计算机科学
COMPUTER SCIENCE

Student Academic Performance Predictive Model Based on Dual-stream Deep Network

XIE Hui, ZHANG Pengyuan, DONG Zexiao, YANG Huiting, KANG Huan, HE Jiangshan, CHEN Xueli

引用本文

XIE Hui, ZHANG Pengyuan, DONG Zexiao, YANG Huiting, KANG Huan, HE Jiangshan, CHEN Xueli. Student Academic Performance Predictive Model Based on Dual-stream Deep Network[J]. 计算机科学, 2024, 51(10): 119-128.

XIE Hui, ZHANG Pengyuan, DONG Zexiao, YANG Huiting, KANG Huan, HE Jiangshan, CHEN Xueli. Student Academic Performance Predictive Model Based on Dual-stream Deep Network[J]. Computer Science, 2024, 51(10): 119-128.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240300097

Student Academic Performance Predictive Model Based on Dual-stream Deep Network

XIE Hui[1,2], ZHANG Pengyuan[1,2], DONG Zexiao[1,2], YANG Huiting[1,2], KANG Huan[1,2], HE Jiangshan[1,2] and CHEN Xueli[1,2,3]

1 Center for Biomedical-photonics and Molecular Imaging, Xi'an Key Laboratory of Intelligent Sensing and Regulation of trans-Scale Life Information, School of Life Science and Technology, Xidian University, Xi'an 710126, China

2 Engineering Research Center of Molecular and Neuro Imaging, Ministry of Education, Xi'an 710126, China

3 Innovation Center for Advanced Medical Imaging and Intelligent Medicine, Guangzhou Institute of Technology, Xidian University, Guangzhou 510550, China

(hxie@xidian.edu.cn)

Abstract Blended teaching is one of the essential teaching methods with the development of information technology. Constructing a learning effect evaluation model is helpful to improve students' academic performance and helps teachers to better implement course teaching. However, a lack of evaluation models for the fusion of temporal and non-temporal behavioral data leads to an unsatisfactory evaluation effect. To meet the demand for predicting students' academic performance through learning behavior data, this study proposes a learning effect evaluation method that integrates expert perspective indicators to predict academic performance by constructing a dual-stream network that combines temporal behavior data and non-temporal behavior data in the learning process. In this paper, firstly, the Delphi method is used to analyze and process the course learning behavior data of students and establish an effective evaluation index system of learning behavior with universality; secondly, the Mann-Whitney U-test and the complex correlation analysis are used to analyze further and validate the evaluation indexes; and lastly, a dual-stream information fusion model, which combines temporal and non-temporal features, is established. The learning effect evaluation model is built, and the results of the mean absolute error (MAE) and root mean square error (RMSE) indexes are 4.16 and 5.29, respectively. This study indicates that combining expert perspectives for evaluation index selection and further fusing temporal and non-temporal behavioral features that for learning effect evaluation and prediction is rationality, accuracy, and effectiveness, which provides a powerful help for the practical application of learning effect evaluation and prediction.

Keywords Blended teaching, Expert perspective indicators, Two-stream information fusion model

中图分类号 G434

1 INTRODUCTION

With the rapid development of Internet technology, more and more schools have gradually changed their education methods from single offline teaching methods to blended teaching methods combining online and offline. However, in the rapid development of blended teaching, there are still some problems that cannot be ignored. At the students level: many students' inattention in the process of online learning and the learning efficiency is low, which leads to a significant downward trend in academic performance compared to offline learning; at the teachers level: teachers are unable to judge the progress of courses because they are not informed of the students' recent learning results in real-time, which in turn leads to problems, such as "difficult" teaching. In this case, the quality of students' learning effects has become an essential concern for society, and accurate prediction and evaluation of students' learning effects is one of the most important ways to ensure the quality of their learning.

The current research on learning effect prediction and evaluation can be mainly divided into two categories: the first is the selection of features related to learning effect indicators; and the second is the research on model algorithm design. Different learning effect evaluation indexes, comprehension, and

到稿日期:2024-03-14 返修日期:2024-07-01

This work was supported by the National Key R&D Program of China (2022YFB3203800), National Natural Science Foundation of China (62007026), National Young Talent Program, Shaanxi Young Top-notch Talent Program, Key Research and Development Program of Shaanxi (2022GY-313), Xi'an Science and Technology Project (23ZDCYJSGG0026-2023) and the Fundamental Research Funds for Central Universities (ZYTS23192).

通信作者:Xueli Chen (xlchen@xidian.edu.cn)

typicality are the prerequisites for significantly improving the accuracy of learning effect prediction results. The evaluation indexes can be divided into two categories: classroom and after-school life indexes. Parack et al.[1] started with the students' classroom performance. They realized the prediction and evaluation of the students' learning effects based on the evaluation indexes, such as the students' classroom attendance rate, semester homework grades, and practical test scores. Dien et al.[2] used traditional classroom indexes, such as the average grade performance of students in the previous semesters and the average grade performance in the earlier semesters. They realized the prediction and evaluation of students' learning effect. El Ahrache et al.[3] performed an in-depth analysis of student demographic data and processed the evaluation metrics of various academic courses to achieve the evaluation prediction function. Zhao et al.[4] added multiple types of after-school life indicators to the classroom evaluation metrics, such as students' work and rest time, to evaluate the students' learning effects. Hung et al.[5] normalized the students' original evaluation indicators. They converted them into rankings to achieve higher accuracy in evaluating and predicting learning effects. The evaluation indicators cover different aspects and stages of students' learning and life. However, these approaches are often based on existing platform foundations or calculated after mathematical model training, with a severe lack of correlation validation.

After selecting evaluation indexes and collecting student data, it is necessary to use appropriate algorithms to analyze and discuss the corresponding model of student data. At this stage, research on the design of model algorithms can be divided into two categories: machine learning algorithms and deep learning algorithms. Akçapınar et al.[6] used the K-nearest neighbor algorithm to predict whether or not 76 students taking a computer hardware course would pass at the end of the semester; Rodriguez et al.[7] implemented an artificial neural network model to predict the academic performance of students in colleges and universities, and the model had an accuracy of 82% in the high-achieving category and 71% in the low-achieving category ; Yang et al.[8] proposed to transform students' course participation into image data and then model and analyzed students' image data based on convolutional neural networks ; Min et al.[9] proposed that an evaluator based on long and short-term memory neural networks could obtain more accurate academic predictions ; Hashim et al.[10] used the main traditional characteristics of students as essential variables in a dataset of machine learning algorithms and concluded that logistic regression algorithms are most accurate in predicting students' academic performance. Chitti et al.[11] reviewed the development of educational data mining techniques in a study that delves into how model "black boxes" makes decisions and the role of interpretable artificial intelligence techniques in making model results interpretable. To summarize, at this stage, researchers focus more on improving of the model algorithm, significantly improving the evaluation results. However, adopting the same treatment for temporal and non-temporal types of data severely limits further improvements in model performance.

In summary, in the current research stage, researchers usually build a prediction model for learning effectiveness evaluation by analyzing students' learning behavior data and using academic performance as the model label. Kou et al.[12] pointed out that combining several learning units generates learning behavior and that there is often an interrelationship between these units. A large number of corresponding studies have pointed out that there is indeed a great connection between the trajectory of students' learning behaviors and their academic performance[13-16]. Therefore, existing studies have also identified the behavioral trajectory evaluation indexes corresponding to essential factors affecting students' learning outcomes through online learning data prediction models [17-20]. However, since classroom learning involves not only the process of students' independent learning, but also the process of teachers' teaching, it is evident that relying solely on simple, un-screened indicators for learning effect evaluation prediction cannot yield more interpretable results. Therefore, in the selection and processing stage of learning effect evaluation indicators, incorporating the teaching experience of teacher experts for further manual processing plays a vital role in improving the accuracy of model results and the rationality of data processing. At the same time, although there are also researchers who have discovered the importance of sequence data in the learning process and used lagged sequence analysis to explore students' behavioral patterns. Jeong et al.[21] from Florida State University used sequence analysis to help analyze the sequences of student group information feedback in an online learning environment. They combined the analysis results to propose a framework for evaluating students in a computer-mediated environment, models, etc. Brooks et al.[22] proposed a generalized approach to educational technology for building predictive models based on time series log data. However, the sequence data in the current study often shows the frequency information of switching from one behavior to another and does not reflect the temporal information of the learning process, making the corresponding data support in the construction of the model missing. In addition, the virtual simulation learning practice characterized by new information technology has been used increasingly along with the popularization of education information technology. However, there is still a lack of research systems demonstrating the impact of virtual simulation learning factors on learning effects.

This paper proposes an innovative learning effect evaluation method to address the methodological problems in previous studies. The process mainly consists of two main parts.

1) The establishment of the evaluation index system from the perspective of experts. Firstly, from the background of the blended teaching model based on the Delphi method[23], the perspective of teacher experts is introduced, and a preliminary evaluation characteristic index system is established. Subsequently, mathematical-statistical techniques such as the Mann-Whitney U-test and complex correlation analysis are used to conduct a fundamental analysis of these evaluation indicators. Through these methods, the rationality and validity of the evaluation indexes are preliminary verified.

2) Establishment of evaluation and prediction methods. This study proposes a two-stream information fusion learning effect evaluation model that integrates the temporal and non-temporal behavioral features produced by students in the classroom to realize the prediction and evaluation of students' learning effects. We can assess students' learning effects more accurately by fusing these two features in the processing. In this study, we collects online and offline learning behavior data from 3 852 students from different universities, different majors, and different courses, construct an evaluation index system with typical relevance, and establish a deep network structure that integrates the temporal and non-temporal learning behavior data in a two-stream manner, which provides theoretical support and practical application for improving the learning effects of students and the teaching level of teachers.

2 METHODOLOGY

Most undergraduate students live on campus and are trained in blended online and offline teaching. For offline teaching scenarios, the teacher's end will collect and accumulate some of the behavioral data. At the same time, as the frequency of using the Rain Classroom APP in online teaching in colleges and universities increases, for online teaching scenarios, the backend side of the software collects more behavioral data, such as the number of words posted in class, the number of discussions in class, the length of the virtual simulation operation, the number of conversations after class, and the number of questions asked after class. In this study, we first screen the data using an expert perspective to select effective metrics; then, we construct a two-stream temporal and non-temporal deep network structure to evaluate the learning effect using the screened data. Fig. 1 shows the overall flow of this study.

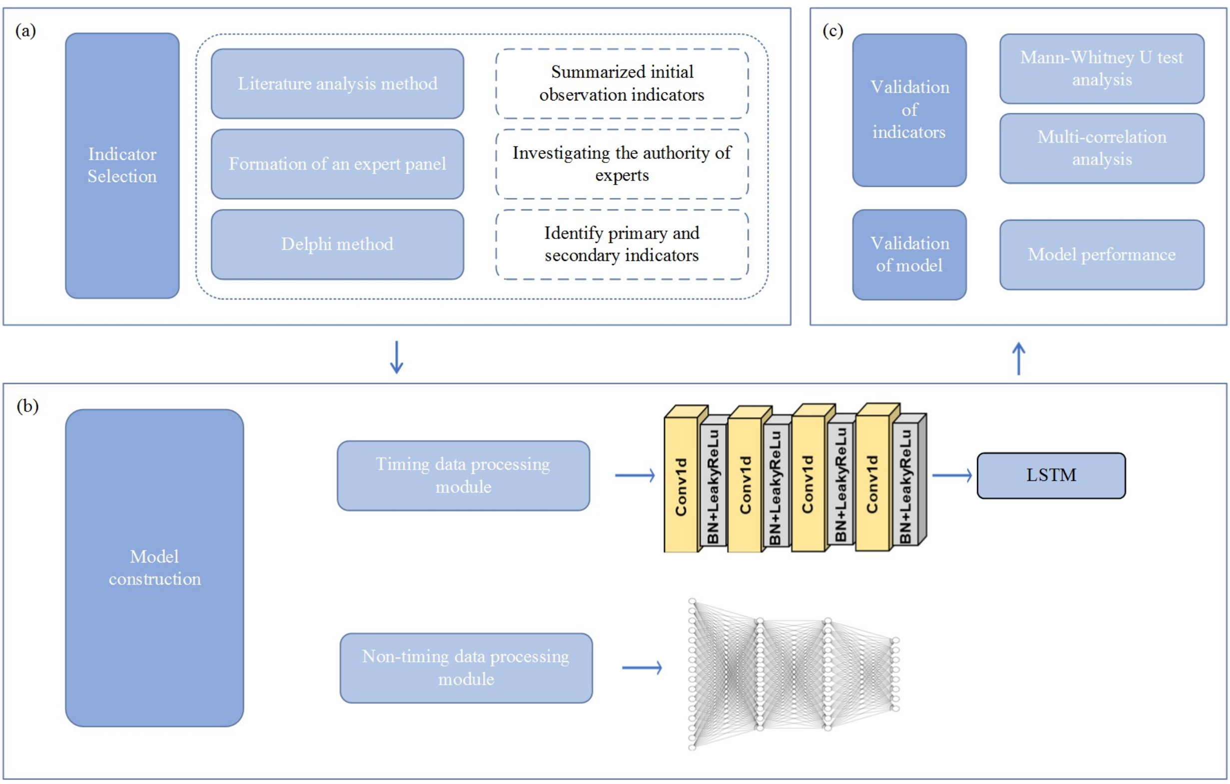


Fig. 1 General overview of evaluation methodology

2.1 Dataset

The data selected for this study comes from the online and offline behavioral data of 3 852 students from 6 universities, dozens of majors, and different courses. It is worth noting that, to avoid the problem that students' final grades may not be able to truly reflect students' learning effects due to the randomness of the difficulty and level of the questions asked by a single teacher for the small group of major and elective courses, the different courses selected for this study mainly focus on the entire course group with a standardized set of questions. The curriculum group standardizes the proposition of the university-wide unified offering of multi-parallel shifts of large classes of basic courses or professional basic courses. In addition, unlike traditional research, the current research on the prediction of academic performance mainly focuses on the teaching of theoretical classes. However, with the continuous maturity of virtual simulation technology, there are already a large number of theoretical courses in the teaching process to integrate the relevant knowledge points of the virtual simulation online practice learning. Moreover, no relevant research has been seen on predicting virtual simulation learning factors in evaluating students' learning effects, so this study introduces a virtual simulation learning module for the course.

The course mainly focus on the introductory natural sciences course, such as advanced mathematics, university physics, and other introductory course, and to carry out further research. This study randomly selects a student's data for display in the specific form shown in Table 1 and Table 2. Each student's data can be categorized into two types: temporal behavioral data and non-temporal behavioral data. Non-temporal behavioral data refers to the information recorded by Rain Classroom. In contrast, temporal behavioral data relates to the data obtained by students during regular learning sessions on the self-designed virtual simulation platform, including the duration of the virtual simulation, the number of questions asked, and the frequency of operations performed. As learning progresses, new temporal behavioral data continues to emerge.

Table 1 Examples of data on students' non-temporal behaviors

Student ID	Percentage of completion/%	…	Number of in-class postings	…	Number of after-school discussions
12	89.8	…	19	…	4
212	59.4	…	5	…	0
2074	32.3	…	5		4

Table 2 Examples of data on students' temporal behaviors

Student ID	Number of virtual simulations 1	Number of virtual simulations 2	…	Number of virtual simulations 13
12	2	1	…	22
212	4	1		12
2074	4	5	…	4

2.2 Indicator Selection

To ensure sufficient sample capacity for the algorithm, this study initially collects data from multiple courses under the complete blended teaching model from different pilot schools, totaling 5 859 students. Only the behavioral data of students' learning under the full-process blended teaching mode are selected, removing the data with incomplete construction of blended course resources, the serious missing of practical features, and learning data generated in a short period concentrated at the end of the semester. At the same time, based on the regular learning behavior characteristics generated by the learning platform, combined with the teaching experience of several senior first-line teacher experts, the Delphi method is applied to determine the final learning effect evaluation index system. The specific process of the Delphi method in this study is as follows:

1) Literature analysis method. According to the research results on online learning/blended learning, learning evaluation systems, and course evaluation systems at home and abroad, we analyze, generalize, and summarize 89 initial observation indicators for evaluation. Through consulting 11 experts and 37 first-line teachers who have been teaching for more than 10 years, we determine 63 effective observation indicators for assessment.

2) Formation of an expert panel. The expert panel of this study consists of 11 teaching masters who have been deeply involved in online teaching and blended teaching for a long time. This study first combines the expert self-assessment questionnaire to survey the degree of specialist authority. The survey results show that 8 experts have an authority coefficient of $Cr \geqslant 0.89$, and 3 experts have an authority coefficient of $Cr \geqslant 0.73$, all in line with the requirements of the Delphi method of an expert authority coefficient of $Cr \geqslant 0.7$.

3) Combining the expert assessment opinions, the basic framework of the learning effect evaluation index systemis determined under the blended teaching mode, including 4 first-level indexes and 13 second-level indexes.

4) The established frameworkis distributed to 11 experts, respectively. After several rounds of soliciting opinions and adjusting the relevant indicators, the 11 experts finally agree to form an effective indicator system for evaluating the learning effect of courses based on the blended teaching mode, including 3 first-level indicators and 10 second-level indicators.

The specific learning effect evaluation index system is shown in Fig. 2. The evaluation indexes finally used in this study are all based on the first-level and second-level evaluation indexes and filtered by teachers' expert recommendations. The finally evaluation indicators are divided into two categories: non-temporal evaluation indicators and temporal evaluation indicators. The non-temporal evaluation indicators include the pre-course learning video completion rate, the total length of the pre-course learning video, the number of posts in the class, the number of words in the post, the number of replies in the class, the number of words in the reply, the number of pop-ups in the class, the number of words in the pop-ups, the number of unintelligible words in the class, the number of discussions in the class, length of discussions in the class, total number of words in the discussions, the average score of the test in the class, length of virtual simulation operation, the number of virtual simulation operations, the number of clicks on the virtual simulation, and an average score of the test in the class, the average score of virtual simulation operation, the number of virtual simulation clicks, the average score of virtual simulation operation, the average score of virtual simulation theory, the number of post-class discussions, the length of post-class discussion, the number of words for post-class debate, the number of post-class questions, the number of post-class replies, the average score of post-class assignments, the average score of the chapter test; temporal evaluation indicators are based on some of the non-temporal evaluation indicators, which can be realized in detail for each classroom. The evaluation indexes include the length of learning in each classroom, in-class discussion, in-class test, virtual simulation operation length, the number of virtual simulation operations, the number of virtual simulation clicks, the score of virtual simulation operations, the score of virtual simulation theories, the post-course discussion, the post-course homework, the chapter test, and the average score of the chapter test.

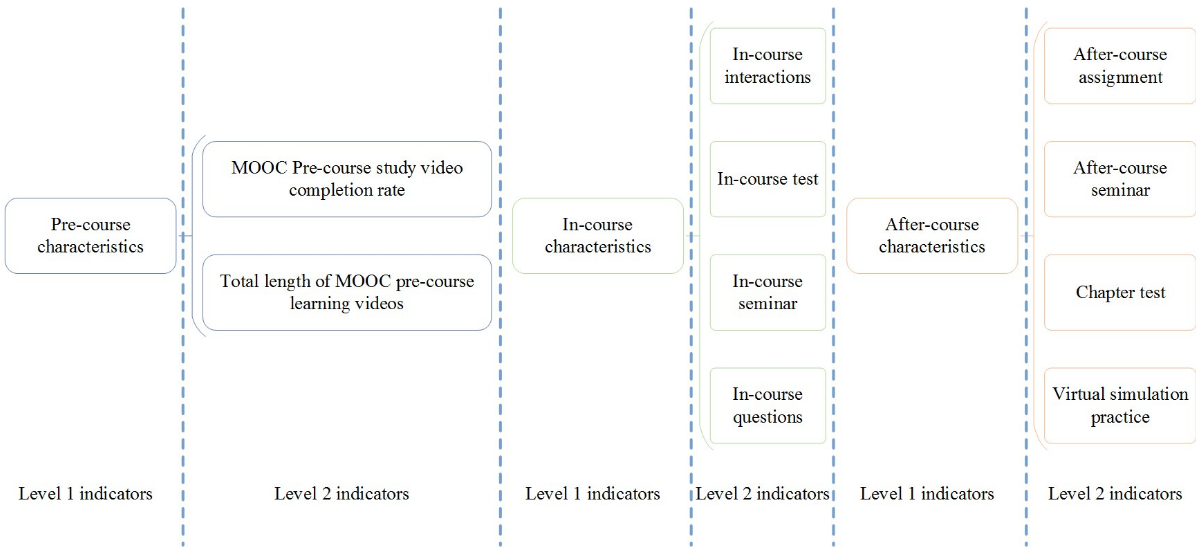


Fig. 2 Evaluation index system of course learning effects in blended teaching mode

To minimize the influence of subjective factors in the prediction of learning effects evaluation and to highlight the advantages and objectivity of the two-stream information fusion learning effect evaluation model proposed in this study in large-scale learning effect assessment, this studydoes not carry out the assignment of the weights of the indexes relying on the experience of experts.

2.3 Definition of Academic Performance

In the current study, college students' course grades are generally composed of ordinary grades (A) and exam grades (B), generally A+B=100. According to different course requirements, ordinary grades often cover multiple features (attendance, online learning data, classroom performance, group seminar, homework completion, chapter test, etc.). Therefore, in order to avoid the final grades given by some course instructors, they have included some of the features of this study in the final grades determination, which leads to the problem of false accuracy of the intelligent evaluationgrades caused by the intersection of the features between the final grades of the course and the intelligent evaluation grades in the study. In this study, the course grades of the 3 852 eligible students used to conduct the study are defined as final grades.

2.4 Network Construction

This study proposes a two-stream information fusion model for student learning effect evaluation prediction in the hope that student behavioral data's temporal and non-temporal features can be processed in parallel. Then, the feature information can be extracted to realize the fusion and complementarity to enhance the scientific accuracy of the model prediction. The framework of the two-stream information fusion learning effect evaluation model proposed in this study is shown in Fig. 3, which is an integrated system of parallel processing of information and mainly includes two modules: the temporal feature extraction module and the non-temporal feature extraction module.

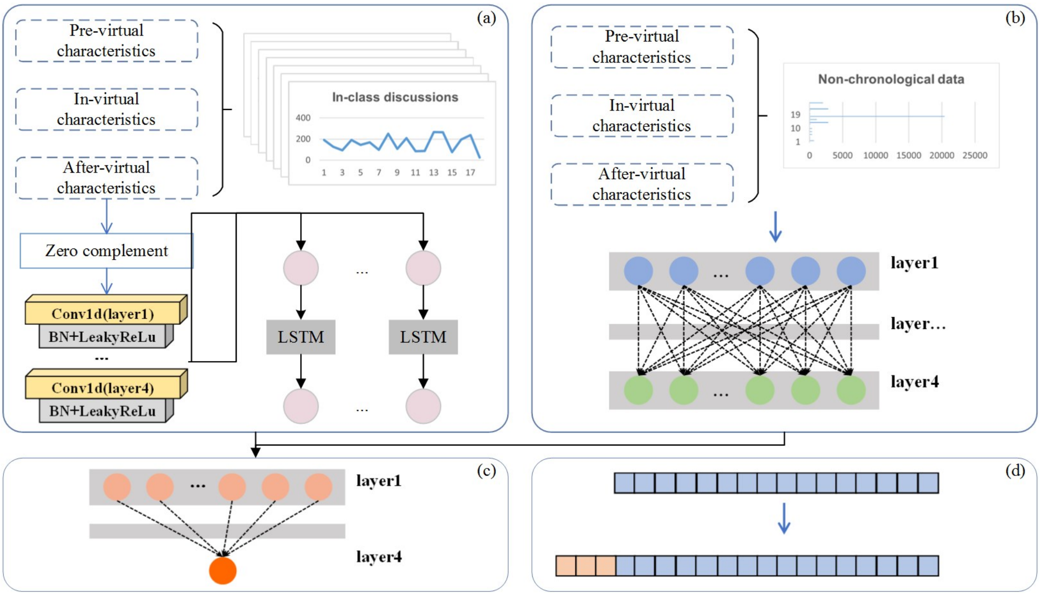


Fig. 3 Overview of algorithm construction section

Firstly, the model needs to perform a time node alignment operation for the length of the time axis of the temporal behavior data, and the specific alignment operation used in this study is forward zero complements which are shown in Fig. 3(d). At the same time, it needs to increase the dimensions of the one-dimensional temporal behavioral data of different categories and splice the temporal behavioral data of different categories in the increased second dimension, which is then input into a 1DCNN for extracting the feature information and the extracted features are further input into the LSTM[24] neural network; then, for the non-temporal behavioral data, it is input into the fully connected neural network to extract the feature information; next, the extracted feature information of the two tributaries is spliced and input into the final fully connected neural network; and the loss is computed to achieve the parameter training according to the MSE loss function.

In the temporal feature extraction module, input $X_i = (X_{i1}, X_{i2}, \cdots, X_{ij}, \cdots, X_{iN})$, which X_i represents the temporal behavioral data of the i^th classmate, N represents the number of temporal features and X_{ij} represents the temporal feature vectors with different lengths of time.

The corresponding timeline lengths are inconsistent, such as when the virtual simulation operation length contains 13 time nodes and the after-class discussion contains 18 time nodes. To avoid the defect of redundant network structure caused by inputting different temporal behavioral data into multiple LSTM neural network with other time periods, this study employs forward padding alignment operations that can be forgotten by the LSTM "gate" structure to address the issue of inconsistent lengths of the time axis in behavioral data across different periods.

To further extract sufficient information among features in X_i^{Cop} after forward padding operations, temporal behavioral data between different periods of the same feature are initially processed through convolutional operators in 1DCNN to extract high-level features. This structure comprises 4 modules, each containing a convolutional layer, a batch normalization layer, and a LeakyReLU activation layer. The computation process is illustrated in Equation(1):

$$X_i^{\mathrm{LSTM}} = LkReLU(BN(W_i X_i^{\mathrm{Cop}} + b)) \quad (1)$$

where LkReLU represents the LeakyReLU activation function, BN represents the batch normalization layer, W_i represents the convolutional kernel, and b denotes the bias term depicted as Equation (2)—Equation (7):

$$I_{i_t} = \sigma(X_{i_t}^{\mathrm{LSTM}} W_{xi} + H_{t-1} W_{hi} + b_i) \quad (2)$$

$$F_{i_t} = \sigma(X_{i_t}^{\mathrm{LSTM}} W_{xf} + H_{t-1} W_{hf} + b_f) \quad (3)$$

$$O_{i_t} = \sigma(X_{i_t}^{\mathrm{LSTM}} W_{xo} + H_{t-1} W_{ho} + b_o) \quad (4)$$

$$\widetilde{C}_{i_t} = \tanh(X_{i_t}^{\mathrm{LSTM}} W_{xc} + H_{t-1} W_{hc} + b_c) \quad (5)$$

$$C_{i_t} = F_t \odot C_{t-1} + I_t \odot \widetilde{C}_t \quad (6)$$

$$H_{i_t} = O_t \odot \tanh(C_t) \quad (7)$$

where F_{i_t} represents the output of the forget gate, I_{i_t} denotes the output of the input gate, $\widetilde{C}_{i_t}$ stands for the candidate memory cell, O_{i_t} signifies the production of the output gate, C_{i_t} represents the memory state, and H denotes the hidden state.

In the non-temporal feature extraction module, the input $Y_i = (Y_{i1}, Y_{i2}, \cdots, Y_{ij}, \cdots, Y_{iN})$, where Y_i represents the non-temporal behavioral data of the i^th classmate, N represents the number of non-temporal features and Y_{ij} represents the specific value of the j^th non-temporal feature of the i^{th} classmate. In the non-temporal feature extraction module, we utilize a multi-layer perceptron (MLP) consisting of 4 layers, with the number of neurons in each layer being 64, 32, 16, and 8, respectively. The computation process is illustrated in Equation(8):

$$A_q = \sum_{p=0}^{M} W_{pq} Y_{pq} \quad (8)$$

where p serves as the index of the neurons in the previous layer, q is the index of the neurons in the current layer, and M represents the number of neurons in the current layer.

Finally, the network concatenates the temporal and non-temporal information across the channel dimension at the feature level and processes it through a four-layer MLP. Employing a multi-layer neuron structure facilitates effectively blending the concatenated features and further making assessments.

3 RESULT

3.1 Indicator Validity Test

To preliminarily investigate the rationality and validity of the evaluation index system obtained by screening under the Delphi method, teacher-expert perspective in this study, firstly, students are categorized into two major groups: excellent and sound, with a cut-off of 75 points for students' academic performance for the Mann-Whitney U-test[25] to observe whether there is a significant difference between the different categories of students under the selected evaluation indexes. The Mann-Whitney U-test is a non-parametric statistical test, which is used to compare whether there is a significant difference in the medians of two independent samples.

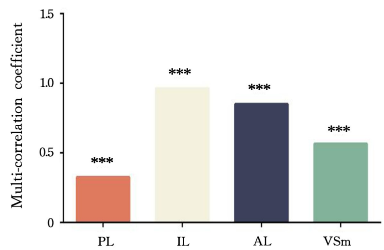
This study conducts a normality test for the selected dataset and then determines the required significant test. Since the sample size of the dataset is 3852, the Kolmogorov-Smirnov test is used. The results show that the significance p-values of the K-S test results are less than 0.001, meaning that all evaluation indicators do not satisfy the normal distribution. Therefore, this study next analyzes the evaluation indicators using the Mann-Whitney U-test. The analysis results show that the significant p-values of the Mann-Whitney U-test are all less than 0.001. There are substantial differences between different categories of students under each type of evaluation index in the evaluation index system proposed in this study. From this, the various evaluation indexes in the system have a certain degree of validity in categorizing students' academic performance.

3.2 Indicator Correlation Test

After completing the Mann-Whitney U-test analysis and initially verifying the validity of the evaluation index system, since the pre-testing process is mainly based on the classification of students' academic performance for statistical analysis, the next part of this study is to use the students' specific academic performance as a metric to further determine whether the evaluation index system is in line with the theory of statistical analysis.

Since the first-level and second-level indicators in the evaluation index system are interrelated and affect each other, this study only utilizes regression models to carry out compound correlation analysis based on the 3 broad categories of indicators in the evaluation index system[26]. Compound correlation analysis is employed to investigate the intricate relationships among multiple independent variables and a dependent variable. This method allows for the simultaneous consideration of the impact of multiple independent variables on the dependent variable and the interaction effects among these independent variables.

The results are shown in Fig. 4. The p-values of the regression models are less than 0.001, indicating that the analyzed models are valid. In addition, there is a correlation between all two types of evaluation indexes and academic performance, and the correlation between the evaluation indexes of in-class learning and academic performance is the largest, with a compound correlation coefficient of 0.962, which is in line with the actual teaching concept. Meanwhile, the virtual simulation factors are analyzed separately in this study, and the results are shown in Table 3. The results show that the compound correlation coefficient between virtual simulation factors and academic performance is 0.566. Virtual simulation factors also affect students' academic performance to a certain extent, proving the necessity of studying them.



Note: PL, indicators for evaluating pre-course learning; IL, indicators for evaluating in-course learning; AL, indicators for evaluating after-course learning; VSm, evaluation metrics for virtual simulation.

Fig. 4 Correlation indices and whether there is a significant correlation between the indicators of assessment of learning outcomes obtained through screening and academic performance under the Delphi method teacher experts' perspectives

Table 3 Compares 10-fold cross-validation for single modal DNN, single modal 2DCNN, single modal LSTM, single modal 1DCNN-LSTM, and two modal models (TIE)

methods	*MAE*	*RMSE*	*R2*
DNN	11.306	13.642	0.394
2DCNN	13.137	16.115	−0.217
LSTM	12.179	14.777	−0.169
1DCNN-LSTM	12.439	15.046	−0.054
TIE(ours)	3.308	4.574	0.897

The results of the compound correlation analysis in Fig. 4 show that the learning effect evaluation index system obtained from the screening under the perspective of teacher experts by the Delphi method is in line with the scientific nature of the screening by manual teacher experts, and also in line with the theory of mathematical statistics. Therefore, the learning effect evaluation index system proposed in this study can be used as a reference for further research on learning effect evaluation and prediction at a later stage.

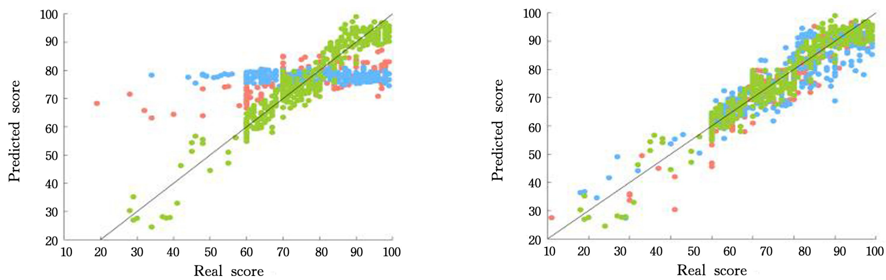
3.3 Validation of TIE Validity

Due to the dual-stream structure adopted by the TIE model, temporally filtered behavioral data and non-temporal behavioral data, from the expert perspective, are merged, complementing each other, and subsequently, the model parameters are trained. Initially, we conduct comparative experiments on the TIE model; specifically, we test the performance differences among the TIE model, single-modal deep neural network(DNN), single-modal 2DCNN, single-modal LSTM, and single-modal 1DCNN-LSTM.

The various models mentioned above are commonly used in educational assessment. The specific mean absolute error (MAE) and root mean square error (RMSE) indices compared with the TIE model are shown in Table 3. The research results indicate that the MAE indices for predicting percentile math scores using the models above are all above 10. In contrast, the proposed TIE model in this study exhibits an average error of only 3.2, demonstrating the effectiveness of this model and the dual-stream data. On the other hand, the R-squared ($R2$) indices for 2DCNN, LSTM, and 1DCNN-LSTM in predicting negative academic performance indicate overfitting, suggesting that using only spatial or temporal models cannot effectively determine academic performance.

The TIE model, which combines the two data types, improves the R2 index to 0.9, suggesting that the model can fit the data exceptionally well based on academic performance. To visually assess the effectiveness of grade prediction, this study uses a scatter plot in Fig. 5(a) to show the prediction results of different models. We can judge the prediction effect by observing the proximity of the scatter points in the scatter plot to the line. The prediction results are better if the points in the scatter plot are closer to the straight line.

According to the results in Fig. 5(a), the predicted values of the two-stream model proposed in this paper are more concentrated on the straight line, which intuitively proves that the two-stream model has significant superiority in grade prediction. Therefore, the proposed TIE model is superior to any unimodal model in terms of performance, and it also proves the necessity and effectiveness of fusing two-stream features for the analysis of predictive modeling for learning effect evaluation.



(a) (b)

Fig. 5 Demonstration of the prediction effect of 1DCNN-LSTM, DNN, TIE, and TIEB, TIEC, TIE models

3.4 Verification of model structure effect

Subsequently, this study conducts a second comparison experiment. Since the TIE model is designed with the powerful function of a convolutional neural network to extract feature information, the 1DCNN module is added to the TIE model in this study. In addition, the TIE model utilizes the "gate" structure principle in the LSTM neural network by placing the zero-complement operation at the start position of the data.

In this study, the model without the 1DCNN is abbreviated as "TIEC", and the model with the zero complementary operation placed at the end position of the data is abbreviated as "TIEB". Experimental comparisons are conducted, and the results are shown in Table 3 and Table 4. The TIEC and TIEB models are better than the TIE model in MAE and RMSE indexes, in which the MAE index reflects the prediction error of the average grade under the percentile system. In the $R2$ index, the TIE model shows higher values, which indicates that the model better fits academic performance through the behavioral data. To visually present the prediction effect of the model, this study adopts a scatter plot to show the prediction results of the three models, shown in Fig. 5(b). By observing the distribution of scatter points in Fig. 5(b), we can see that the prediction results of the model proposed in this paper are more concentrated around the straight line, which is closer to the actual value.

Table 4 Comparison of 10-fold cross-validation results obtained with different zero-completion operations and removal of the 1DCNN fraction

	MAE	*RMSE*	*R2*
TIEC	5.660	7.067	0.763
TIEB	4.007	5.016	0.879
TIE(ours)	3.308	4.574	0.897

The above results show that, based on the learning effect evaluation index system screened under the Delphi method teacher-expert perspective for student behavioral data collection,and after the training of the TIE model proposed in this paper,its prediction effect on MAE and RMSE is better than that of other neural network models,which can realize the effective prediction of students' academic performance in different courses under the blended teaching mode.

4 DISCUSSION

Current research on learning effect evaluation and prediction is limited to single and unreasonable tasks, such as using machine learning algorithms to determine the indicators affecting students' learning effects based on a large volume of platform data and constructing a student learning effect evaluation model based on non-time series feature data. Our research is firstly based on the course that incorporates virtual simulation online practice learning, based on the expert perspective and using the Delphi method to determine the evaluation indexes, combined with statistical analysis methods to evaluate the evaluation index system to assist the side of the validation. Secondly,our research in the fusion of two-stream data features at the same time to join the model optimization strategy, which also makes the model's predictive effect has been further improved,the whole study in terms of rationality and accuracy. The entire research shows excellent results in terms of rationality and accuracy.

Based on the results of the comparative experiments, since the features used in this study have the assistance of expert judgment and data analysis at the same time and show the best results,our future research will be able to formulate corresponding evaluation standards for the evaluation of blended courses, especially those that incorporate virtual simulation practice learning, and then be able to judge the effect of the whole learning process of the students more effectively, and better realize the evaluation of the entire process of the learning of the students,which fills in the gaps of the current assessment and prediction of the learning effect of the factors on the virtual simulation of learning.

At the same time,the model proposed in this paper has great potential for application in the future. Since the initial training of the model is based on the training and testing of different courses,the model can be well generalized to various classes for the same students to carry out the corresponding intelligent evaluation, which also makes the model well used in the process of student learning evaluation and then pushes personalized course learning programs for students to address the problems encountered by the students in the process of learning. The model can be applied to students' process learning evaluation and push personalized course learning plans for students to solve the pain points and difficulties encountered by students in the process of learning and improve students' academic performance; at the same time, the model prediction results can be pushed to the teacher's end in a more timely manner, so that the teacher can also achieve a timely understanding of the students' learning situation and control, and then to do the teaching to enhance the overall learning atmosphere and learning level of the class.

In addition,there is still room for further improvement in this study. In the researcher's previous research work, it is found that the level of academic performance is not only affected by the student's learning behavior data but also has a specific correlation with the students' background factors, such as family factors, metacognitive ability factors, learning strategy factors. Consequently,there still needs to be a gap in evaluating and predicting learning ability's learning effect. In the following research step, we will optimize the questionnaire,collect and accumulate more data,and then construct a reasonable neural network model algorithm to incorporate the students' background factors under the Delphi hybrid course learning evaluation index system to get better prediction results. We are also committed to developing a real-time evaluation platform to provide algorithmic support for the application and promotion of evaluation and the prediction of students' learning effects.

5 CONCLUSION

Exploring how to scientifically and accurately evaluate students' learning outcomes in blended learning is an emerging but significant field. Our study firstly utilizes the Delphi method to determine the indicator system of students' learning effect evaluation, and secondly uses statistical analysis methods to further validate the evaluation system with data, and discovers the reasonableness of using the Delphi method for indicator screening as well as the importance of the virtual simulation factors on academic performance. Moreover, the two-stream information fusion learning effect evaluation model constructed by combining temporal and non-temporal features can show the best performance on the predictive effect compared with other models, which shows that the integration of temporal data has a significant impact on the performance improvement of the learning effect evaluation model. Our study reflects that combining expert perspective and multimodal data for learning effect evaluation and prediction can produce more scientific and accurate evaluation results, which provide specific theoretical and applied values for the further development and promotion of blended teaching and corresponding learning effect evaluation and prediction.

REFERENCES

[1] PARACK S,ZAHID Z,MERCHANT F. Application of data mining in educational databases for predicting academic trends and patterns[C]//2012 IEEE International Conference on Technology Enhanced Education(ICTEE). IEEE,2012:1-4.

[2] DIEN T T,DUY-ANH L,HONG-PHAT N,et al. Four Grade Levels-Based Models with Random Forest for Student Performance Prediction at a Multidisciplinary University[C]// Complex,Intelligent and Software Intensive Systems:Proceedings of the 15th International Conference on Complex,Intelligent and Software Intensive Systems(CISIS-2021). Springer International Publishing,2021:1-12.

[3] EL AHRACHE S I,BADIR H,TABAA Y,et al. Massive Open Online Courses:a new dawn for higher education? [J]. International Journal on Computer Science and Engineering,2013,5(5):323-327.

[4] ZHAO I,CHEN K,SONG J,et al. Academic performance prediction based on multisourle,multifeature behavioral data[J]. IEEE Access,2020,9:5453-5465.

[5] HUNG J L,SHELTON B E,YANG J,et al. Improving predictive modeling for at-risk student identification:A multistage approach[J]. IEEE Transactions on Learning Technologies,2019,12(2):148-157.

[6] AKÇAPINAR G,ALTUN A,AŞKAR P. Using learning analytics to develop early-warning system for at-risk students[J]. International Journal of Educational Technology in Higher Education,2019,16(1):1-20.

[7] RODRÍGUEZ-HERNÁNDEZ C F,MUSSO M,KYNDT E,et al. Artificial neural networks in academic performance prediction:Systematic implementation and predictor evaluation[J]. Computers and Education: Artificial Intelligence,2021,2:100018.

[8] YANG Z,YANG J,RICE K,et al. Using convolutional neural network to recognize learning images for early warning of at-risk students[J]. IEEE Transactions on Learning Technologies,2020,13(3):617-630.

[9] MIN W,FRANKOSKY M H,MOTT B W,et al. DeepStealth:Game-based learning stealth assessment with deep neural networks[J]. IEEE Transactions on Learning Technologies,2019,13(2):312-325.

[10] HASHIM A S,AWADH W A,HAMOUD A K. Student performance prediction model based on supervised machine learning algorithms[C]// IOP Conference Series:Materials Science and Engineering. IOP Publishing,2020,928(3):032019.

[11] CHITTI M,CHITTI P,JAYABALAN M. Need for interpretable student performance prediction[C]// 2020 13th International Conference on Developments in eSystems Engineering (DeSE). IEEE,2020:269-272.

[12] KUO Y H,CHEN J N,JENG Y L,et al. Real-time learning behavior mining for e-learning[C]// The 2005 IEEE/WIC/ACM International Conference on Web Intelligence(WI'05). IEEE,2005:653-656.

[13] HORN D,KISS H J. Which preferences associate with school performance? —Lessons from an exploratory study with university students[J]. PloS one,2018,13(2):e0190163.

[14] CUADRA-PERALTA A,VELOSO-BESIO C,MARAMBIO-GUZMAN K,et al. Relationship between personality traits and academic performance in university students[J]. Interciencia,2015,40(10):690-695.

[15] CONARD M A. Aptitude is not enough:How personality and behavior predict academic performance[J]. Journal of Research in Personality,2006,40(3):339-346.

[16] LI X,ZHANG Y,CHENG H,et al. Student achievement prediction using deep neural network from multi-source campus data [J]. Complex & Intelligent Systems,2022,8(6):5143-5156.

[17] FENG S,WONG Y K,WONG L Y,et al. The Internet and Facebook usage on academic distraction of college students[J]. Computers & Education,2019,134:41-49.

[18] CAO X,MASOOD A,LUQMAN A,et al. Excessive use of mobile social networking sites and poor academic performance:Antecedents and consequences from stressor-strain-outcome perspective[J]. Computers in Human Behavior,2018,85:163-174.

[19] MUUSSES L D,FINKENAUER C,KERKHOF P,et al. A longitudinal study of the association between compulsive internet use and wellbeing[J]. Computers in Human Behavior,2014,36:21-28.

[20] MALHI P,BHARTI B,SIDHU M. Use of electronic media and its relationship with academic achievement among school going adolescents[J]. Psychological Studies,2016,61:67-75.

[21] JEONG A. A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication[J]. Distance Education,2005,26(3):367-383.

[22] BROOKS C A,THOMPSON C,TEASLEY S D. Towards A General Method for Building Predictive Models of Learner Success using Educational Time Series Data[C]// LAK Workshops. 2014.

[23] OKOLI C,PAWLOWSKI S D. The Delphi method as a research tool:an example,design considerations and applications[J]. Information & Management,2004,42(1):15-29.

[24] GRAVES A,SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM networks[C]// Proceedings of 2005 IEEE International Joint Conference on Neural Networks. IEEE,2005:2047-2052.

[25] TAMURA R. Multivariate nonparametric several-sample tests [J]. The Annals of Mathematical Statistics,1966:611-618.

[26] HOTELLING H. Relations between two sets of variates[M]// Breakthroughs in Statistics:Methodology and Distribution. New York,NY:Springer New York,1992:162-190.

XIE Hui,born in 1986,Ph.D,professor. His main research interests include brain functional cognition and multi-modal data online learning effect multi-dimensional evaluation.

CHEN Xueli,born in 1984,Ph.D,professor. His main research interests include advanced biomedical photonics imaging and detection technology,intelligent analysis of brain cognitive function combining behavior-ability-physiological data,and intelligent health platform based on non-contact vital signs intelligent perception.

(责任编辑:杨雪敏)