



Online Learner Categories Recognition Based on MoE-LSTM Model

Minghu Wang¹ · Yanhua Gong¹ · Zhikui Shi¹

Received: 26 July 2023 / Accepted: 3 March 2024
© The Author(s) 2024

Abstract

Online learner behaviour patterns are comprehensive indicators reflecting the learning status and learning outcomes of online learners, which have a guiding role in the design, implementation and improvement of online education in the post-epidemic era, but the methods for their identification still need to be improved. To improve the accuracy of online learner identification, a model of online learner behaviour identification based on improved long and short-term memory networks with deep learning is established. Firstly, the online learner behaviour characteristics are extracted according to the behavioural science theory, secondly, the online learner type is identified based on the "hybrid expert system-long and short-term memory network" model, and then compared with other identification models, and finally, the results are outputted by the progressive gradient regression tree GBRT in the stacking integration framework and validated using The results were validated using a ten-fold crossover. The results show that the method is effective in portraying online learners, and its accuracy and robustness are improved compared to other algorithms.

Keywords Online learning · Long and short-term memory networks · Hybrid expert systems · Progressive gradient regression trees · Learner profiling

1 Introduction

In the Internet era, with the rapid development of information technology, the scale of online learning is also gradually expanding. At present, China has entered the new normal of "mass entrepreneurship and innovation", and both the state and the society have put forward higher requirements for higher education, and universities should actively change the mode of talent training, improve the quality of education and teaching, and cultivate comprehensive quality talents. According to the Statistical Report on the Development of China's Internet, as of June 2021, China's Internet

users reached 1.011 billion and the Internet penetration rate reached 71.6%. Online education accounts for 302.5 billion.

The Ministry of Education is accelerating the transition by exploring new opportunities and new models of cross-border integration in specific areas. It is developing online education centers to guide university faculty in new directions of online education change, such as Gaotu's opening of Gaotu Academy, which focuses on adult training, the launch of the new brand "Bufan Classroom" by Homework Help, and the establishment of an adult education division by NetEase Youdao to establish an adult education-centered education center to support the new mission of "teaching for the masses" is to improve the professional knowledge and professional value of adults. The 56th China Education Expo proposed "data-driven high-quality development of higher education and empowering education governance innovation", and data-driven education reform, innovation and upgrading are effective ways to improve higher education. Personalized teaching, scientific evaluation, intelligent decision-making, and accurate prediction based on big data have become important means to integrate higher education resources, implement intelligent online education and improve the teaching level.

Minghu Wang, Zhikui Shi have contributed equally to this work.

✉ Yanhua Gong
676224167@qq.com
Minghu Wang
wangminghu@xauat.edu.cn
Zhikui Shi
szk1357234291@xauat.edu.cn

¹ School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, Shaanxi, China

At present, online learning has become an important new learning method. The 2020 sudden epidemic has objectively and quickly pushed online education to universalization and popularization. Since online learning has the remarkable characteristics of time and space separation, enjoying the convenience of the high rate of shared resources and not limited by time and space, but also the problems of lack of humanized communication and insufficient self-adaptability, how to effectively classify the types of online learners and improve the learning efficiency of online learners are the focus of continuous attention in online education. In recent years, scholars have used machine learning methods to analyze online learning behavior, identify online learner types, and thus predict online learning efficiency has become a hot research topic among scholars. The field of machine learning has developed extensively, and its prediction methods are abundant, and the researchable space is very wide. However, most of the existing studies are limited to the prediction of individual models, which cannot effectively grasp the dynamic changes of online learners' behavioral states. Ming-Yan Zhang [et al. 1] combine student behavior pattern analysis and introduce LSTM-Autoencoder to feature processing of student behavior time series to build an early warning model of student online learning and identify at-risk students and earliest intervention points. Zhang, Xiaoqing et al. [2] proposed a novel recognition method MoE-LSTM for reversed gates by fusing the hybrid expert model MoE and the long and short-term memory network LSTM, which improved the accuracy of reversed gate recognition. In this paper, based on this method for different learners' learning behavior differences, we use the MoE model to build multiple expert (Expert) networks to process learning behavior data from different sources, improve the model's recognition effect for online learner types, and use LSTM to learn the contextual relationship of action segments for learning behavior characteristics of the same learner in different learning periods.

The data in this paper consists of 480 online learners, 4 behavior types and a database of 8160 behavioral data. A model for online learner-type identification is proposed, which combines the advantages of fitting hybrid expert networks to different data sources and the long-term nature of long and short-term memory networks to samples to identify online learner behavior classification prediction problems.

2 Related Research

As early as 1900, scholars at home and abroad began to study online learning, and Pedro Isaias first proposed the concept of online education in higher education. With the development of the times and social progress, online learning has received more and more attention and focus and

has become an important research topic in modern distance education. In this paper, we start from an overview of online learning and explore its related issues and development trends. It can be divided into 3 aspects according to the research direction: education theory, technology application, and personalized innovation. According to the research content, it can be divided into two aspects: one is the research on the study of the external environment of online learning, and the other is the analysis of the analysis of the implicit behavior of online learning.

2.1 Study of the Online Learning Exogenous Environment

The factors that affect the learning efficiency of online learners are not only internal to the analysis of learners' own behavior but also external environmental factors arising from external interactions. Ke Zhu [3] proposes a model of online learner interaction, analyzes operational interaction, information interaction, and conceptual interaction, points out the current situation of low engagement of online learners and low level of conceptual interaction, and designs an interaction model that can be used to predict the learning behavior of online learners. Yunzhen Liang constructed a two-dimensional multi-layer ground interaction assessment framework and found that the problem-solving behavior model and knowledge construction behavior model have significant positive effects on learning performance. Gao Wei et al. [4] argued that platform factors are crucial, while teacher management should also be given attention. Alghasab et al. [5] analyzed student-posted content and teacher comments and interactions in an online learning environment and found that student responses in interactions were influenced by instructive content posted by the teacher. Zhang et al. [6] developed a video player based on Javascript technology and autonomously collected video interaction data from online learners. They proposed a model of factors influencing video interaction behavior based on the law of attribution and found correlations between personal ability, effort, learning preference, task difficulty, and learning performance and video interaction behavior.

In addition, many scholars have conducted systematic studies on the evaluation methods of online education. Tianmin Feng and Shilu Zhang [7] pointed out that the purpose of evaluation of online learning is to monitor the learning process, ensure the quality of learning, and promote students' development. Lei Juncheng [8] compared the course evaluation methods of three major MOOC platforms, Udacity, Coursera, and Edx, and analyzed that their evaluation models are similar to the current evaluation models of traditional classroom teaching. Dai Lili and Li Jingshan [9] pointed out that flipped classroom learning needs to design the basic value orientation of the evaluation system, three

key dimensions and index composition, operation methods, and weight distribution.

Carlos J. Asarta et al. [10] studied that the speed of network transmission has a significant impact on students' application of online platforms and that the quality of the network environment directly affects students' motivation to learn online [11] studied the correlation between students' content behavior posting and used regression analysis to predict subsequent browsing and evaluation behavior using the number of times learners posted. moon, Kim et al. [12] applied different functional modules of an online learning platform for learning activities and found that there were differences in learners' learning behaviors under different functional modules. Jie et al. [13] developed a Spark distributed system for a logging system of user operations in a mobile e-learning environment to analyze the behavioral patterns of online learners in generating content.

3 Research on the Analysis of Online Learning Implicit Behavior

Online learning implicit behavior analysis is more based on education theory, focusing on group classification with online learners. Wei et al. [14] used the RFM analysis method in business to construct the classification index system of users, introduced Bayesian fuzzy rough sets to approximate the learning behavior index attributes, used the Apriori algorithm to realize the association rule extraction of the online learning model, and divided online learning into three dimensions from structure, function and mode into the clustering analysis of behavior, the recommendation analysis of personalized courses, and the association between learning behavior and learning effect analysis. Yang et al. [15] argue that some learning behavior characteristics (time and effort invested) cannot accurately predict learning effects, and propose that "learners' learning sequences better reflect behavioral paths and cognitive processes", use cluster analysis to analyze the differences in learning effects of different behavioral sequence characteristics, and propose strategies to improve learning behavior from three aspects: course design, learning resource support, and learning management and evaluation. We propose strategies to improve learning behavior in three aspects: curriculum design, learning resource support, and learning management and evaluation.

Foreign researchers proposed their own modeling ideas for online learning, such as using Bayesian networks to model learner behavior, which can be used to assess current state and instructional decisions. Multi-level latent class models constructed to explore the behavioral patterns and regional differences of platform users were used. Researchers have also made greater use of tool software to assist in analyzing the behavioral state of online learners, revealing

how learners' online learning behaviors occur in learning platforms and using big data and learning analytics techniques to analyze and predict learning data.

Student learning is a step-by-step process, and online learning platforms can record learning data at each stage of the learning process, which can be viewed as time-series data. Recurrent neural networks in artificial intelligence are most suitable for dealing with time series problems; RNNs are characterized by recurrent connections between hidden units; if the input is a time series, it can be expanded in the time dimension, and each unit in it, in addition to processing the input data at the current time point, also processes the output of the previous unit and finally outputs a time series. However, RNN models cannot learn the problem of "long dependence". With the emergence of LSTM (Long Short-Term Memory) to eliminate the problem of unreliable long-term dependency of RNNs, LSTM is a special type of RNN that can learn the dependency of data in a longer time range, and LSTM is designed to avoid the problem of "long dependency" in the future. LSTM has the chain structure of RNN, but the repetition module has a different structure. However, there is a single classification method currently used, and most of the studies only focus on the prediction effect of one model, and the incompleteness and optimizability of a single model are yet to be innovated. Traditional time series prediction models (such as ARIMA models) are simple but require time-series data or data after differencing to be stable, and cannot handle nonlinear data, which cannot cope with the diversity and complexity of online learners' behavioral instability, making the prediction results have large errors. Therefore, this paper uses a long short-term memory network LSTM model to deal with behavioral indicators based on deep learning methods. Taking the advantage that the MoE model can integrate multiple base models to improve classification accuracy, we combine the LSTM with the MoE model and propose the MoE-LSTM recognition method to effectively capture behavioral information and improve the accuracy of learner-type recognition.

4 Model Construction

4.1 MoE Expert Hybrid Network

Mixture-Of-Experts (MoE) is both a neural network and a combined model. The way data is generated from dataset data is different. A normal neural network is trained with multiple models separately based on the data. The difference is that each model is called an expert, and the experts to be used are selected through a gating module, and the actual output of the model is a combination of the output of each model and the gated model weights. Thus, hybrid networks can accomplish the classification of data samples simultaneously using

multi-class learning mechanisms. In this paper, a hybrid intelligent system framework based on a combination of MoE and Petri net theory is constructed. Each expert model can use different functions (linear or nonlinear functions). A hybrid expert system is a system that integrates multiple models into a single task to improve classification accuracy by integrating multiple underlying models. In real life, there may be multiple data sources, which can be effective in predicting human behavior under virtual platforms due to its powerful learning capability and strong fault tolerance. In this paper, we will investigate how to use MoE multi-gate control applied to LSTM models to solve the type judgment problem under online learning platforms. Since the data sources collect action data from different learners, there are some differences in the distribution of all data. Individual models tend to be good at processing data from the same learner, but the results are not satisfactory for data from different learners. The MoE model addresses this phenomenon by using multiple base models (Experts) to process data from different sources, and each Expert network has a data region domain that is good at classifying data, which is better than other Expert networks perform better classification and use Gating to determine the Expert network that processes the input data. Unlike other integrated learning methods such as Bagging and Boosting, the MoE model's unique gating network is set up to effectively filter the output of the underlying model and allows multiple underlying models to process the data in parallel, with the performance of the model improving significantly as the data increases.

MoE can be divided into two architectures: Competitive and cooperative. competitive MoE requires data to exist in a discrete data space, while on the contrary, cooperative MoE does not clearly distinguish data regions and allows data overlap. Considering that different learners may produce the same behavioral data, the Cooperative architecture is used in this study. the process of the MoE model for the i th input sample in the data set is shown in Eq. (1), and the output result of the Expert network is multiplied with the corresponding weights and summed to obtain the output result of the model. The results of the weight calculation are shown in Eq. (2).

$$y(x_i) = \sum_{e=1}^n g(x_i)_e f_e(x_i) \quad (1)$$

$$G_\sigma(x_i) = \text{soft max}(x \cdot W_g) \quad (2)$$

4.2 MoE-LSTM Model

In this paper, we study how to transform a continuous-time signal into a discrete event sequence by using Fisher's criterion-based method, and then use an improved adaptive algorithm to complete the modeling and parameter estimation

of the signal. Long Short-Term Memory Neural Network (LSTM) model records long-term state information by adding cell states to the RNN through three "gate" structures: input, hidden, and output, and is effective in classifying long-term dependent sample data. To improve the ability of the LSTM model to learn from different data sources, this paper combines the basic LSTM model with the MoE idea to establish an MoE-LSTM model. The model consists of the input layer, MoE layer, and output layer. The specific model structure is shown in Fig. 1.

(1) Input layer

The input layer consists of input sample data and multiple different LSTM models. Specifically,

the sample data generated by moment i are input to different LSTM models, and each LSTM model unit passes timing information by using the output of the previous moment t as part of the input to the current moment. The learning results of each LSTM model can be represented by a probability distribution. The mathematical expression is shown in Eq. (3).

$$y_{t+1} = \arg \max_{\gamma} \sum_{i=0}^n \hat{p}_i(y|x_{t+1}, y_t). \quad (3)$$

(2) MoE layer

The MoE layer consists of multiple basic models (Experts). Firstly, the learning results of the upstream input layer are input into the Expert network, each Expert network is responsible for the dynamic selection of LSTM for the upper input samples by the Gating network, Then the output of multiple Expert networks is mapped to a weight vector, look at the i element in g indicates the i th LSTM network output result in the final output result's weight in the final output, the weight calculation formula is shown in Eq. (4).

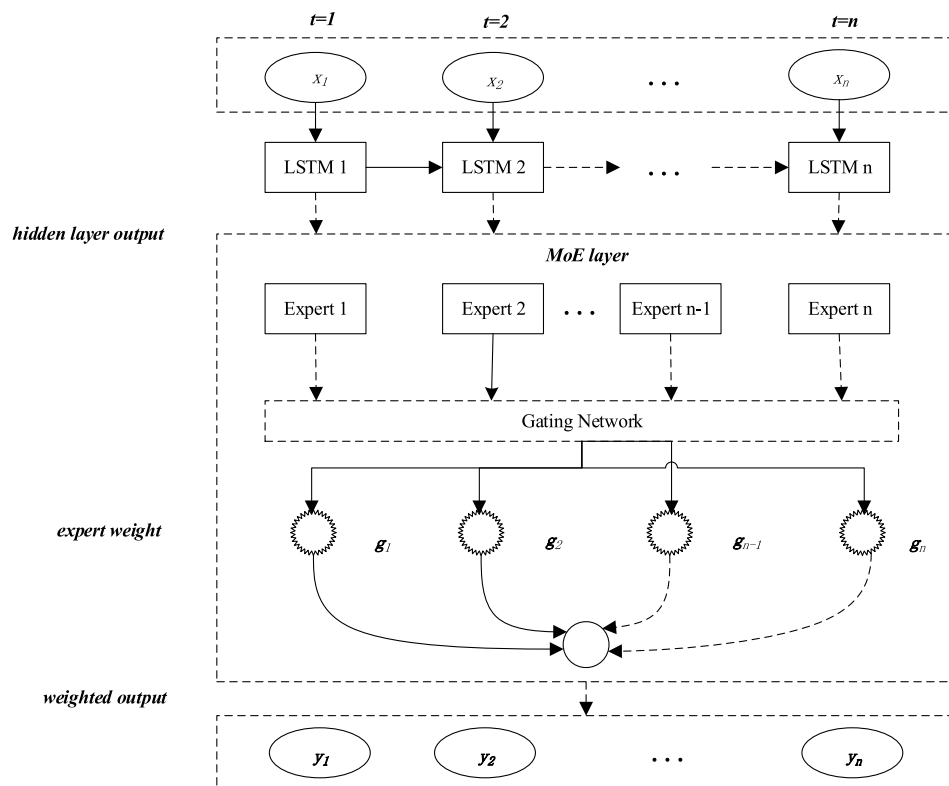
$$L = \sum_I g_i (y - f_i)^2. \quad (4)$$

As shown in Eq. (4), the objective function of y represents the true labels corresponding to the input samples, the output of the first LSTM model, and the weights corresponding to the first LSTM model.

(3) Output layer.

Through the weight calculation of the multi-layer Experts model. The data from different sources are modeled and learned using Eq. (5) and backpropagation. The gradient obtained from Eq. (5) is used to correct the LSTM model. If the expert reliability is low gradient value gradient is small, then the LSTM model needs only a small adjustment. The gradient derived from this Eq. (6) is used to correct the expert reliability, and LSTM model errors that are smaller than the average error indicate that the LSTM model has improved reliability and learns the input data better than other LSTM models. Finally, the random forest algorithm

Fig. 1 MoE-LSTM model structure



is used to optimize the MoE-LSTM network weights and thresholds. The output result of the final model is the product of the output vector matrix of all LSTM model models and the weight vector g .

$$\frac{dL}{df_i} = g_i(y - f_i) \tag{5}$$

$$\frac{dL}{dx_i} = g_i((y - f_i)^2 - L) \tag{6}$$

It should be noted that, in the MoE-LSTM model, the input data, the input result, and the previous moment's output prediction calculation formula are shown in Eq. (3). In summary, the MoE-LSTM model is trained with the aim of minimizing the weighted sum of prediction biases of all LSTM models. Achieved a more accurate and more comprehensive online learner identification and classification. The MoE-LSTM model has the following advantages:

(1) The MoE-LSTM model, by introducing the Mixture of Experts (MoE) mechanism, enables different expert models to collaborate. This not only allows the dynamic adjustment of the weights and parameters of expert models based on the diverse types and behavioral patterns of learners, thus adapting in real-time to learner variations but also enhances the interpretability and generalization capability of the model.

(2) The LSTM network within the MoE-LSTM model possesses its own memory units and gate mechanisms,

enabling effective capture and storage of long-term dependency information in sequential data, and learning patterns and features within time-series data. This capability allows the MoE-LSTM model to better comprehend and identify the learning behaviors and habits of online learners, including learning paths, frequencies, and durations.

4.3 Gradient Boost Regression Tree

A progressive gradient regression tree is an improved algorithm combining a decision tree and a gradient boosting tree. This algorithm can effectively deal with the classification problem of sample data sets under uncertainty. This paper presents the progressive gradient regression tree model and some of its results achieved in practical applications, using the negative gradient of the loss function as an approximation of the residuals in the boosted tree algorithm. The basic procedure of the algorithm is as follows.

(1) Initialize the weak learner.

$$f_0(x) = \arg \min_c \sum_{i=1}^N L(y_i, c). \tag{7}$$

(2) For the sample $i = 1, 2, \dots, N$, the negative gradient is calculated.

$$r_{im} = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}x} \tag{8}$$

(3) The calculated residuals are taken as the true values of the samples, and the data $(x_i, r_{im}), i = 1, 2, \dots, N$ as the training data for the next tree to get the new regression tree $f_m(x)$ whose corresponding leaf nodes are $R_{jm}, j = 1, 2, \dots, J$ where J is used as the number of leaf nodes of the regression tree.

(4) For the leaf region $j = 1, 2, \dots, J$ Calculate the best-fit value.

$$\gamma_{im} = \arg \min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma). \tag{9}$$

(5) Update the learner.

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^J \gamma_{jm} I(x \in R_{jm}). \tag{10}$$

(6) Get the final learner.

$$f(x) = f_M(x) \tag{11}$$

$$f_M(x) = f_0(x) + \sum_{m=1}^M \sum_{j=1}^J \gamma_{jm} I(x \in R_{jm}). \tag{12}$$

5 Experimental Results and Analysis

5.1 Data Sources

The data used in this study comes from the Tianchi Big Data Platform, which is owned by AliCloud. The platform mainly provides a variety of data types for serving school teaching and management. After combing and analyzing the related literature, we found that there are few research results on smart campuses in China. In this paper,

we aim to build a complete system framework and search the xAPI-Edu-Datax public dataset with STUDENT as the keyword, Wang s research is divided into three main categories: (1) basic student information: gender and age; (2) learning behavior information: number of hands raised, access announcements, etc.; and (3) learning effect information: online class classification (L low, M medium, H high). The deep learning framework used in this experiment is Tensorflow 2.4.0, the programming language is Python 3.7.9, and the operating system is Windows 10.

5.2 Behavioral Feature Extraction

Based on behavioral science theories, learners move around the online learning platform and generate learning behaviors according to their needs. These learning behaviors include cognitive, affective, and attitudinal performance, which can be used to assess students' mastery of knowledge and learning effectiveness, as well as an important criterion for evaluating teachers' teaching levels and effectiveness. To analyze the learning behaviors and construct behavioral characteristic indicators, the discrete variables in the dataset—class type—are processed by ONE-HOT coding and do Z-Score normalization, some results of the processing are shown in Fig. 2, and the correlation coefficients of each behavioral characteristic are shown in Fig. 3. Five types of online learning are extracted from the data Episodic behaviors: raise hands frequency raise hands, get resourcesvisitedResources, visit announcementsAnnouncementsView, interaction timesDiscussion, absence daysAbsenceDays, according to the xAPI specification, the behavioral features are demanded to describe, see Table 1. where raise hands, visited Resources, AnnouncementsView, and Discussion are all positively related to the learning behavior classification results, and Absence Day is inversely related to the learning behavior classification results.

	raisedhands	VisiTedResources	AnnouncementsView	Discussion	gender_F	gender_M	NationalTy_Egypt	NationalTy_Iran
raisedhands	1.000000	0.691572	0.643918	0.339386	0.149978	-0.149978	-0.024964	-0.094925
VisiTedResources	0.691572	1.000000	0.594500	0.243292	0.210932	-0.210932	0.005028	-0.081024
AnnouncementsView	0.643918	0.594500	1.000000	0.417290	0.052139	-0.052139	0.031622	-0.051854
Discussion	0.339386	0.243292	0.417290	1.000000	0.124703	-0.124703	0.010264	-0.003192
gender_F	0.149978	0.210932	0.052139	0.124703	1.000000	-1.000000	-0.040886	-0.046264
...
ParentschoolSatisfaction_Bad	-0.297015	-0.363835	-0.298744	-0.061104	-0.093478	0.093478	-0.047985	0.024970
ParentschoolSatisfaction_Good	0.297015	0.363835	0.298744	0.061104	0.093478	-0.093478	0.047985	-0.024970
StudentAbsenceDays_Above-7	-0.463882	-0.499030	-0.312134	-0.218778	-0.209011	0.209011	0.044519	0.061775
StudentAbsenceDays_Under-7	0.463882	0.499030	0.312134	0.218778	0.209011	-0.209011	-0.044519	-0.061775
label	-0.231016	-0.161748	-0.143996	-0.121971	-0.123675	0.123675	0.012831	0.069589

Fig. 2 Data normalization results

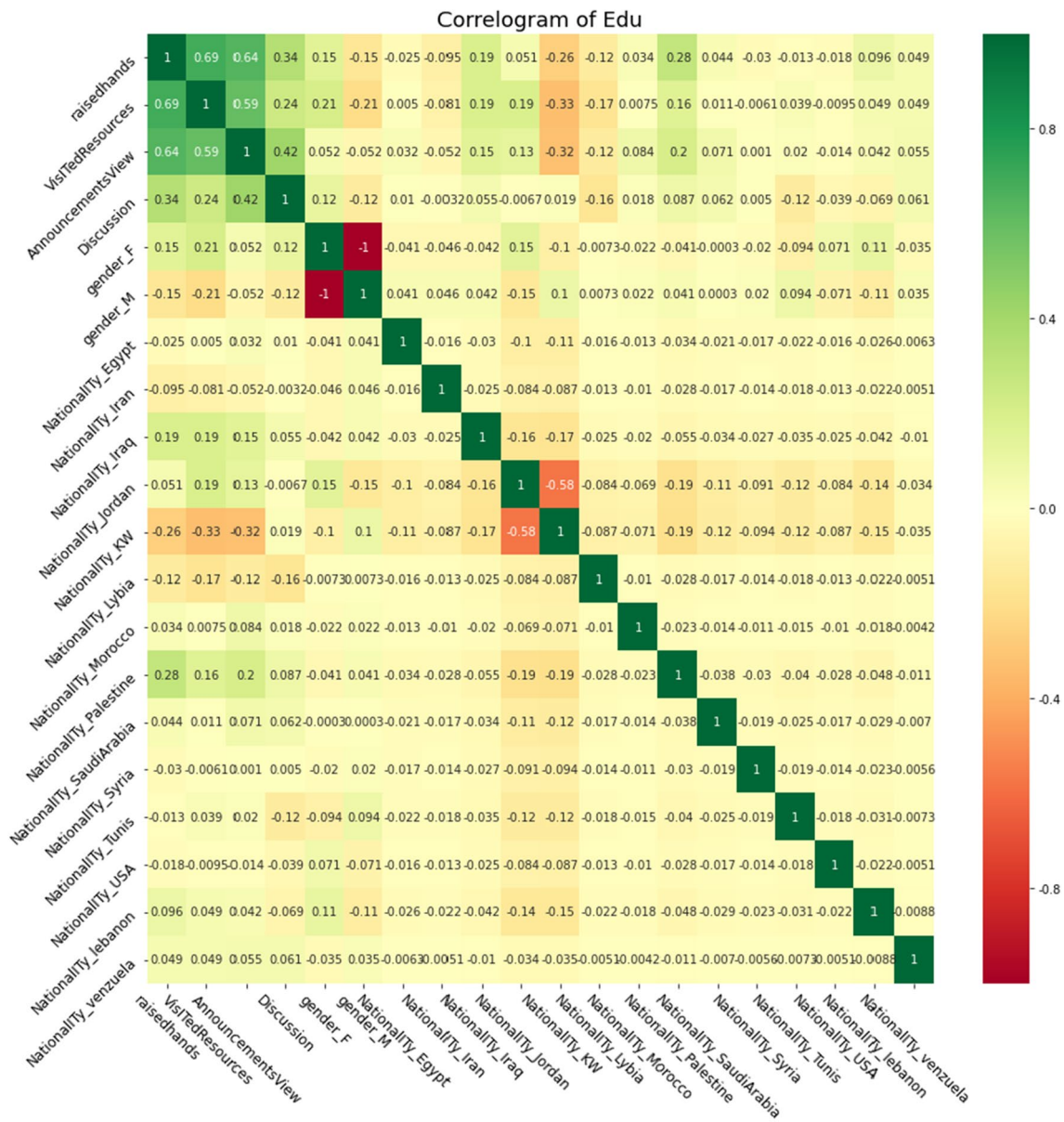


Fig. 3 Feature Correlation

Table 1 Online learning behavior feature extraction

Requirement Description	Learning behavior characteristics
Raise your hand to ask questions based on the need for answers	Frequency of hand raising
Generate resource access behavior based on the demand for the resource	Get Resources
The need to view lesson plans, learning progress and results, resulting in the act of viewing announcements	Access Announcement
Generate discussion behavior based on the need to interact with them collaboratively	Number of interactions
Record days of absence based on the need to check student learning status and engagement	Number of days absent

Table 2 Prediction results of training set and validation set

Student Category	Training true value	Training Predicted Values	Verify the true value	Validation of predicted values
L (low efficiency)	121	112	33	30
M (medium)	108	104	22	23
H (High efficiency)	159	172	37	39

5.3 Model Results

It can be seen that as the dataset grows, the MoE integration becomes more effective. Table 3 Comparison of performance results on different validation sets Based on the performance of MoE-LSTM, the number of Expert networks was determined. The highest accuracy of model prediction results in the training set was achieved when the number of Expert networks was set to 3, respectively, so the number of Experts was set to 3. Then prediction analysis and comparison experiments were done with these two models on the test set. The results show that compared with the traditional linear regression method, MoE-LSTM can better reflect the

nonlinear features and time-varying characteristics in the actual problem, has a certain generalization ability, and can effectively improve the classification performance of the dataset, thus effectively improving the prediction accuracy. The method finally verifies the above conclusions by simulation results. Table 2 shows the classification of the training set and the validation set.

The number of iterations is uniformly determined as 20 and the number of neurons is uniformly determined as 32 according to the degree of convergence of the loss function in the training process. The experimental results show that the method has better performance compared with other algorithms.

Figure 4a, b shows the performance of the model in terms of accuracy and loss values in the training and validation sets, and the accuracy of the validation set is the highest when the number of iterations is 20, which is 0.82609%.

5.4 Experimental Comparison

To analyze the advantages of the MoE-LSTM model over other models, this study compares the MoE-LSTM model with the base classification methods of LSTM and LSSVM and the improved LSSVM method. The evaluation metrics

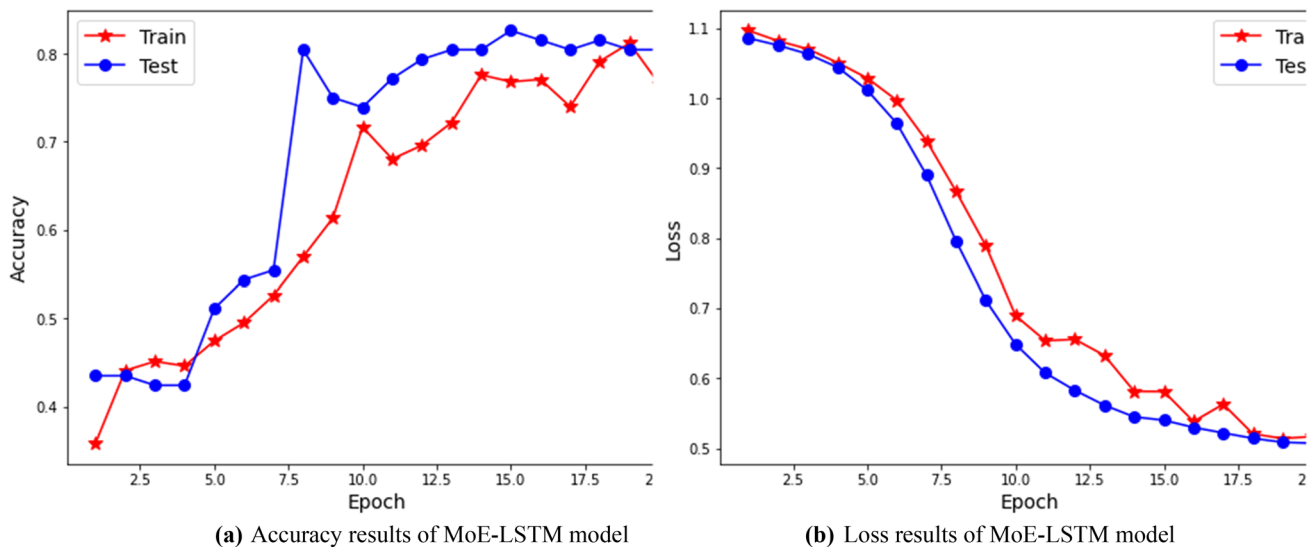


Fig. 4 a Accuracy results of MoE-LSTM model, b Loss results of MoE-LSTM model

Table 3 Comparison of performance results on different validation sets

Serial number	Performance Indicators	MoE-LSTM	LSTM	LSSVM	Improved LSSVM
1	Accuracy	0.8261	0.8043	0.7717	0.7826
2	Accuracy rate	0.8359	0.8175	0.7920	0.8048
3	Average Recall Rate	0.8352	0.8129	0.7737	0.7788
4	Average F1 value	0.8346	0.8150	0.7805	0.7886

include accuracy, precision, average recall, and average F1 value, which are shown in Table 3. It can be seen that the overall accuracy of the MoE-LSTM and LSTM-based models is much higher than that of the traditional LSSVM model and the improved LSSVM model, indicating that the MoE-based integration has certain advantages. Meanwhile, the MoE-LSTM model is more accurate relative to the original model; and the LSTM model outperforms the MoE-LSTM model when the input data volume is small. In addition, the MoE-LSTM model outperforms the LSTM model, and the fused LSTM in the MoE model performs better than the fused convolution operation to extract the temporal features implied in the original data, which has a positive impact on identifying the online learner types. The comparison with the results of previous experiments shows that the MoE integration becomes more effective as the dataset grows.

In this study, the Gradient Boost Regression Tree (GBRT) was chosen as the meta-learning algorithm to integrate the results of the other four sub-models, and the "Gradient-BoostClassifier" package was applied in anaconda3 software for modeling. " package for modeling. The number of regression trees, the depth of each tree, the loss function and the learning rate of the model are the important parameters of the model. Adding more numbers to the model and increasing the depth of the trees will, to some extent, allow the model to capture more bias and increase the complexity of the model, but with the consequent problem of overfitting the model. GBRT can solve the overfitting problem by adjusting the tree structure, convergence, and randomness. The depth of the number basically determines the degree of interaction between variable features, but it is usually set low because the interaction between feature variables cannot be predicted in advance. Setting a smaller learning rate as the convergence strategy will improve the prediction performance of the model, which learns by converging the predictions made by each tree.

In this study, the parameters of the model are set as "learning = 0.1, learning tree n_trees = 4, depth max_depth = 5", and the final output of the fitted model is GBRT_score: = 0.8586956522739131. The performance indicators of the ten-fold cross-validation are shown in Table 4. Performance metrics are shown in Table 4: Combining the accuracy, precision, average recall, and F1 values of the prediction model (see Table 3 for specific results), it can be shown that the model after Stacking framework integration fits the time series better than the four sub-models.

6 Summary

This paper aims to improve the teaching effectiveness of online learning in the post-epidemic era, proposes a prediction model based on MoE-LSTM to identify the type of

Table 4 Details of performance results for tenfold cross-validation

Number of times	Accuracy	Accuracy rate	Recall Rate	F1 value
1	1.0	1.0	1.0	1.0
2	1.0	1.0	1.0	1.0
3	1.0	1.0	1.0	1.0
4	0.9792	0.9744	0.9855	0.9793
5	1.0	1.0	1.0	1.0
6	1.0	1.0	1.0	1.0
7	1.0	1.0	1.0	1.0
8	1.0	1.0	1.0	1.0
9	0.7708	0.8087	0.7751	0.7701
10	0.8125	0.8569	0.8359	0.8347
Average value	0.9563	0.6940	0.9577	0.9584

online learners, and compares and analyzes the prediction results of this method with other models, the results show that (1) the five online learner behavioral characteristics of hand raising frequency, access to resources, access to announcements, number of interactions, and days of absence have the highest relevance to student classification and can be used as important indicator for identifying student types; (2) the accuracy, precision, recall, and F1 values of the prediction results of the MoE-LSTM model are better than those of LSTM, LSSVM, and improved LSSVM, indicating that the MoE module optimizes LSTM better; (3) the fitting effect of using the asymptotic gradient enhancement tree GBRT as a combined model integrating the results of the other four submodels is better than that of the four sub-models, and GBRT is more suitable for being in multi-dimensional data and improving prediction accuracy. Based on the above findings, an effective portrayal of online learners can provide teachers with corresponding teaching suggestions to help students better grasp knowledge and improve teaching quality, and can also provide more targeted personalized recommendation services for universities. In addition, since only the characteristics of learners' own behavioral factors are considered, in the future, we can consider the perspective of environmental factors of online learning behavior and optimize the parameters of the recognition model to further improve the generalization ability and recognition accuracy.

Acknowledgements This work is supported by the Key Industry Innovation Chain Project of Shaanxi Provincial Department of Science and Technology [grant number 22ZZDLGY06] and the Key Project of Philosophy and Social Sciences of Shaanxi Provincial Department of Education (21JZ035).

Author Contributions All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by Minghu Wang, Yanhua Gong, and Zhikui Shi. The first draft of the manuscript was written by Minghu Wang and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Conflict of Interest To the best of our knowledge, the named authors have no conflict of interest, financial or otherwise. The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Zhang, M.Y., Du, X., Li, H.: A study on early warning of achievement combined with student behavior pattern analysis. *Comput. Eng. Appl.* **58**(01), 99–105 (2022)
- Zhang, X., Xiao, W., Guo, Y., Liu, B., Han, X., Ma, J., Gao, G., Huang, H., Xia, S.: Fusing LSTM and MoE for inverted gate operation identification. *J. Syst. Simul.* **34**(08), 1899–1907 (2022)
- Ke, Z.: Research on the analysis model and application of learner interaction in e-learning space. *Electrochem. Educ. Res.* **38**(05), 43–48 (2017)
- Wei, G.: Research on the interaction between teachers and students' speech acts in classroom teaching. *Educ. Res. Exp.* **05**, 43–49 (2009)
- Alghasab, M., Hardman, J., Handley, Z.: Teacher-student interaction on wikis: Fostering collaborative learning and writing. *Learn. Cult. Soc. Interact.* **21**, 10–20 (2019)
- Zhang, G., Li, Z.: A study on the attribution of high dropout rate of MOOCs based on qualitative analysis. *Electrochem. Educ. Res.* **39**(1), 29–35 (2018)
- Wang, X., Lu Wang, W., et al.: A study of ozone variation trend within area affecting human health in Hong Kong. *Chemosphere* **52**(9), 1405–1410 (2003)
- Xintong, G., Hongquan, S., Liuko, L., et al.: Spatial and temporal variation characteristics of ozone concentration in China from 2015–2017. *Meteorol. Environ. Sci.* **43**(03), 41–50 (2020)
- Tao, W., Likun, X., Peter, B., et al.: Ozone pollution in China: a review of concentrations, meteorological influences, chemical precursors, and effects. *Sci. Total. Environ.* **575**, 1582–1596 (2017)
- Asarta, C.J., Schmidt, J.R.: The effects of online and blended experience on outcomes in a blended learning environment. *Internet Higher Educ.* **44**, 100708 (2020)
- Money, W.H., Dean, B.P.: Incorporating student population differences for effective online education: a content-based review and integrative model. *Comput. Educ.* **138**, 57–82 (2019)
- Moon, J.W., Kim, Y.G.: Extending the TAM for a World-Wide-Web context. *Inform. Manag.* **38**(4), 217–230 (2001)
- Jie, W., Hai-Yan, L., Biao, C., et al.: Application of educational data mining on analysis of students' online learning behavior. In: 2017 2nd International Conference on Image, Vision and Computing (ICIVC). IEEE, (2017)
- Wei, L., Li, Y., Wei, Y.: Representative information retrieval algorithm based on PageRank algorithm and mapreduce model. *Int. J. Database Theory Appl.* **9**(3), 25–36 (2016)
- Xiaojuan, Y., Xiujuan, Z., Jinzhi, Ai.: The two-way integration of information technology in primary and secondary schools under “integrated” education. *Mod. Educ.* **07**, 22–26 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.