalized course recommendation methods and solve the problem of large limitations. XLNet based on the autoregressive language model can effectively optimize the information missing problem under the Mask mechanism in the BERT model and realize bidirectional prediction. The temporal attention mechanism can change the importance of different words by means of probability weight distribution, thus improving the quality of feature extraction in the hidden layer and the accuracy of personalized course recommendations. This method is of great significance to solve the problem of low accuracy and limitations of personalized course recommendation methods in the current educational big data environment. Future work will further study the relationship between course reviews and courses. On this basis, consider mining the information from course reviews to discover the relationship between courses from a diversified perspective and to achieve more accurate personalized course recommendations.
Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References


