A Prediction Method of Students' Output and Achievement in Higher Vocational English Online Teaching Based on Xueyin Online Platform

Dan Wang(✉) and Lihua Sun

Liaoning Petrochemical College, Jinzhou 121000, China
wdd6162023@126.com

Abstract. By predicting students' achievements in advance, teachers and administrators can make and improve teaching plans, optimize teaching resources and improve teaching results in advance. Aiming at the problem of insufficient accuracy of traditional prediction methods, this paper studies a prediction method of students' output scores in online English teaching in higher vocational colleges based on the online platform of Xueyin.Collect and sort out student behavior data on the online platform of Xueyin, and divide learning behavior into two categories, namely, student basic data and online learning behavior data.Implement data cleaning, data transformation and missing value filling for student behavior data.The attention mechanism is introduced into LSTM to build an Att-LSTM prediction model. The attention mechanism helps LSTM quickly filter out important information from a large number of feature data, focus LSTM on the information that is most helpful for completing the current task, and improve the prediction accuracy of the model by filtering out unimportant data.The results show that the average absolute error and root mean square error are smaller and the coefficient of determination is larger under the application of the research method, which shows that the research prediction method has good effect and higher accuracy in predicting student performance.

Keywords: Xueyin Online Platform · Vocational English · Student Output Performance · Att-LSTM Prediction Method

1 Introduction

With the popularization of the Internet and the progress of information technology, more and more groups accept the teaching and sharing of knowledge through digital information. The new coronal epidemic in early 2020 further pushed online teaching to the track of mainstream teaching mode. Online teaching is no longer only the main way for primary and secondary students to cram lessons, but also the inevitable choice for colleges and universities to change the teaching mode under the epidemic. At present, there are

L. Yun et al. (Eds.): ADHIP 2023, LNICST 547, pp. 275–290, 2024.
https://doi.org/10.1007/978-3-031-50543-0_19

many network teaching systems for teachers to teach courses through the network. However, online teaching is not as intuitive as classroom teaching, which leads to teachers' less accurate judgment of students' learning than classroom teaching. Therefore, how to effectively analyze the effect of online teaching has always been a hot issue in teaching research. Many researches on online teaching show that the data related to recording students' learning paths and learning performance can be used as an important basis for teachers to evaluate students' learning effectiveness and diagnose students' learning difficulties. These data have valuable research value [1]. Therefore, the effective evaluation of students' academic performance is an urgent problem to be studied. This research proposes an integrated analysis algorithm, which processes these heterogeneous large-scale learning records and integrates multiple perspectives to analyze these learning record information, so as to identify students' learning behavior, and predict students' possible learning effects according to their current learning situation. So that teachers can provide auxiliary teaching strategies to students who may have learning difficulties according to these predicted information.

At present, relevant literature has conducted research on methods for predicting student grades. Reference [2] applies machine learning technology to educational systems, constructs a prediction and analysis model for student academic performance data, and develops a data analyzer to predict students' future academic performance by analyzing their academic performance data. However, the information mined by this method is not accurate and comprehensive enough, and the prediction accuracy needs to be further improved. Reference [3] uses deep learning techniques, CNN models are used to extract local features, and LSTM models consider the advantages of global text order. By classifying educational texts on online learning platforms and analyzing fine-grained emotional tendencies, factors that affect academic performance are explored to achieve prediction of students' online learning performance. But this model requires more parameters, which means that deep learning requires more training data, so it is not suitable as a universal algorithm.

This research proposes a prediction method of students' output performance in online English teaching in higher vocational colleges based on the online Xueyin platform. Through students' learning behavior data on the online Xueyin platform, we can observe students' various behavior performances, and let students' learning behavior data speak, so as to find out students' learning behavior that has greater relevance to students' learning effect after learning. Finally, a performance prediction model is constructed to study the behavioral characteristics that affect students' performance.

2 A Study on the Prediction of Students' Output in Online English Teaching in Higher Vocational Colleges

2.1 Collection of Student Behavior Data on the Online Platform of Xueyin

Student achievement prediction is a process of discovering the potential effective information in teaching data through mining and analysis based on the data of curriculum, students' historical achievements, students' behavior, etc., and then evaluating and inferring students' performance in the future learning stage. Based on this, data collection is the first step of performance prediction.

"Xueyin Online" platform is developed and operated by Beijing Xueyin Online Education Technology Co., Ltd., a wholly-owned subsidiary of Superstar Group Co., Ltd. It is a new generation of open learning platform based on the concept of credit bank jointly launched by Superstar Group and National Open University, and a public platform for higher education, vocational education and lifelong education, It is also one of the selection and operation platforms of national high-quality online open courses. "Xueyin Online" provides a new educational structure and learning mode characterized by "lifelong learning", "ubiquitous learning" and "future learning" for social learners, builds personal end learning files based on credit banks that realize the storage, certification, accumulation and conversion of learning achievements, and helps build a learning society. "Xueyin Online" provides learners with diversified and personalized learning services by integrating a large number of high-quality digital learning resources and courses from various schools and educational institutions, and authenticates, accumulates and converts in accordance with a unified learning achievement framework and standard to achieve the goal of "learning without boundaries, credits can be accumulated, achievements can be converted, and quality can be trusted". The online Xueyin platform provides support for online English teaching in higher vocational colleges. Its teaching scheme covers three teaching links before, during and after class. Pre class preparation, student preview, online teaching and software practice operation in class, review, assessment and teaching evaluation after class, through this whole process, students can realize the integration of knowledge. At the same time, the online platform's real-time data collection, cloud processing analysis and real-time feedback of results are conducive to the innovation and reform of teachers' teaching mode and teaching organization form, thus forming a complete online platform curriculum framework (see Fig. 1).

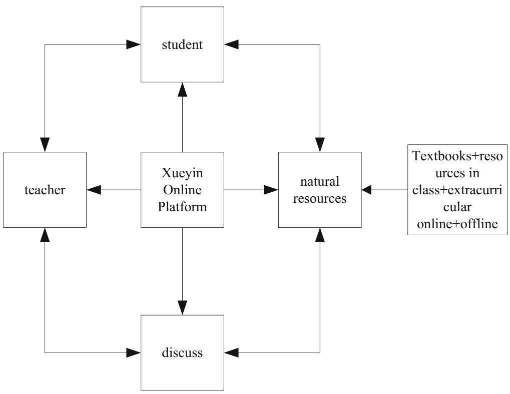

Fig. 1. Course framework of online platform of Xueyin

It can be seen from this that the online platform of Xueyin has completely saved the whole learning process of students, including their basic information, student number, name, social occupation, initial education background, learning terminal, learning duration, number and score of physical examination, credit, number of students' posts and replies, etc.; The database of the system records the interaction between students and the platform server in detail, which can truly and completely record the online learning process of students, and can obtain a large amount of student data at any time.

This paper collects students' learning behavior data through our school's online teaching platform, sorts out and classifies the factors that may affect students' performance according to the collected data, and then divides the learning behavior into two categories. The first category is students' basic data: student student number; The second category is the online learning behavior data of students: task point completion rate, video task point completion rate, video viewing duration (minutes), chapter test completion rate, chapter test average score, chapter learning times, homework completion rate, homework average score, course interaction volume, sign in completion rate, and score, of which the score is the score prediction category.

In order to facilitate the subsequent performance prediction and analysis, the online learning behavior data of students are described, as shown in Table 1.

Table 1. Characteristics and Meaning of English Online Learning Behavior

Characteristics of learning behavior	meaning
Completion amount of video task points	Number of completed watching videos
Video viewing duration	Total time spent watching videos
Chapter quiz completion count	Number of completed chapter tests
Number of completed assignments	Number of assignments completed
Chapter learning frequency	Number of learning times for each chapter
Number of completed check-in	Number of times students attend classes on time
Course interaction volume	Number of interactions between teachers and students
Chapter Test Average Score	Average score of all chapter tests
Average homework score	Average score of all assignments
Final exam	Final Exam Total Score

The student performance data of higher vocational colleges is saved in the school's educational administration system, which stores the scores of each student and each course as well as their comprehensive rankings. With the consent of relevant departments, this paper obtained the final scores of students in 7 different majors through the school educational administration system of the school. Student performance data includes the examination results, classroom results and credits of each course obtained by students in different courses. The fields in the obtained performance data include academic year, semester, student student number, course code, examination results, classroom results and credits. For the score information, this paper obtained the final exam scores of all students from the educational administration system with the approval of the relevant departments of the school. The score information generally includes the student's ID, name, course name, course hours, course credits, assessment methods, assessment scores and other information [4]. In this paper, the test scores are summarized for the purpose of studying the relationship between students' online behavior on campus and their scores.

Next, in order to build a more effective performance prediction model, we must further refine and process these data, and analyze whether each learning behavior feature is related to the performance and the correlation between each behavior feature. For the applicability of data types and correlation coefficients involved in this study, three major statistical correlations are selected.

The Pearson correlation coefficient, one of the coefficients, is shown in Formula (1) to analyze the relevant attributes of this article and verify the analysis results.

$$s = \frac{n\sum x_i C_i - \sum x_i \sum C_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum C_i^2 - \left(\sum C_i\right)^2}} \tag{1}$$

where, x_i Represent sample data, such as age, learning duration, etc.; C_i Represent the learning effect, i.e. the end result; n Number of representative samples; s Represents the correlation coefficient between the sample data and the final grade. correlation coefficient s The size of represents the correlation between the sample data and the final grade, which is linear and most commonly used [5].

2.2 Data Pre-Processing

The obtained original data has many types, a wide range of sources and a large amount of data, which makes it inevitable that there are some problems in the data, such as data missing, data exceptions, etc. Data pre-processing refers to the process of filtering, reviewing and integrating the acquired data before conducting research and operation. In order to ensure the reliability of data mining results, the data is preprocessed to eliminate the redundancy, inconsistency and other problems in the original data. Data cleaning, data transformation, missing value filling, etc. are common methods of data preprocessing [6].

(1) Data cleaning

In the process of data collection, the preliminary work is only based on the teaching business of the school and the needs of teachers and students.

Combine and improve, integrate and process the data according to the actual meaning of the data fields of the three party data sources, and the data will be.

The integrated data is cleaned and improved with the research of this paper. Due to the existence of null value, data noise, data redundancy and other data format problems, it is not suitable for data mining [7]. Due to the particularity of online academic education, the courses chosen by students are also very different. Some students fail to take the final exam due to reasons such as dropping out, repeating a grade, or not registering, or some students fail to take the final exam due to other reasons such as work. Although they have student information in the information database, their scores are all empty, so students who repeat a grade or drop out only consider their scores, All absent students fill in the score as default, which is used to smooth the noise data. The general principles of data cleaning in this paper are as follows:

When selecting data mining objects, because there are many courses offered by schools and most people choose different courses, the selected data are the public courses selected by most people and the corresponding student and learning information.

The students who lack the final grade of the course are meaningless for analyzing learning behavior and constructing the grade prediction model, so they are deleted directly. The data set in this paper excludes invalid data of more than 10000 people,

Due to the special reasons of online academic education, the scattered places where students sign up, and the different branch campuses, these data that have nothing to do with their scores are directly deleted. This is based on the results of educational psychology and research on the factors affecting their scores in the previous research, such as credits, semesters, branch campuses, etc. The combination of manual and script procedures is mainly used to reduce the data dimension.

(2) Data transformation

Most of the data in the original dataset belong to different unit magnitudes or have different dimensions, such as gender, region, pre school education and other attributes. Or because the data types are different, these data cannot be calculated. In view of these problems, we use numerical methods to conduct numerical operations on character data. The main work is to conduct numerical processing on character data, standardize the data, and summarize and overview the data,A part of data standardization and summary has been done, so data normalization [8] is mainly carried out in this section. After comparing a large number of literature, we choose the minimum maximum normalization method in this paper, so that the normalized data fall within the 0–1 interval, which is convenient for subsequent data analysis and mining. For example, students' age, online learning duration, learning times, physical examination completion, number of posts and replies and other relevant data, see formula (2) for conversion.

$$x_i = \frac{\dot{x}_i - \min \dot{x}}{\max \dot{x} - \min \dot{x}} \tag{2}$$

where, x_i is the normalized data value, $\dot{x}_i$ Is the original data value of this type, $\min \dot{x}$ is the maximum value in the sample, $\max \dot{x}$ is the maximum value in the sample.

(3) Missing value filling

Due to the particularity of distance education, adult education and school running history, it is inevitable that there is no data or missing data in the integrated data, such as the students' score ratio, course completion ratio, learning times, online days, number of posts and replies, and learning duration. For individual students, when there are multiple missing learning data, based on the method of ignoring tuples, all relevant information of the student is directly eliminated without analysis. For students with few missing attributes, multiple interpolation method is used to fill the missing values [9]. For some attributes, the overall missing value has exceeded 50%. According to previous research, it is considered that they are not worth studying in this dataset.

2.3 Performance Prediction Model

The research on the prediction model of student output performance should first clarify the purpose, then carry out correlation analysis from the perspective of students, extract the independent variables of the correlation function, and finally establish the relationship between students and characteristic behaviors. Under the guidance of the previous theoretical research, based on the learning behavior selected in Table 1, as the input of the performance prediction model, students' final scores are taken as the output results of the prediction model. Similarly, the input-output relationship of the final score prediction model is as follows (3).

$$Y(x) = \{x_i | i = 1, 2, ..., 9\} \tag{3}$$

This paper takes students' achievement data and curriculum knowledge points as the main research object, and analyzes the collected student achievement data by constructing a student achievement prediction model. First, preprocess the test score data accumulated by students, and calculate all the change indicators [10] of students in this test with each test as the basic unit. In this paper, the problem of student achievement prediction is regarded as a time series prediction problem, and a short-term memory artificial neural network achievement prediction model integrating attention mechanism is proposed. By integrating the attention mechanism into the long-term and short-term memory artificial neural network time series model, the model can screen key information from a large number of feature data, so as to focus on the feature information that is helpful for students' performance prediction tasks.

Similar to the commonly used neural network, attention mechanism is a method of using neural network to process input data. Attention mechanism is derived from the inspiration generated by the observation process of things. People will focus on a specific position and ignore other relatively unimportant parts. Similarly, the attention mechanism will focus on the key areas after obtaining the overall overview, which is a model similar to human focus. The characteristics of the input model will first be assigned a different importance value, indicating that the importance of each feature is different, and then find the feature that has the greatest impact on the results from all the features, so that the model can make better decisions [11]. Generally, for the same model, increasing the number of input parameters of the model can improve the prediction ability of the model. Correspondingly, this model needs to store more intermediate data. Sometimes, due to too much intermediate data, it is easy to put a lot of pressure on the model, resulting in reduced model effect. By introducing the attention mechanism, we can focus on the feature information that can improve the prediction ability of the current model, capture a large amount of feature information, and filter out the feature information that has little influence. This reduces the pressure on the model to a certain extent, thus improving the overall processing capacity of the model. The use steps of attention mechanism can be roughly divided into three stages:

In the first stage, according to Query and available methods such as vector dot product, Key selects an appropriate method to calculate the correlation coefficient between Query and Key. Vector similarity method and MLP neural network method. The formula of vector similarity method is as follows:

$$B_{(Q,K_i)} = \sqrt{\frac{Q^T K_i}{\|Q^T\| * \|K_i\|}} \tag{4}$$

Among them, $B_{(Q,K_i)}$ Represents the value of the correlation coefficient, $*$ Denotes modular operations, K_i Indicates different keywords, Q Represents a query. about K_i Correlation coefficient value of $B_{(Q,K_i)}$ Can be used D_i Means that $B_{(Q,K_i)} = D_i$.

In the second stage, the Softmax method is used to process the previous correlation coefficient to obtain the weight coefficient. This processing has two advantages. One is to process the weight of all keywords as a probability distribution with a sum of 1. The other is to increase the importance of important keywords through its own functions. The Softmax formula is as follows:

$$d_i = soft\max(D_i) = \frac{\ln D_i}{\sum_{j=1}^{L} \ln D_i} \tag{5}$$

Among them, d_i express K_i Corresponding weight coefficient; L Indicates the total length of the input feature.

In the third stage, the weight coefficient d_i and the attention weight of each position W_i Calculate to get the final attention value, and the formula is as follows:

$$Att_{(Q,K,W)} = \sum_{i=1}^{L} d_i W_i \tag{6}$$

Long term and short term memory network (LSTM) is a special recurrent neural network (RNN) used to process long time series. It is designed to solve the problem of long-term dependence. LSTM adds screening of past states on the basis of RNN, so that it can effectively select more influential states, and extract long-term dependence information from long series data. The LSTM neural network has effectively avoided the problems of gradient disappearance and explosion, and has achieved certain success in speech modeling, translation, recognition and picture description. LSTM model is mainly used to obtain information in input feature data. The attention mechanism helps the model quickly filter out important information from a large number of feature data, and focuses the model on the information that is most helpful for completing the current task [12]. Improve the training efficiency of the model by filtering out unimportant data. The Att-LSTM model is designed and implemented based on LSTM model and attention mechanism. This model has a good effect on predicting students' achievements. The Att-LSTM performance prediction model consists of five layers:

Input layer: Input students' test data into the model. In Chapter 3, various data related to students and courses have been screened and constructed through feature engineering [13]. According to the historical information of examination papers, knowledge points and achievements, the feature vectors that can be recognized are constructed. By setting the exam times window to m, get continuous m Score training sample of the second exam Y_m.

$$Y_m = \{y_1, y_2, ..., y_m\} \tag{7}$$

$$y_m = \{x_1, x_2, ..., x_9\} \tag{8}$$

Among them,y_j Indicates the student's j Times of examination results, x_i It is the student behavior characteristics on the online platform of the Xueyin corresponding to this exam. See Table 1.

LSTM layer: It is controlled by three gates, which are called forgetting gate, input gate and output gate respectively. LSTM needs to filter out the input that needs to be discarded from the cell state first[14]. This operation is carried out through an activation function of forgetting the door. It passed the last time $t-1$ Hidden state of E_{t-1} And current time t Input information for x_t, to calculate a vector in which the value range of each value is between [0, 1]. Indicates the retention degree of corresponding data in the cell state. 0 means that the information is not reserved at all, and 1 means that it is reserved at all. The specific calculation of this process is as follows:

$$g_t = f\left(w_g x_t + R_g E_{t-1} + \delta_g\right) \tag{9}$$

$$f(\cdot) = \frac{1}{1+e^{-\cdot}} \tag{10}$$

Among them, g_t Indicates the forgotten door, x_t Represents the input of the current time, w_g, R_g Indicates the importance parameter, δ_g Indicates the correction offset parameter. E_{t-1} Indicates the last operation $t-1$ the hidden state of LSTM cells. Formula (10) is a sigmoid expression.

After the forgetting gate calculation is completed, the next step is to calculate what new data should be input for the cell state. First, by calculating the last time $t-1$ Hidden state of E_{t-1} and x_t At the input door p_t To determine the data to be updated. Then use E_{t-1} and x_t Obtain new candidate memory cells through tanh function [15] q_t. The calculation is as follows:

$$p_t = f(w_i x_t + R_i E_{t-1} + \delta_i) \tag{11}$$

$$q_t = \tanh\left(w_q x_t + R_q E_{t-1} + \delta_q\right) \tag{12}$$

Among them, p_t Indicates the current operation t Input door of, w_i, R_i, w_q, R_q Represents the importance parameter,δ_q, δ_i Indicates the correction offset parameter, x_t Indicates the current operation t Input data, E_{t-1} Indicates the last operation $t-1$ Hidden state of LSTM cells,q_t Indicates that the current operation may be selected to remember the cell status, tanh() Represents a hyperbolic tangent function.

After calculating the memory cells that may be selected q_t After that, the original memory cells need to be updated $\dot{q}_{t-1}$, determining memory cells $\dot{q}_t$ Subsequent values. The process is when x_t get into g_t When,g_t Will pass $f()$ discard E_{t-1} Some data of When x_t adopt p_t When, p_t Will use $f()$ hold q_t Some data of $\dot{q}_t$ Medium. The calculation is as follows:

$$\dot{q}_t = g_t \otimes \dot{q}_{t-1} + p_t \otimes q_t \tag{13}$$

Among them, $\dot{q}_t$ Indicates the updated memory cell status, q_t Indicates the status of candidate memory cells, g_t Stands for Forgotten Gate, $\dot{q}_{t-1}$ Indicates the last operation $t-1$ The state of the memory cell when.

Calculate the cell state $\dot{q}_t$ After, according to E_{t-1} and x_t To calculate the hidden layer information that needs to be sent to the next cell by E_{t-1} and x_t stay r_t And then calculate the vector through tanh layer, which is the same as r_t The final result is obtained by calculating the obtained result. This step is as follows:

$$r_t = f(w_r x_t + R_r E_{t-1} + \delta_r) \tag{14}$$

$$E_t = r_t \otimes \tanh(\dot{q}_t) \tag{15}$$

Among them, r_t Represents an output gate. w_r, R_r Indicates the importance parameter,δ_r Indicates the correction offset parameter, ${}_r x_t$ Indicates the current operation t's data source, E_{t-1} Indicates the last operation $t-1$ Hidden layer state of, E_t Indicates the current operation t The hidden state of Attention mechanism layer: capture input feature information and filter out feature information with little influence.

Full connection layer: aggregate the results of attention mechanism layer into final prediction data $\hat{y}_j$.

Full connection layer output prediction results $\hat{y}_j$ Then it is necessary to calculate the MAPE value between the actual score and the training sample. When the MAPE value is less than the set error threshold, the optimal Att-LSTM model is obtained and saved. Otherwise, the Att-LSTM model parameters need to be adjusted until the training meets the end requirements.

3 Method Test

3.1 Data Set Preparation and Processing

Att-LSTM is used to establish a prediction model for experimental training, which is used to predict the final scores of online academic education students and guide a vocational college to improve teaching strategies in a timely manner. This paper selects 6000 students' relevant data for model training. In the stage of experimental results display, this paper randomly selects the original behavior data of 100 students on the "Xueyin Online" platform to show the actual prediction effect, and analyzes the relevant data. Next, on the basis of the basic data set in Table 1, the original behavior data of 100 students is processed into a data set conforming to P-MIML. The data set contains information about students' learning behavior and performance in all semesters. It should be noted that the data set contains the information of students' required courses and optional courses. When forecasting, only the same courses learned by students in the same data set are predicted, because the students in the same major of online academic education in this school have the same required courses, but some optional courses are different. If an optional course is given for forecasting,There will be some students who do not choose this course, and their grades are missing, which will lead to the problem of missing marks in the subsequent prediction, and will also affect the prediction results. Therefore, this paper only forecasts public courses in the same major.

3.2 Evaluation Index

Obtaining a model with strong generalization ability is a continuous goal of machine learning tasks. In the process of continuously improving the model, the model evaluation method guides the direction of model improvement. Choosing appropriate model evaluation methods and indicators can judge the performance of the model more objectively and accurately. This paper uses three evaluation indicators to evaluate the student achievement prediction model, as follows.

(1) Mean absolute error (ς)This indicator can reflect the true state of the difference between the predicted value and the actual value, and can also deal with the problem of the offset between the positive and negative errors. Generally, the smaller the MAE value, the better the model fitting effect. The formula is as follows:

$$\varsigma = \frac{\sum_{j=1}^{m} \left|\hat{y}_j - y_j\right|}{m} \tag{16}$$

Among them, $\hat{y}_j$ Is the predicted value of the sample, y_j Is the true value of the sample, m Is the number of predicted samples.

(2) Root mean square error (ζ), used to measure the difference between the predicted value and the true value, which represents the sample standard deviation between the predicted value and the true value. Its value is equal to the square root of the mean square error. The formula is expressed as follows:

$$\zeta = \sqrt{\frac{\sum_{j=1}^{m} \left(\hat{y}_j - y_j\right)^2}{m}} \tag{17}$$

(3) Coefficient of determination (ξ) The formula is as follows. Its molecule is similar to the mean square error, that is, the total error generated by the prediction model. At the same time, its denominator can also be understood as a prediction model, and all the predicted values of the model are considered as the average value of the sample. If $\xi = 0$ means that the forecast model is equal to the benchmark model; If $\xi = 1$, it means that the prediction model will not have any error and the prediction result is perfect. ξ The value range is between [0, 1]. In general, the coefficient of determination ξ The closer the value of is to 1, the better the prediction model is.

$$\xi = 1 - \frac{\sum_{j=1}^{m} \left(\hat{y}_j - y_j\right)^2}{\sum_{j=1}^{m} \left(\overline{y}_j - y_j\right)^2} \tag{18}$$

Among them,$\overline{y}_j$ Represents the average of the true values of all samples.

3.3 Prediction Results

The prediction adopts a cross validation method of 50 times, which involves five cycles to obtain five test results. Finally, these five results are used as the average of the model's prediction results. To verify the feasibility and performance of each module, attention mechanism and LSTM model were removed from the complete model, and only LSTM model and traditional machine learning methods were used for prediction. Compare the prediction results after removing the LSTM model with the prediction results after removing the attention mechanism and the complete model, and evaluate the contribution of the LSTM model and attention mechanism to prediction accuracy. And compare the method studied in this article with traditional machine learning based performance prediction methods and deep learning based performance prediction methods to verify the prediction effect of this method. The specific results are shown in Figs. 2, 3 and 4 below.

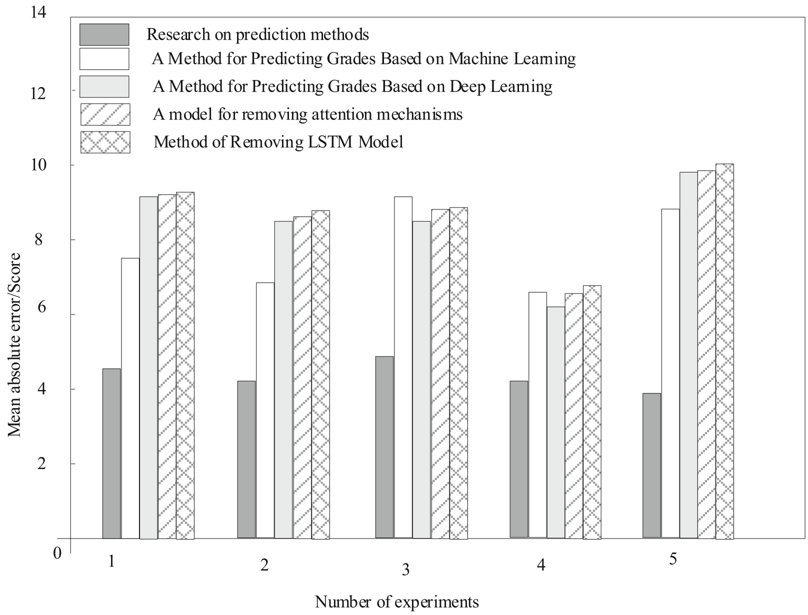

Fig. 2. Comparison Diagram of Average Absolute Error

From Figs. 2, 3 and 4, we can see that compared with LSTM model, traditional machine learning method, traditional machine learning based performance prediction method and deep learning based performance prediction method, the Mean absolute error of the research method is smaller, which indicates that the prediction method in this paper has higher accuracy in predicting student scores, and can more accurately predict student scores; The Root-mean-square deviation is small, which indicates that the difference between the prediction results of this method and the true value is small, and the prediction accuracy is higher; The larger Coefficient of determination indicates that the method in this paper can better explain the variance of student performance changes

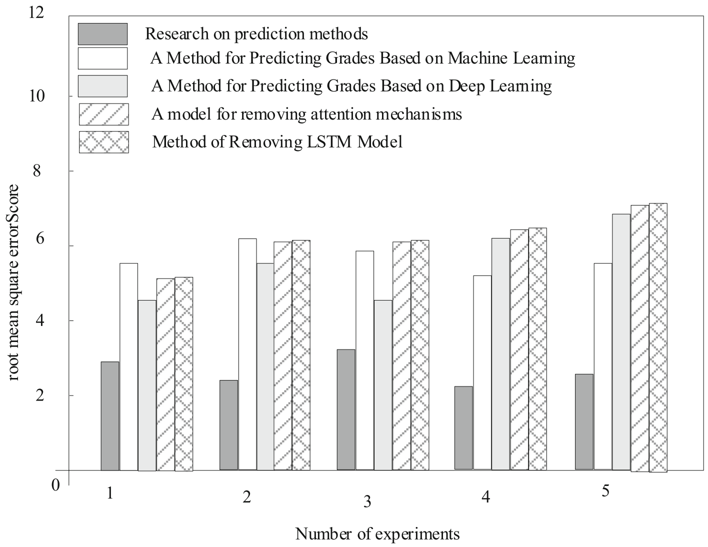

Fig. 3. Comparison Diagram of Root Mean Square Error

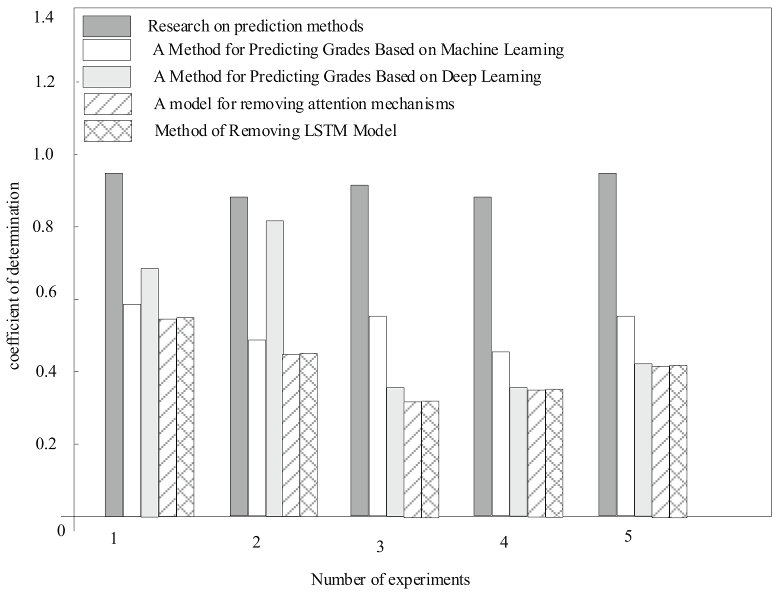

Fig. 4. Comparison Diagram of Determination Coefficient

and has a better fitting effect. This is because attention mechanisms can help models pay more attention to important information and improve the accuracy of predictions; The LSTM model can capture long-term dependencies in time series data, thereby improving the accuracy of prediction.

In order to verify the applicability and effectiveness of the method proposed in this article, three additional groups of 100 students' raw behavior data were randomly selected, labeled as dataset 1, dataset 2, and dataset 3 as the research objects. After processing the data, the prediction under the method proposed in this article was completed. The number of experiments is set to 5 cycles, and the predicted results are shown in Fig. 5.

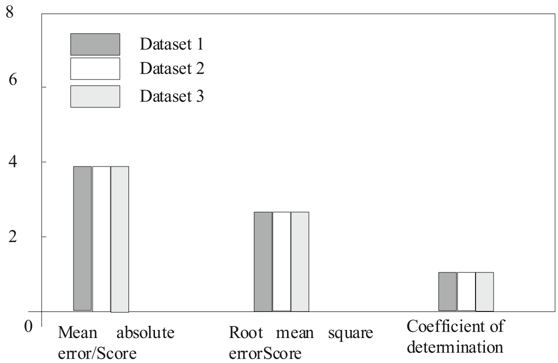

Fig. 5. Comparison of Prediction Results under Different Datasets

It can be seen from Fig. 5 that the results of Mean absolute error, Root-mean-square deviation and Coefficient of determination of experimental data sets 1, 2 and 3 are consistent after five cycles of testing. This indicates that the proposed method has consistent prediction performance and stability across different datasets. By verifying multiple datasets, the generalization ability and applicability of the method can be more reliably evaluated.

4 Conclusion

Student achievement prediction is one of the important research topics in the field of education. Accurate prediction results can help improve learners' academic performance and help managers make scientific decisions. It has important research significance and application value in the field of personalized service and adaptive learning. Through analysis, it is found that there are still some deficiencies in the existing performance prediction models in the use of various attribute characteristics and other factors to predict student performance. In view of this, this paper first proposes a data preprocessing method, and from the perspective that students are affected by different factors to different degrees, proposes a student performance prediction method based on the self attention mechanism to achieve effective prediction of student performance;Secondly, it analyzes the relationship between students' final grades and historical scores, and proposes a prediction method of students' output scores in online English teaching in higher vocational colleges based on the online platform of Xueyin, which makes more comprehensive and accurate use of various attribute features, thus further improving the prediction ability of the model;Finally, through the visualization of the attention results of each attribute feature on the final grade, the personalized analysis and guidance of students can be realized. This research also provides a new idea for students' performance prediction. The main work of the paper includes the following two aspects:

(1) In order to overcome the shortcomings of existing machine learning based performance prediction research methods under the online learning environment, this paper introduces an efficient performance prediction model. Compared with other traditional machine learning algorithms, this algorithm can be efficiently and accurately applied to the online learning behavior dataset of students with large amount of data, sparse features and high dimensions. Secondly, based on the experimental results of the constructed performance prediction model, combined with Pearson's phase.

    The relationship number together analyzes the impact of different learning behavior characteristics on students' online learning performance, so as to provide meaningful guidance for the construction and optimization of online learning.
(2) The self attention mechanism is introduced into LSTM to fully explore the relationship between attribute characteristics and their importance to the final grade, so as to give different attention weights to different attribute characteristics, and solve the problem that different factors have different degrees of influence on the same student and different students have different degrees of influence by the same factor.

This research not only solves the problem of individual differences ignored in the current research on the prediction of students' performance using various attribute characteristics, improves the accuracy of performance prediction and the interpretability of the model, but also realizes the personalized analysis and guidance for students, which also provides a new idea for solving the problem of performance prediction. However, there are still some deficiencies in the research, which need to be further improved:

(1) Only the relationship between the attribute characteristics and the final grade has been preliminarily excavated, and the combined characteristics may also be related to the final grade to some extent. Next, more consideration and design can be given to the combination of different characteristics or the influence of higher-order characteristics on the performance prediction results to improve the prediction accuracy of the model.
(2) The selected data set is relatively mature and the data volume is not large enough. In the next step, we can consider to carry out experimental verification on the constructed model in the data set with larger data volume. It is better to conduct further thinking and research after cleaning the actually collected data.
(3) When the system or indicators of the platform change, it is necessary to re collect and reorganize student behavior data, and conduct corresponding data processing and Feature engineering. In addition, it is necessary to adjust the parameter configuration of the prediction model based on new systems or indicators, and may even require retraining the model. Only after adapting to new changes in data and model adjustments can the accuracy and precision of the prediction method be ensured. Therefore, in practical applications, it is necessary to combine the collected data for research, establish a flexible process and mechanism, update data and models in a

timely manner, and analyze individual differences of students more accurately to maintain the effectiveness of prediction methods.

References

1. Yu, J.: Academic performance prediction method of online education using random forest algorithm and artificial intelligence methods. Int. J. Emerg. Technol. Learn. (iJET) **16**(5), 45 (2021)
2. Li, G.: Decision system for performance prediction based on machine learning technology in the context of intelligent education. J. Ningde Normal Univ. (Nat. Sci.) **33**(1), 36–41 (2021)
3. Zheng, A., Wang, Y., Hao, C.: Research on online academic achievement prediction with deep learning. Comput. Era (12), 69–72, 75 (2021)
4. Yekun, E.A., Haile, A.T.: Student performance prediction with optimum multilabel ensemble model. J. Intell. Syst. **30**(1), 511–523 (2021)
5. Maurya, L.S., Hussain, M.S., Singh, S.: Developing classifiers through machine learning algorithms for student placement prediction based on academic performance. Appl. Artif. Intell. **35**(3), 1–18 (2021)
6. Liang, K., Liu, J., Zhang, Y.: The effects of non-directional online behavior on students' learning performance: a user profile based analysis method. Future Internet **13**(8), 199 (2021)
7. Klingner, M., Fingscheidt, T.: Online performance prediction of perception DNNs by multi-task learning with depth estimation. IEEE Trans. Intell. Transp. Syst. **22**(7), 4670–4683 (2021)
8. Yong, H.E., Chung, Y., Min, S.P., et al.: Deep learning-based prediction method on performance change of air source heat pump system under frosting conditions. Energy **228**(1), 120542 (2021)
9. Lee, I., Sawant, A., Shin, J., et al.: Accurate gyrotron performance prediction based on full 3-d magnetic field and electron beam information altered by nonideal factors. IEEE Trans. Electron Devices **68**(3), 1276–1283 (2021)
10. Hao, M., Xiaoling, D., Shulong, W., Shijie, L., Jincheng, Z., Yue, H.: GaN JBS diode device performance prediction method based on neural network. Micromachines **14**(1), 188 (2023)
11. Chen, W., He, W., Zhu, H., Zhou, G., Mu, Q., Han, P.: A processor performance prediction method based on interpretable hierarchical belief rule base and sensitivity analysis. Comput. Mater. Continua **74**(3), 6119–6143 (2022)
12. Rossopoulos Georgios, N., Papadopoulos, C.I.: A journal bearing performance prediction method utilizing a machine learning technique. Proc. Inst. Mech. Engineers Part J: J. Eng. Tribol. **236**(10), 1993–2003 (2022)
13. Luo, Y., Han, X.: A model for predicting student performance in hybrid courses based on incremental learning algorithm. E-educ. Res. **42**(7), 83–90 (2021)
14. Wei, X.: The evaluation and prediction of academic performance based on artificial intelligence and LSTM. Chin. J. ICT Educ. **28**(4), 123–128 (2022)
15. Zhang, B., Chen, G., Chen, X., Lu, H.: Hardware implementation of tanh function based on second-order approximation and error compensation. Microelectronics **51**(6), 905–909 (2021)