

Research on the Academic Early Warning Model of Distance Education based on Student Behavior Data in the Context of COVID-19

Early Warning Model of Distance Education

Yi Qu^{1*}

The Open University of Shaanxi
Sino Swiss School of Tourism and
Hospitality, Xi'an 710119, China

Zhiyuan Sun²

The Open University of Shaanxi
School of Marxism
Xi'an 710119, China

Libin Liu³

The Open University of Shaanxi
School of Basic Subjects
Xi'an 710119, China

Abstract—The COVID-19 epidemic has caused great impact on the entire society, and the spread of novel coronavirus has brought a lot of inconvenience to the education industry. To ensure the sustainability of education, distance education plays a significant role. During the process of distance education, it is necessary to examine the learning situation of students. This study proposes an academic early warning model based on long- and short-term memory (LSTM), which firstly extracts and classifies students' behavior data, and then uses the optimized LSTM to establish an academic early warning model. The precision rate of the optimized LSTM algorithm is 0.929, the recall rate is 0.917 and the F value is 0.923, showing a higher degree of convergence than the basic LSTM algorithm. In the actual case analysis, the accuracy rate of the academic early warning system is 92.5%. The LSTM neural network shows high performance after parameter optimization, and the academic early warning model based on LSTM also has high accuracy in the actual case analysis, which proves the feasibility of the established academic early warning model.

Keywords—COVID-19; Student behavior data; Distance education; Academic early warning model

I. INTRODUCTION

Due to the impact of the COVID-19, teachers' teaching methods and students' learning models in many regions have been transformed into distance education [1]. The online learning mode of distance education involves computer and other technologies. This new teaching mode requires students and teachers to change traditional teaching methods and learning concepts, and adapt to online education and learning mode through new teaching and learning practice approaches [2]. Distance education ensures the continuity of education and provides an effective way for students to learn professional knowledge [3]. Although distance education provides students with a new way of learning, there are obvious problems in online education, which lead to the poor effect of online education. First of all, in online education, students lack the learning atmosphere of offline teaching, and teachers are difficult to manage students' learning behavior, so they generally show a lack of learning initiative. Some research shows that keeping the classroom active during online learning is conducive to improving students' academic performance [4].

In addition, in the survey of distance education for students, it is found that the learning level and self-confidence of students in online education environment are lower than those in offline environment. [5]. This has brought great challenges for students to successfully complete their homework. In order to solve these problems and improve the teaching effect of online teaching, this research uses artificial intelligence technology to establish a learning management system to evaluate students' academic achievements and homework [6]. A reasonable evaluation method can improve students' confidence in learning. An effective evaluation method will strengthen the management effect of teachers on students. The research establishes the academic early warning model of online learning platform using neural network technology, and confirms the validity of the model through experiments. The model can give early warning to students with poor learning conditions with high accuracy. The research is of positive significance to the improvement of academic performance and other comprehensive abilities, and it can help students successfully complete their studies.

II. LITERATURE REVIEW

Early warning systems have been effectively used in medical, ecological and other fields. Ginestra et al. established a sepsis early warning system using machine learning to help nurses and patients predict sepsis symptoms and severe septic shock. Experiments have shown that the early warning system can effectively identify alerts for sepsis symptoms, helping nurses and patients to detect sepsis symptoms in advance [7]. The academic early warning system plays an important role in the analysis of students' learning situation. Through the academic early warning system, the early warning of students' problems in their learning can be realized, and intervention policies and measures can be proposed based on the obtained information and data to help students successfully complete their studies [8]. In the existing literatures, researchers have investigated students' learning motivation and used the learning analysis dashboard to analyze students' learning motivation and academic performance. By taking e students' academic performance as a reference, it is possible to intuitively analyze students' academic performance through the early academic early warning system. However, the early academic warning

*Corresponding Author.

system has a poor degree of automation, cumbersome operation and low accuracy, so it has not been popularized in practical applications [9]. The application of machine learning technology in the early warning system can effectively improve the accuracy of early warning, and realize automatic monitoring and rapid response. Cho optimized and improved the early warning system on the basis of deep learning. The optimized early warning system had a higher accuracy rate and recall rate, which can be used to accurately judge patients' condition [10]. Lin et al. established a bridge scouring early warning system composed of the Internet of things and artificial intelligence. The researchers encapsulated the water level sensor in a stainless-steel ball. The early warning system uses neural networks for deep learning and real-time recording according to the changes of water level [11]. In the application of students' academic early warning system, Rashid et al. used a hybrid system including recurrent neural network and gray wolf optimizer to identify students' weaknesses, predict students' academic performance, and put forward corresponding suggestions. This improved academic early warning system can help teachers to improve the teaching model and improve the learning experience of students at the same time. Although this model has good performance, it is the same as most such systems. It mainly serves offline teaching, and the application effect in online teaching is not good [12]. In general, the academic early warning system has been applied to a certain extent at present, but most of them serve offline, and the accuracy and timeliness need to be improved.

LSTM is a neural network model, which belongs to the field of machine learning. It can effectively save and analyze time series information with strong dependence on time. Among various forecasting models, LSTM-based hybrid models can be used for forecasting foreign exchange fluctuations. Researchers have used LSTM technology to optimize the forecasting model, and the experimental results show that the method combined with LSTM has a high accuracy in foreign exchange volatility forecasting [13]. LSTM combines the convolutional neural network, retains the advantages of the LSTM method itself and time-frequency information, which can be used for the prediction of remaining lifespan and help solve health management problems [14]. In the early warning of industrial systems, Dorgo et al. used LSTM to predict rare operations in chemical technology when processing the information obtained by data mining, and at the same time analyzed time series information to predict the possible consequences of rare operations [15]. Based on its excellent analytical ability, LSTM is also used in the field of education. There are many courses in online education, and the information is miscellaneous. When students face many course data, it is difficult to accurately locate the courses. The researchers proposed a learning framework based on LSTM, and introduced attention mechanism on the basis of LSTM to predict and recommend courses according to students' behavior data [16].

The LSTM neural network retains time-dependent time series information, which can optimize the model in the establishment of early warning models. It is widely used in the establishment of early warning systems in medical, geological, economic and other fields. Students' performance and relevant data in the online learning process show obvious time series

nature, so LSTM network is a more suitable analysis tool. However, after the discussion of literature review, it is found that the application of LSTM in distance education academic early warning model is less. Therefore, this study proposes to use LSTM neural network to analyze students' behavior data, and establishes an academic early warning model for students' learning evaluation in online education.

III. RESEARCH ON THE ACADEMIC EARLY WARNING MODEL OF DISTANCE EDUCATION BASED ON LSTM NEURAL NETWORK

A. Academic Early Warning Model based on LSTM Neural Network

The information is input to the signal X_t in the conventional recurrent neural network and output as h_t , and the input and output of the information are continuously circulated to realize the preservation of the information. However, the recurrent neural network does not have the ability to process long time series. When the distance between the input and the output is large, the output quality of the network will be reduced. LSTM can improve the capacity of the cyclic neural network in time series processing, and store long-term series information by increasing the complexity of the cyclic neural network structure [17]. The standard LSTM consists of three parts: input gate, output gate and forget gate, as shown in Fig. 1.

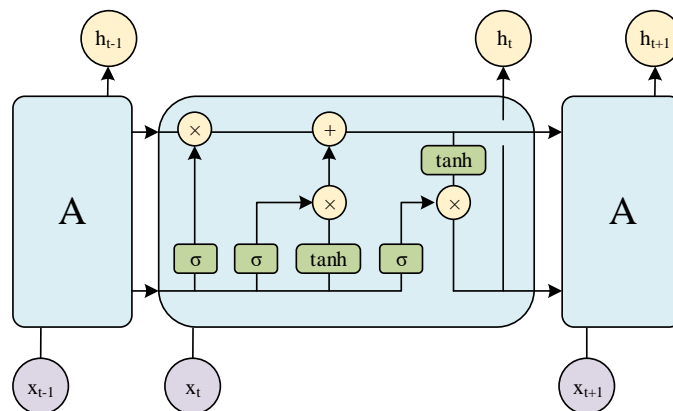


Fig. 1. LSTM Neural Network.

Although LSTM overcomes the gradient loss problem of traditional recurrent neural network, it still faces the problem of gradient decrease when processing longer-period time series [18]. LSTM has a large amount of computation and a deeper network, which takes more time in computation and training [19]. Combining the superiority of LSTM algorithm, this paper presents an academic early warning model of distance education based on student behavior data, and optimizes the method for the problems existing in LSTM. First, collect and clean the relevant distance education data of students, and organize the cleaned data into the format required for input. Second, high-order mapping of data is performed, so that the impact of the experiment on the neatness of the data is reduced. Next, feature extraction is performed according to the data characteristics of students, and classification and training are performed. Then, the long- and short-term neural network input

incentives are optimized, and the results of the optimized network output are cascaded and fused so as to improve the accuracy of the output. Fig. 2 shows the schematic diagram.

To reasonably distribute the attention of administrators to students' studies and facilitate students' self-awareness of their own studies, the academic warnings are divided into green, yellow, orange and red warnings according to the actual situation. Different early warning methods are used for students at different levels in the early warning system, and factors that actually affect learning, such as family circumstances, are fully considered in the early warning process.

B. Research on Feature Extraction and Classification Methods

The deepening of datafication has reformed students' learning methods, and distance education has been widely used in the context of COVID-19. It is necessary to mine data information such as students' achievements in distance education. Based on the distance education network database, the scores of each subject of the students are collected in the cloud. The score table of each student is generated using Excel, and each student is abstracted into a class containing variables such as score information and corresponding credits. See formulas (1) and (2) for the calculation formulas of the average score and the weighted average score of the students' original grades.

$$M = \frac{a_1 + a_2 + a_3 + \dots + a_n}{n} \tag{1}$$

$$B = \frac{a_1m_1 + a_2m_2 + a_3m_3 + \dots + a_nm_n}{m_1 + m_2 + m_3 + \dots + m_n} \tag{2}$$

Where a represents the score of the corresponding subject, the number of subjects is n , and the weight value of the subject

score is m . The variance of the student's grade can be obtained through the above formula, as shown in formula (3).

$$S = \frac{\sum (a_i - M)^2}{n}, i = 0, 1, \dots, n \tag{3}$$

The calculated average scores and weighted average scores of the corresponding subjects of the students are added to the abstract class. The grades and credits corresponding to each subject of the students are saved as float data, and the average scores and weighted average scores are saved as double data. In distance education, in addition to collecting students' subject scores for academic evaluation, students' online course learning information and other information are also collected for academic evaluation. Amplify the corresponding variables in the formed class to generate a 9*N source data matrix. However, information loss is easy to occur during data collection. In this experiment, the RBF kernel function is used to process the data information. In this method, low-dimensional data is mapped to generate high-dimensional data, and high-dimensional vector operations are maintained, so as to solve the problem of accuracy decline caused by missing data. Using data mapping can ignore the problem of linear inseparability, while preserving the integrity of the data. In the source data matrix, the information vector of each row of students is mapped to the high-dimensional space using formula (4) to generate the corresponding vector label.

$$k(a', b') = (\phi(a'), \phi(b')) = e^{-ha^2} \begin{bmatrix} 1 \\ \sqrt{\frac{2h}{1!}} a' \\ \sqrt{\frac{(2h)^2}{2!}} a'^2 \\ \vdots \end{bmatrix} e^{-hb^2} \begin{bmatrix} 1 \\ \sqrt{\frac{2h}{1!}} b' \\ \sqrt{\frac{(2h)^2}{2!}} b'^2 \\ \vdots \end{bmatrix} \tag{4}$$

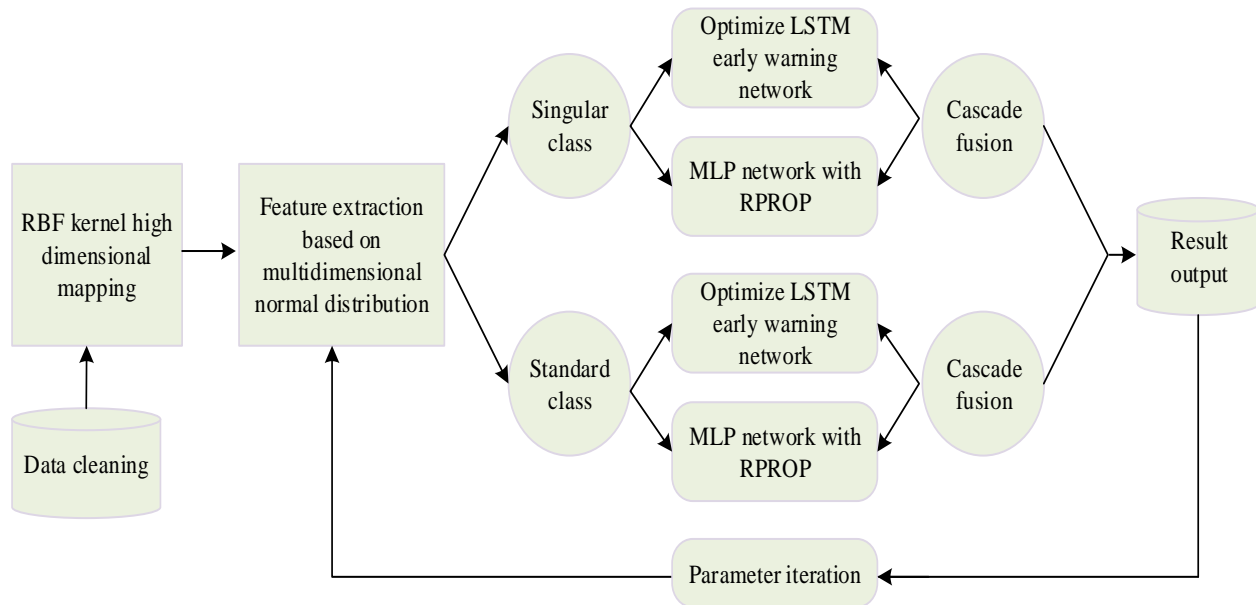


Fig. 2. Early Warning Model based on LSTM.

Where, $a^i = A_i, (i = 1, 2, \dots, n)$ and $b^i = B_i, (i = 1, 2, \dots, n)$ represents the student information vector represented by each row in the source data matrix, and h is a constant value. The representation method of elements in high-dimensional space is shown in formula (5).

$$e_n(a) = \sqrt{\frac{(2\alpha)^2}{n!}} a^n e^{-\alpha a^2}, n = 0, 1, 2, \dots \quad (5)$$

Where, the kernel parameter gamma value of the RBF kernel function is expressed as α . Through function mapping, the vacant data information in the source data matrix can be filled to obtain a full data matrix. In the evaluation of academic tests, the normal distribution can be used to grade the outstanding and backward situations, so as to determine the corresponding number of people, score lines, and winning rates. When the students' grades conform to the normal distribution, the RBF function in formula (6) is used to map the grade information to the high-dimensional space for feature extraction of the multi-dimensional normal distribution, and the students who meet the multi-dimensional normal distribution are divided into standard classes. Students who meet the multidimensional normal distribution are classified as singular.

$$f_a(a_1, \dots, a_k) = \frac{1}{\sqrt{(2\pi)^k |\sum a_i|}} \exp(-\frac{1}{2} (a - \mu)^T \sum a_i (a - \mu)) \quad (6)$$

Where, the mean value is expressed as μ , k is a constant. In the classification based on multi-dimensional normal distribution, data feedback is used to adjust the parameters of multi-dimensional normal distribution, as shown in Fig. 3.

The selection of the threshold in the normal distribution has an important impact on the accuracy of network training results. The optimal solution of the threshold parameter is selected by the parameter iteration method of gradient descent. After repeated attempts of the gradient descent method, the global optimal solution μ of the average threshold can be obtained, which is then normalized by formula (7).

$$X = (Y - \mu) / (Max - Min) \quad (7)$$

Where, Y represents the matrix that needs to be normalized, X represents the matrix obtained after normalization, and Max, Min represent the maximum and minimum vector values of all elements in X , respectively.

C. Research on Early Warning Algorithm based on Improved LSTM

LSTM performs well in processing long time series information. In distance education, the information is recorded as a long time series, which shows a strong dependence on time. The modelling of Multilayer perceptron (MLP) network is easy to realize, and neurons can be constructed according to the actual situation [20]. In this experiment, the MLP network is used to train students' grades. The MLP-based LSTM network uses time to divide the length of information in different dimensions, which reduces the complexity and polymorphism of information on the basis of retaining data features. The other data is back-propagated using the function to increase the speed at which the network operates. At the same time, the LSTM neural network is used to train other data, and the weighting function is introduced to solve the problems of gradient disappearance and gradient explosion. Then, cascade fusion is used to train students' behavioral data to further improve the accuracy of prediction results and solve the problem of academic early warning.

By using multi-dimensional normal distribution to extract the features from student data matrix, two matrices of singular class and standard class can be obtained. The data with high time-series dependence is more sensitive to time. The LSTM with adaptive excitation function is used to train the data with high time-series dependence. The schematic diagram of the neural network unit is shown in Fig. 4. The sum function is introduced into the adaptive excitation function, and the expression of **relu** and **tan** is shown in formula (8).

$$z = \frac{1}{v} \text{relu}(w \frac{x^t}{h^{t-1}}) + \frac{1}{u} \tanh(w \frac{x^t}{h^{t-1}}) \quad (8)$$

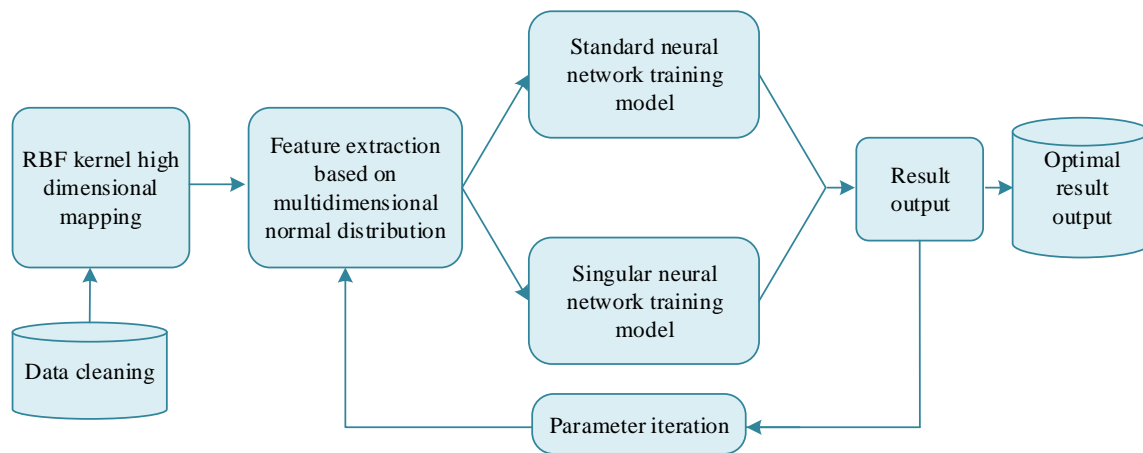


Fig. 3. Feedback-based Parameter Adjustment Process.

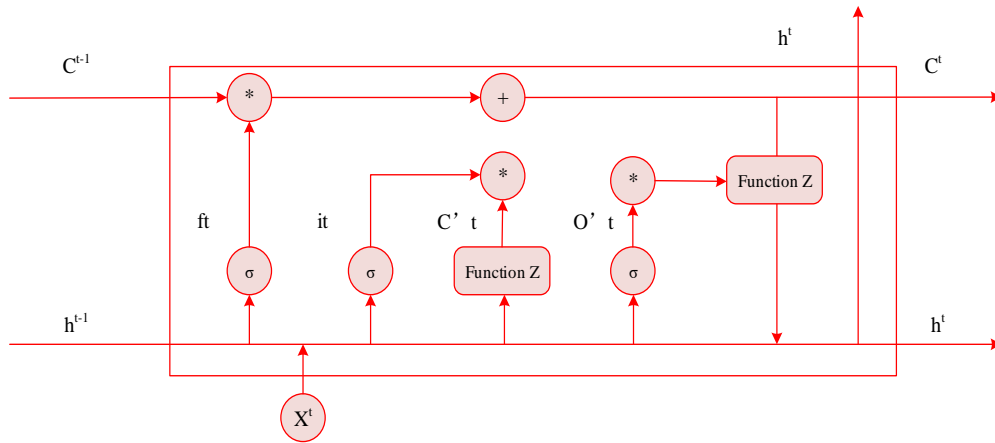


Fig. 4. LSTM Unit with Adaptive Excitation Function.

Where, W represents the function vector, v and u are the weight coefficient, $v+u=1$. The input module of LSTM includes the previous round of information carrier c^{t-1} , the previous round of state quantity output h^{t-1} and the current round of input x^t ; the output module includes the current round of information carrier c^t , the current round of state quantity output h^t and the current round of output y^t . Inputting \tanh function yields a vector z , and then, the state quantity output h^{t-1} of the previous round and the input x^t of the current round are combined into a vector, which is multiplied by W and then by the forgetting gate's matrix weight W^f , the input gate's matrix weight W^i , and the output gate's matrix weight W^o respectively, thus obtaining the output z^f and z^i sum z^o of each door. Thus, the calculation formulas of the current round information carrier, the current round state quantity output and the current round output can be obtained, as shown in formulas (9) to (11).

$$c^t = z^f c^{t-1} + z^i z \quad (9)$$

$$h^t = z^o * \left(\frac{1}{v} \operatorname{relu}\left(w \frac{x^t}{h^{t-1}}\right) + \tanh\left(w \frac{x^t}{h^{t-1}}\right) \right) * c^t \quad (10)$$

$$y^t = \sigma(wh^t) \quad (11)$$

The adaptive excitation function is to obtain the weighted average function after the function is weighted. The adaptive excitation function in formula (12) is used to excite the incoming data in each gate, and the data with high time series dependence is trained. Academic warnings are classified as follow.

$$f(x) = \frac{1}{u} \frac{e^x - e^{-x}}{e^x + e^{-x}} + \frac{1}{v} \max(0, x), u + v = 1 \quad (12)$$

Use formula (8) to iteratively calculate the values of weight coefficients v and u in the adaptive excitation function. When $v=0.643$ and $u=0.357$, the optimal data training results are obtained. RProp is a resilient backpropagation algorithm, and the MLP network with RProp can be used to train data with low time-series dependence such as the scores of various subjects. First, assign an initial value to the weight change and determine its acceleration and deceleration factors. Then, according to the feedback that the error gradient sign change or does not change during the iteration process, the training is decelerated or accelerated respectively to ensure the stable convergence of the system. Finally, backpropagation is performed by combining the sign of the error gradient with the changing step size. When the sign of the error gradient is negative, the weight needs to be appropriately increased to obtain the minimum value, as shown in formula (13).

$$\begin{aligned} \Delta w_{i,j} &= -\eta_{i,j}(t), \text{if } g(y) > 0 \\ \Delta w_{i,j} &= +\eta_{i,j}(t), \text{if } g(y) < 0 \\ \Delta w_{i,j} &= 0, \text{otherwise} \end{aligned} \quad (13)$$

Where, $\eta_{i,j}(t)$ denotes the learning rate of the algorithm. When the directions of the error gradient functions of two adjacent moments are inconsistent, it indicates that the minimum value of the current moment is missed, and the learning rate $\eta_{i,j}(t)$ of the current moment is smaller than the learning rate $\eta_{i,j}(t-1)$ of the previous moment, then multiply $\eta_{i,j}(t-1)$ by η^{up} in the range (0, 1) to obtain the current learning rate. When the directions of the error gradient functions of two adjacent moments are consistent, it indicates that the minimum value of the error function has not been reached. At this time, multiply $\eta_{i,j}(t-1)$ by η^{down} with a range greater than 1 to obtain the current learning rate, as shown in the formula (14).

$$\begin{aligned}\eta_{i,j}(t) &= \eta^{up} \eta_{i,j}(t-1), g(t-1)g(t) > 0 \\ \eta_{i,j}(t) &= \eta^{down} \eta_{i,j}(t-1), g(t-1)g(t) < 0 \\ \eta_{i,j}(t) &= \eta_{i,j}(t-1), otherwise\end{aligned}\quad (14)$$

The neural network system performs classification operations, and then combines the output results and uses formula (15) to fuse the results, so as to present different academic warnings for students.

$$p * a + (1 - p) * b = c \quad (15)$$

Where, P is the weight coefficient, a represents the LSTM prediction result, b represents the MLP prediction result, and the c represents graduation probability.

IV. SIMULATION ANALYSIS OF ACADEMIC EARLY WARNING MODEL BASED ON LSTM

The construction and training of the academic early warning system model of distance education is the key to the development of the system, and the accuracy and running time of the system are affected by the model construction and training results. Determining the number of layers and the number of nodes in the neural network is the main step that affects the construction of the system. In the experiment, the root means square error (RMSE) and the mean absolute error (MAE) are used to verify the number of layers in the neural network. The LSTM-based academic early warning model is repeatedly trained in the training set, and the parameters of the model are continuously adjusted in the process. Fig. 5 shows the training trend of RMSE and MAE values when different models predict time-dependent time series information. The abscissa represents the number of iterations of model training, and LSTM-1 to LSTM-5 represent the number of layers in the neural network, which are 1 to 5, respectively.

In Fig. 5(a), the number of iterations for LSTM-1 and LSTM-2 to start to converge is about 200. The minimum RMSE value of LSTM-1 with 1 neural network layer is 31.84, and the minimum RMSE of LSTM-2 with two neural network layers is 31.84. When the number of layers of the neural network increases to three, the RMSE value decreases more significantly. When the number of iterations is 225, LSTM-3 obtains the lowest RMSE value of 20.17, and the RMSE value remains at this level in subsequent iterations of training. When the number of layers of the neural network increases to four and five layers, the RMSE value has no obvious downward trend. In Fig. 5(b), the change trends of MAE value and RMSE value are consistent. The number of iterations when LSTM-1 and LSTM-2 begin to converge is about 200, the minimum MAE

value of LSTM-1 is 22.91, and the minimum MAE value of LSTM-2 is 18.96. When the number of layers of the neural network increases to three, the downward trend of the MAE value is more significant. When the number of iterations is 225, LSTM-3 obtains the lowest MAE value of 15.17, and the MAE value remains at this level in subsequent iterations of training. When the number of layers of the neural network increases to four and five layers, the MAE value has no obvious downward trend. It can be seen from Fig. 5 that in the training data set, the lowest error values RMSE and MAE are obtained when the number of layers of the LSTM neural network is set to 3 layers. When the number of layers is one or five, underfitting or overfitting may be encountered.

When the LSTM neural network is embedded in the academic early warning model, the embedding size is also one of the important factors affecting the performance of the early warning model. In the experiment, the number of layers of the neural network is set to three, and each neural network layer has different embedding sizes. Fig. 6 shows the training variations of the RMSE and MAE values of the LSTM neural network layers under different embedding sizes. In Fig. 6, LSTM-100 represents a three-layer neural network with an embedding size of 500-250-100, LSTM-150 is 500-300-150, LSTM-200 is 500-350-200, and LSTM-250 is 500-400-250, 500-400-300 for LSTM-300. The number of iterations at which the LSTM-100 and LSTM-150 models start to converge in Fig. 6(a) are 225 and 200, respectively, and the minimum RMSE values are 30.89 and 24.76, respectively. The RMSE value converges faster in the LSTM-200 model, reaching the lowest value of 20.68 after 250 iterations. In LSTM-250 and LSTM-300, there is basically no downward trend of the RMSE value, and the performance of the model is no longer improved. In Fig. 6(b), the MAE value and the RMSE value have the same change trend. The number of iterations at which the LSTM-100 and LSTM-150 models start to converge are 200 and 225, respectively, and the minimum MAE values are 20.03 and 14.98, respectively. The MAE value converges faster in the LSTM-200 model, reaching the lowest value of 13.09 after 200 iterations. When the embedding size increases, there is basically no downward trend of the MAE value. When the LSTM neural network is embedded in the academic early warning model, the size of the embedding size is proportional to the predictive ability of the model. The larger the embedding size, the stronger the data fitting ability. Fig. 6 shows that when the embedding size is 200, the minimum RMSE and MAE are obtained. Combined with the experimental results of the algorithm in different configurations, and considering the complexity and prediction effect, this study selects the embedded size of 200 and the number of neural network layers of three as the default settings for subsequent verification experiments.

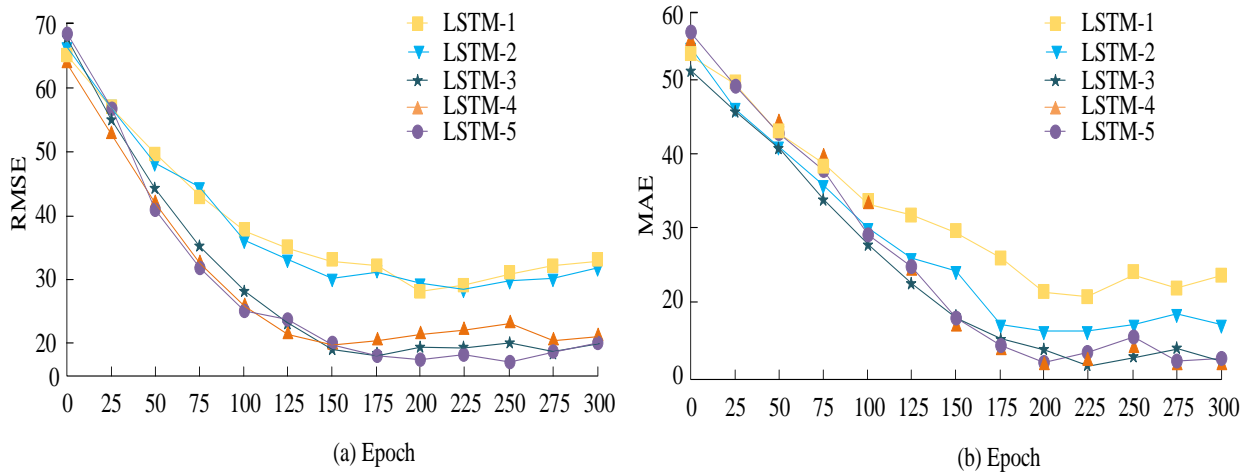


Fig. 5. RMSE and MAE Trends of Graph Neural Network with 1-5 Layers on Training Set.

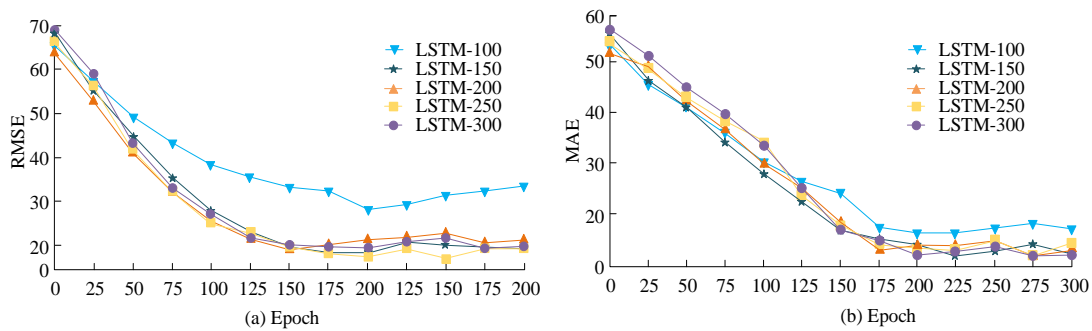


Fig. 6. Change Trends of RMSE and MAE under different Embedding Sizes.

TABLE I. COMPARISON OF EXPERIMENTAL RESULTS OF TIME SERIES STATE PERCEPTION

Algorithm	P	R	F
HC	0.673	0.622	0.646
HN	0.612	0.530	0.568
HM	0.405	0.492	0.444
LC	0.831	0.806	0.818
LN	0.678	0.713	0.695
LM	0.462	0.562	0.507
LSTM	0.886	0.812	0.847
This paper	0.929	0.917	0.923

The basic LSTM algorithm and the optimized LSTM algorithm proposed in this experiment are comparatively analyzed in terms of accuracy (P), recall (R) and the ratio F. Table I shows the experimental results of the HTM algorithm, the basic LSTM algorithm and the optimization algorithm proposed in this experiment in time series state perception. In Table I, H represents the HTM algorithm, L represents the basic LSTM algorithm, and LSTM represents the prediction model based on the basic LSTM algorithm; C represents the system parameter CPU, M represents the system parameter MEM, and N represents the system parameter NET. The sensitivity of the algorithm to abnormal situations can be judged by the accuracy rate, the detection ability of the algorithm to abnormal situations can be judged by the recall rate, and the F value is used to evaluate the overall performance of the algorithm. The

results in Table I show that the accuracy of the optimized LSTM algorithm is 0.929, the recall is 0.917 and the F value is 0.923, which are significantly higher than those of the basic LSTM algorithm. The comparative experimental results show that the accuracy rate of the optimized algorithm is higher, and the probability of false positives of the model is significantly reduced; the recall rate of the optimized algorithm is outstanding, and the probability of false negatives of the model is significantly reduced; the accuracy rate and recall rate of the optimized algorithm are higher, indicating that the overall performance of the optimized algorithm is higher. Based on the experimental parameters set above, the convergence rates of the basic algorithm and the optimized algorithm are compared.

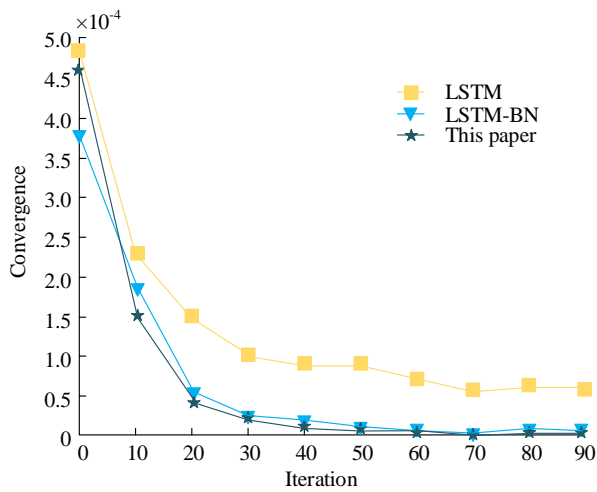


Fig. 7. Comparative Analysis of Convergence Rate.

Whether the model satisfies the convergence condition is mainly judged by comparison with the preset value, judgment of the weight value and the comparison of the maximum number of iterations. The purpose of training the model is to reduce the loss value, and it is necessary to set a threshold in advance to determine whether the training is over. The preset threshold will affect the convergence of the neural network model. When the loss value is less than the preset value, it is considered to end the training. When there is no significant change in the weights between two adjacent iterations, it can be considered that the model has reached a state of convergence with no need to adjust the weights. There are many preset applications for the maximum number of iterations. The iteration period is selected according to the actual situation, and the model is trained to achieve the optimal solution. Fig. 7 shows the comparison results of the convergence rate of the basic method and the optimized method in this experiment. It can be clearly seen that the convergence effect of the optimized method is significantly better than that of the basic method. 80

college students in distance education were randomly selected for the comparative evaluation of the academic early warning system, and the results obtained by the academic early warning prediction model based on the optimized LSTM algorithm were compared with the actual situation.

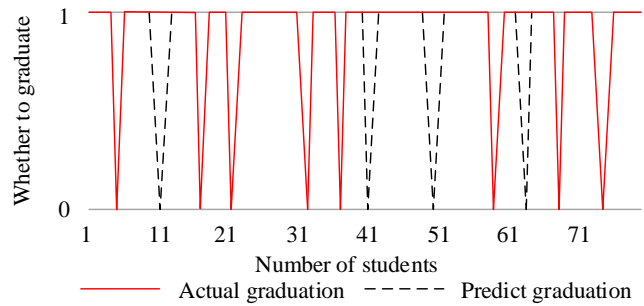


Fig. 8. Comparison of the Actual Results and the Results Predicted by the Academic Early Warning Prediction Model based on Optimized LSTM Algorithm

In Fig. 8, the abscissa represents the number of students, 0 in the ordinate represents ungraduated, and 1 represents graduation. The figure shows that among the 80 college students, six are inconsistent with the actual number of graduates, and the accuracy rate of the academic warning system is 92.5%. In addition, the academic early warning system is used to predict students' enthusiasm for active learning in distance education, and the predicted value is compared with the real situation. Fig. 9 shows the results.

Fig. 9 shows that during distance education, the number of submissions in the academic early warning system is consistent with the actual trend, reflecting that prediction results of learning enthusiasm by the academic early warning system are consistent with the actual situation. The academic early warning system in distance education designed in this experiment has high prediction accuracy and can be applied to students' academic early warning.

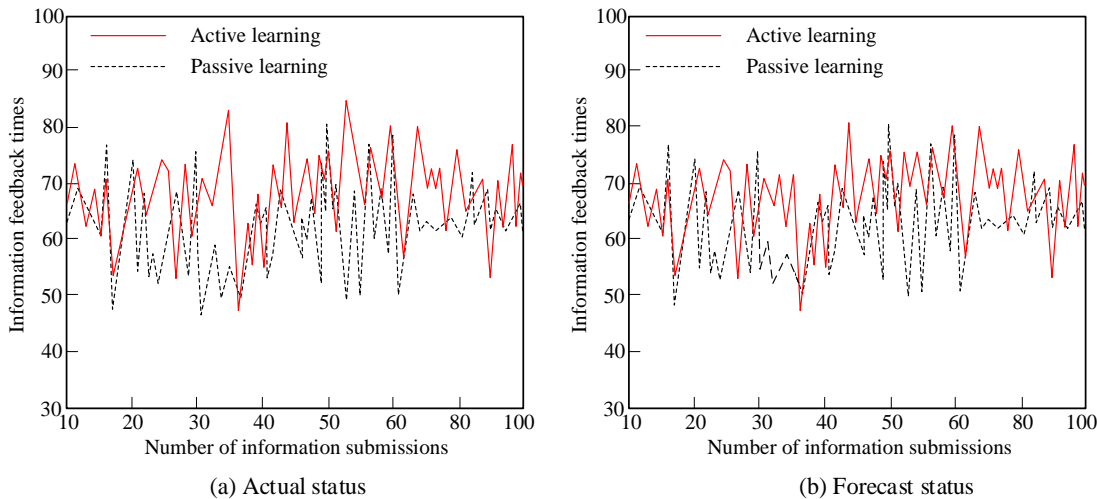


Fig. 9. Interactive Comparison Results of the Two Systems.

V. DISCUSSION

At present, the proportion of middle line education in the whole education industry is gradually increasing. In the online education environment, students' learning behavior is difficult to be monitored. And according to the information in literature [4] and [5], students under online education generally lack initiative and learning self-confidence, and their overall performance is lower than that of offline teaching. The early warning system is mainly a system to detect signs and warn relevant personnel at the early stage of or before the beginning of a negative situation. The quality of current academic early warning systems is uneven. The early warning system in literature [9] has been applied in some colleges and universities, but its operation has a low degree of automation, and its accuracy also has a large room for improvement. The early warning system in [12] and the proposed early warning model both use neural network models, but the performance of the latter is slightly higher than that of the former. In addition, the early warning model in literature [9] and [12] pays more attention to the academic early warning of offline teaching. When the teaching scenario is online, its performance will be further limited. The early-warning model proposed in this study is specially designed for online teaching, which has great application potential in the current social context. From the results of the performance test, the model proposed in this study also has a high degree of early warning accuracy and automation, which can be applied to online teaching scenarios.

VI. CONCLUSION

In the simulation analysis experiment of the academic early warning model based on LSTM, the number of layers of the neural network is set to three and the embedding size is set to 200 through parameter optimization and comprehensive consideration. The accuracy rate of the optimized LSTM algorithm is 0.929, the recall rate is 0.917 and the F value is 0.923, which are significantly higher than those of the basic LSTM algorithm, which has the accuracy rate of 0.886, the recall rate of 0.812 and the F value of 0.847. Comparing the results obtained by the academic early warning prediction model based on the optimized LSTM algorithm with the actual situation, the accuracy of the academic early warning system is 92.5%. To predict the enthusiasm of students for active learning, the number of submissions in the academic early warning system is consistent with the actual trend. The LSTM-based academic early warning model proposed in this experiment has high performance after parameter optimization, and also has high accuracy in actual case analysis, which can be applied to academic early warning in distance education. This research discusses the application of LSTM network in online teaching evaluation, and broadens the application field of LSTM network. At the same time, the research results have brought a practical model to the field of online education, which can improve the teaching effect through accurate and timely warning, and help students learn and teachers teach. However, the running time of the model was not verified, and it was not compared with other existing models in terms of the running time. Therefore, the proposed model may not be the optimal solution in terms of time, and there is still room for further improvement. In the follow-up work, we need to further improve the relevant experimental content.

VII. FINDINGS

The research is supported by: Innovation Team of Digital Communication and Cultural Soft Power (Grant No. TD2021002); Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No. 22JZ0016).

REFERENCES

- [1] K. Cicha, M. Rizun, P. Rutecka, et al. "COVID-19 and Higher Education: First-Year Students' Expectations toward Distance Learning", *Sustainability*, Vol. 13, pp. 1889-1908, 2021. DOI:10.3390/su13041889.
- [2] J. A. Vilchez, J. Kruse, M. Puffer, et al. "Teachers and School Health Leaders' Perspectives on Distance Learning Physical Education During the COVID-19 Pandemic", *Journal of School Health*, vol. 91, pp. 541-549, 2021. DOI: 10.1111/josh.13030.
- [3] V. Domenici, "A Course of History of Chemistry and Chemical Education Completely Delivered in Distance Education Mode during Epidemic COVID-19", *Journal of Chemical Education*, 2020, vol. 97, pp. 2905-2908, 2020. DOI:10.1021/acs.jchemed.0c00739.
- [4] J. E. Baker, "Maintaining an Active Organic Class during the COVID-Induced Online Transition at Two Undergraduate Institutions", *Journal of Chemical Education*, vol. 97, pp. 3235-3239, 2020. DOI:10.1021/acs.jchemed.0c00759.
- [5] J. Wilhelm, S. Mattingly, V. H. Gonzalez, "Perceptions, Satisfaction, and Performance of Undergraduate Students During Covid-19 Emergency Remote Teaching", *Anatomical Sciences Education*, 2021, vol. 15, pp. 42-56, 2021. DOI: 10.1002/asc.2161.
- [6] B. J. Santiago, J. M. O. Ramírez, J. Rodríguez-Reséndiz, et al. "Learning Management System-Based Evaluation to Determine Academic Efficiency Performance", *Sustainability*, vol. 12, pp. 4256-4272, 2020. DOI:10.7763/IJEEEE.2013.V3.233.
- [7] J. C. Ginestra, H. M. Giannini, W. D. Schweickert, et al. "Clinician Perception of a Machine Learning-Based Early Warning System Designed to Predict Severe Sepsis and Septic Shock", *Critical Care Medicine*, vol. 47, pp. 1477-1484, 2019. DOI: 10.1097/CCM.0000000000003803.
- [8] A. S. Sein, M. Daniel, A. Fleming, et al. "Identifying and Supporting Students to Prevent USMLE Step 1 Failures When Testing Follows Clerkships: Insights From 9 Schools", *Academic Medicine*, vol. 95, pp. 1338-1345, 2020. DOI: 10.1097/ACM.0000000000003272.
- [9] S. J. Aguilar, S. A. Karabenick, S. D. Teasley, et al. "Associations between learning analytics dashboard exposure and motivation and self-regulated learning", *Computers & Education*, vol. 162, pp. 104085, 2021. DOI:10.1016/j.compedu.2020.104085.
- [10] K. J. Cho, O. Kwon, J. M. Kwon, et al. "Detecting Patient Deterioration Using Artificial Intelligence in a Rapid Response System", *Critical Care Medicine*, vol. 48, pp. 285-289, 2020. DOI: 10.1097/CCM.0000000000004236.
- [11] Y. B. Lin, F. Z. Lee, K. C. Chang, et al. "The Artificial Intelligence of Things Sensing System of Real-Time Bridge Scour Monitoring for Early Warning during Floods", *Sensors*, 2021, vol. 21, pp. 4942-4959, 2021. DOI: 10.3390/s21144942.
- [12] T. A. Rashid, D. K. Abbas, Y. K. Turel, "A multi hidden recurrent neural network with a modified grey wolf optimizer", *PLoS ONE*, vol. 14, pp. 1-34, 2019. DOI: 10.1371/journal.pone.0213237.
- [13] G. Jung, S. Y. Choi, "Forecasting Foreign Exchange Volatility Using Deep Learning Autoencoder-LSTM Techniques", *Complexity*, vol. 1, pp. 1-16, 2021. DOI:10.1155/2021/6647534.
- [14] M. Meng, M. Zhu. "Deep Convolution-based LSTM Network for Remaining Useful Life Prediction", *IEEE Transactions on Industrial Informatics*, vol. 17, pp. 1658-1667, 2020. DOI:10.1109/TII.2020.2991796.
- [15] G. Dorgo, J. Abonyi, "Learning and predicting operation strategies by sequence mining and deep learning", *Computers & Chemical Engineering*, vol. 128, pp. 174-187, 2019. DOI:10.1016/j.compchemeng.2019.06.006.
- [16] X. Ren, W. Yang, X. Jiang, et al. "A Deep Learning Framework for Multimodal Course Recommendation Based on LSTM+Attention", *Sustainability*, vol. 14, pp. 2907-2920, 2022. DOI:

<https://doi.org/10.3390/su14052907>.

- [17] D. R. Liu, S. J. Lee, Y. Huang, et al. "Air pollution forecasting based on attention-based LSTM neural network and ensemble learning", *Expert Systems*, vol. 37, pp. 1-16, 2020. DOI:10.1111/exsy.12511.
- [18] O. Jogunola, B. Adebisi, K. V. Hoang, et al. "CBLSTM-AE: A Hybrid Deep Learning Framework for Predicting Energy Consumption", *Energies*, vol. 15, pp. 810-825, 2022. DOI:10.3390/en15030810.
- [19] B. Lagesse, S. Wang, T. V. Larson, et al. "Predicting PM2.5 in Well-Mixed Indoor Air for a Large Office Building Using Regression and Artificial Neural Network Models", *Environmental Science & Technology*, vol. 54, pp. 15320-15328, 2020. DOI:10.1021/acs.est.0c02549.
- [20] K. Y. Almansi, A. Shariff, B. Kalantar, et al. "Performance Evaluation of Hospital Site Suitability Using Multilayer Perceptron (MLP) and Analytical Hierarchy Process (AHP) Models in Malacca, Malaysia", *Sustainability*, vol. 14, pp. 3731-3766, 2022. DOI: 10.3390/su14073731