Journal of Nonlinear and Convex Analysis
Volume 26, Number 6, 2025, 1841–1860

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

LEARNING STATUS PREDICTION METHOD BASED ON TRANSFORMER MODEL FOR SPORTS MAJOR STUDENTS

GUOBIAO YANG, WANYING YANG, AND YONGXIANG WANG*

Abstract. Accurately predicting the learning status of sports major students plays an important role in personalized teaching. It prevents learning obstacles, enhances learning motivation, and optimizes the allocation of educational resources, directly influencing educational quality and comprehensive development of students. The commonly used methods, such as online learning behavior and cognitive psychology analyses, currently face challenges in data collection and processing, insufficient application and feedback of prediction results, and mismatch between technological implementation and actual needs. Therefore, this study investigated the learning status prediction methods using the Transformer model. It preprocessed student learning data through text data collection, segmentation and encoding, and sequence standardization. Based on these data, a core encoder and a decoder were designed to complete the construction of the Transformer model. Furthermore, the established model was trained, optimized, and evaluated by defining a loss function. Finally, the model was validated using a sophomore student from an engineering college as the research object. The verification results showed that the model could achieve good discrimination for various learning status of students and had stable predictive ability.

1. Introduction

The increasing awareness toward physical health in society, coupled with the continuous development of the national sports industry, has enhanced the significance of nurturing sports talents in higher education. The establishment of sports major students not only reflects the diversified development of higher education but also helps nurture a significant number of professional sports talents for society, especially in comprehensive universities. However, the learning process for sports major students is extremely complex, involving not only the study of cultural courses but also the simultaneous acquisition of sports skills. Skillfully balancing these two aspects has become the key to enhancing the learning efficiency and overall quality of sports major students.

Sports major students face several challenges in their academic development. First, balancing between academic knowledge and physical training is a major challenge for many sports major students. These students need to not only study traditional theoretical courses such as exercise physiology, sports psychology, and sports

2020 *Mathematics Subject Classification.* 68T05, 68T07.

Key words and phrases. Information attenuation perception, learning status prediction, multi-head attention mechanism, personalized education, transformer model.

This study was supported by Regular Scientific Research Project of Shaanxi Provincial Sports Bureau (2023012) and the Innovation Fund of Xidian University (YJSJ24018).

Key Research and Development Program of Shaanxi (Program No. 2025SF-YBXM-146).

*Corresponding author.

nutrition but also invest a considerable amount of time in physical training and competitive events. The demanding nature of such high-intensity physical training and practical activities often leaves less time for theoretical studies, impacting students ’ academic performance and the depth of their knowledge acquisition.

Second, personalized learning needs pose another significant challenge for sports major students. The traditional teaching methods and standardized educational models often fail to address the individual needs of these students due to variations in their physical fitness, athletic skills, and study habits. Some students may excel in certain courses but struggle in others. Providing targeted teaching and tutoring based on the diverse needs of students has become a pressing issue for teachers to address. The traditional one-size-fits-all teaching method currently employed in universities for sports majors cannot resolve the conflict between academic learning and physical training. Furthermore, this teaching method is even more inadequate in meeting their personalized development needs due to individual differences among students.

Abandoning the traditional one-size-fits-all teaching method and adopting more precise teaching strategies is imperative to achieve personalized education. Precise teaching requires educators to deeply understand the specific learning status and cognitive style of each student, enabling them to tailor teaching content and intensity to best suit individual needs. Currently, a new round of technological revolution represented by artificial intelligence (AI) has been deeply embedded in the learning and life of college students. However, it is changing their learning methods and cognitive styles from various aspects. However, many universities still adopt a unified teaching method of syllabus and textbooks for the basic compulsory physical education courses of the entire college or even the entire school to facilitate teaching management. This one-size-fits-all traditional teaching method seriously ignores the huge differences in the learning foundation, cognitive level, learning style, and personal interests of students, especially in the learning management and academic prediction of sports major students. Consequently, the students are unable to find a suitable rhythm in learning. This not only reduces teaching efficiency but may even weaken the learning interest of students, especially for those whose personal learning abilities do not align with the teaching content and speed. Students with strong learning abilities may also lose motivation to learn if the learning content is too simple.

Teachers should accurately grasp the learning status of each student to achieve personalized education. They can effectively predict the current and future learning development directions, identify potential learning challenges, and then adopt the most appropriate teaching methods based on the specific needs of students by delving into the basic data of their learning activities. It not only enhances the learning interest and efficiency of students but also provides timely assistance when students encounter difficulties, protects their learning interest and continuous learning motivation, improves the personalization and accuracy of education, and ultimately promotes educational equity through technological power, thus providing scientific data support and decision-making basis for comprehensively improving educational effectiveness and quality. However, the learning status of students is currently predicted using qualitative research methods such as artificial online learning behavior and cognitive psychology analyses. Obtaining quantitative prediction results is difficult, hindering the future implementation of personalized education. The advent of the era of Big Data, coupled with rapid advancements in Internet technology and campus informatization, has led to the continuous generation of massive data covering the learning behavior, life trajectory, and psychological behavior of students. These massive and targeted data provide new perspectives and methodologies for analyzing the learning status of college students.

The development of AI technology has provided new tools and methods for addressing these challenges. Traditional machine learning algorithms, such as support vector machine, decision tree, random forest (RF), and so forth, have been applied to predict students ’ learning status. However, these methods face challenges such as cumbersome feature extraction processes and limited model expression capabilities, making it difficult to effectively capture the intricacies of student learning behavior. The emergence of deep learning technology in recent years has brought new opportunities for learning status prediction. The Transformer model is a deep neural network model based on a self-attention mechanism. It can effectively capture long-range dependencies in sequential data and has strong feature extraction and model expression capabilities, making it suitable for predicting student learning status [5].

Since its launch in 2017, the Transformer model has achieved significant success in natural language processing. Its core feature, the self-attention mechanism, enables it to focus on any part of a sequence when processing sequential data, thus effectively capturing long-range dependencies. This ability is particularly suitable for analyzing complex student learning data with time-series characteristics, even in situations where the data are incomplete and can be effectively analyzed [19]. Therefore, the Transformer model has immense potential for analyzing and processing learning behavior data in the field of education. In February 2008, Yahoo introduced Hadoop technology to improve the efficiency of massive data retrieval, applying it to internal server clusters [1]. This technology has promoted the development and application of data processing systems for large-scale data processing. Data storage and analysis have become a hot research topic in the academic community after the development of information technology and breakthroughs in data processing technology. In terms of business, Facebook [23] has integrated Hadoop with technologies such as Hive and MapReduce to build a cluster environment for processing large amounts of data. This approach supports tasks such as analyzing user activity, status updates, and logs. At present, Facebook uses a fully distributed Big Data platform for data storage, management, and mining [16]. In addition, the academic community has shown significant interest in Hadoop technology [12]; many research institutions and universities have begun to study Hadoop clusters and offer related courses [13]. Judijanto uses AI systems to understand the unique preferences, needs, and learning styles of individuals through data analysis and intelligent algorithm modeling; this approach prioritizes personalized learning and enables teachers to present teaching materials more effectively tailored to the comprehension levels of students [11]. Some universities have also applied Hadoop technology for research on student behavior data mining [30]. For example, University of Indianapolis designed and developed a “hotspot map” to monitor student activity hotspots [14].

These measures have effectively improved the development of Big Data technology in the business and academic fields. The learning behavior analysis system designed by Lake First University in Canada [3] focuses on analyzing the learning behavior of students. The schools can use this software to understand students ' recent learning progress and provide tailored guidance via email. When analyzing the learning behaviors of students in smart classrooms, Wang established an end-to-end framework incorporating a deformable attention mechanism to improve target detection capabilities, enhance the capture of detailed behavioral features, simplify the detection process, and deliver test results superior to those of other deep learning methods [29]. In addition, some universities [25] mainly analyze and mine the academic performance data accumulated by students during their time in school. They provide professional advice to students based on this data, predict students who may exhibit abnormal behavior, and issue timely warnings to provide appropriate guidance to students [27].

Research on using Transformer models for predicting student learning status started relatively late in China but has progressed rapidly in recent years. Many universities and research institutions have recognized the potential of Transformer models in educational data analysis and conducted a series of related studies [17, 22, 31]. Central China Normal University has made significant progress in using the three-space fusion intelligent learning system to innovate the mechanism and path of promoting information technology strategy, promote the deep integration of information technology and education and teaching, enhance the deepening application of information technology in school business management and services, and support the internationalization development strategy of the school. The Smart Learning Experimental Zone and Experimental School of Beijing Normal University provide innovative experimental venues for education and teaching [4, 35]. They explore the optimization and innovation of the teaching process and promote the development of teaching models toward a more intelligent, personalized, and interactive direction by introducing new technologies, methods, and concepts [26]. These experimental areas and schools provide a practical foundation and data support for educational and teaching research. Important references and support are provided for educational theory and practice by observing and researching student learning behavior and teacher effectiveness [34]. The Joint Laboratory of Artificial Intelligence and Education Integration Innovation at Northeast Normal University is committed to integrating AI technology with education and conducting cutting-edge educational technology research and innovative practices [7, 20]. It explores new theories, methods, and models in education and promotes innovative development in education and teaching by integrating technologies such as AI, machine learning, and data mining [32]. The research team at Tsinghua University [21] has also leveraged advanced technologies such as AI to promote the diversified development of teaching modes, streamline student management, and improve the personalization and innovation of teaching content. It has modeled students' personalized learning characteristics using knowledge tracing techniques to collect and analyze student behavior data. This approach enables dynamic tracking and prediction of student ' s knowledge mastery, [33]. In addition, Shanghai Jiao Tong University [10], Shandong University [24], Peking University [9], Zhejiang University [28], South China

Normal University [15], Renmin University of China [6], Beihang University [18], and others have all achieved various research results.

In summary, the use of AI technology to predict the learning status of students has significantly promoted the innovation of educational technology [2]. However, student image data, learning behavior data, and life trajectory data cannot be collected in a regular manner compared with standardized and structured time-series data [8]. Also, the collection times and intervals for different types of data and even for various features within each type are inconsistent. This inconsistency leads to challenges such as missing data and misaligned dimensions across time. In addition, the nonuniform time intervals in various dimensions have resulted in uneven spacing between various data points in the sequence data. In summary, the temporal data of sports majors students examined in this study had the following significant characteristics: (1) data loss caused by uneven sampling time in each dimension of various types of data and uneven loss in each dimension; (2) inconsistent time intervals between data points in time-series data caused by the inability to control them in terms of time. This issue is also a core reason why AI-driven advancements in higher education remain limited. This study addressed the aforementioned issue by exploring the implementation path of the learning status prediction of sports major students based on the information attenuation perception Transformer model. It employed the advantages of the Transformer model in handling sequence dependencies. A special information attenuation mechanism was added when analyzing input information data, which not only effectively mitigated long-term dependency issues in the data but also improved the sensitivity and recognition ability of the model to key behavioral features. Finally, the effectiveness of the established model was demonstrated through practical application cases, including analyses of training loss, final accuracy, and predictions of the learning status of excellent, average, and poor students. These applications verified the model ' s accuracy and utility in real educational scenarios.

2. Establishment of transformer model based on information attenuation perception

We collected and organized data on the learning status of students and then conducted feature analysis. These data exhibited the characteristics of multimodal low-density sequential data. We selected a relatively suitable model framework based on these data features. The system design scheme proposed in this study covered data preprocessing, model design, training and evaluation, and ways to deploy and use the system. It specifically included three basic parts: data collection and processing, model design, and model training and evaluation.

Data collection and Processing

The steps involved in data collection and processing were as follows:

(1) Collect data on student sex, attendance rate, competition participation, number of failed subjects, classroom test scores, and Cattell 16PF personality factors.

(2) Convert all non-numerical data into numerical data, such as the 16PF factor.

(3) Standardize numerical features to eliminate dimensional effects and ensure the effectiveness of model training;

(4) Divide the dataset into 80% training set and 20% testing set to evaluate the generalization ability of the model.

Transformer Model Design Process

(1) Use a Transformer structure including multiple encoding layers, each layer including a multi-head self-attention mechanism and a feedforward network.

(2) Configure a self-attention mechanism to allow the model to capture long-distance dependencies between features.

(3) Map the output of the last Transformer layer to the final classification result through a fully connected layer.

Model Training and Evaluation

(1) Choose the appropriate loss function and optimizer.

(2) Apply gradient cropping technology to prevent gradient explosion.

(3) Record the loss and accuracy of the training set at the end of each training cycle, and plot the loss and accuracy curves.

Simplification of Personality Trait Scores

Individual personality traits are often complex and difficult to distinguish in education and learning environments. This study classified these personality factors into three levels based on the evaluation results to determine their potential impact on learning. This classification helped more accurately analyze the behavioral motivations of students and provide more accurate conclusions.

The study categorized the results of 16PF into three levels: beneficial for learning, neutral impact, and potentially hindering learning, as shown in Figure 1. This classification aimed to help the program identify which personality traits might promote student learning, which might not have a significant impact, and which might have a negative impact on learning, and also provide a rough rating of the personality of the participants.

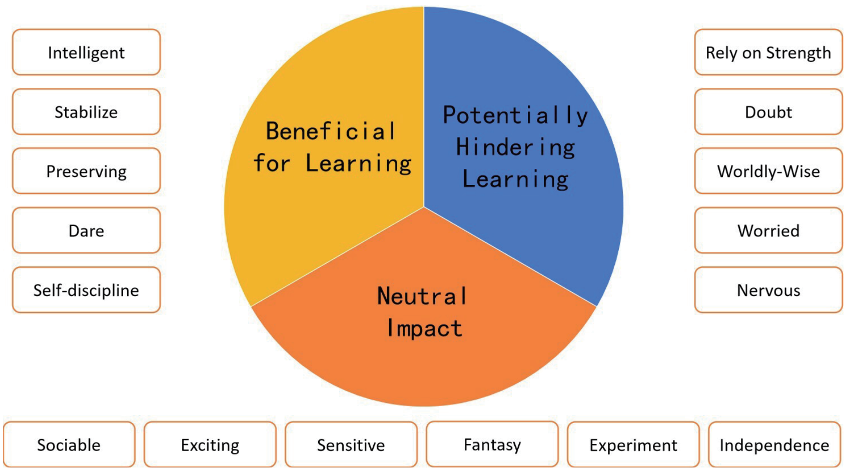

FIGURE 1. The impact of different personality traits on learning outcomes.

Datasets and Preprocessing

The dataset used in this system comprised the following fields:

StudentId: Student's ID number.

Sex: Sex of the student, where 0 represents female and 1 represents male.

Attendance rate: Attendance rate of a student, expressed as a percentage.

Challenge: Number of times a student participates in subject-related competitions.

Failed: Number of subjects a student has failed in their area of study.

Quiz: Average score of a student's test.

16PF (personality factor score): Personality trait score of students based on the Cattell ' s 16 Personality Factor Questionnaire.

Score: Comprehensive rating of a student, used to measure their learning ability.

We deliberately simulated the situation of data loss, such as intentionally removing some students' test scores or personality trait ratings, to address the potential missing data in the dataset. Several common strategies were adopted, including using the mean, median, and mode of the fields for filling and ultimately selecting the average filling strategy that performed the best in accuracy testing, on facing missing numerical data (as shown in Table 1), ensuring the accuracy of data processing and the reliability of the prediction model.

Table 1. Training loss and accuracy with different imputation strategies over 10 cycles

	Original values	Mean	Median	Mode
Training loss	0.1260	0.1417	0.1477	0.6365
Accuracy(%)	88.06	83.58	80.60	79.10

Datasets and Preprocessing

This study evaluated the performance of the model on an independent dataset by dividing the data into training and testing sets, with 80% of the data used for training and the remaining 20% used for testing the performance of the model. Such partitioning helped evaluate the generalization ability of the model on unseen data, ensuring its robustness.

The data preprocessing included missing value filling, data standardization, and partitioning the dataset into training and testing sets. The quality and consistency of the data were ensured through these preprocessing steps, laying a solid foundation for subsequent model training and evaluation.

Model Definition

The timeliness of information is crucial for temporal data. Data with varying delays inevitably lead to a decline in their credibility. However, temporal models such as RNN (Recurrent Neural Network) cannot perceive the aforementioned nonuniformity and information loss. Data filling and other operations are also required to adapt to the structure of the model, which has a significant impact on the performance of the model in predicting and detecting the learning status of students. In addition, people's demand for the interpretability of models in the field of machine learning, especially deep learning, has become increasingly evident. In the context of this project, it is imperative to not only perceive, detect, and predict the learning status of students but also help "students with learning difficulties." This requires the model to reasonably express the learning barriers of students based on the output results by calculating the contributions of input data from various dimensions to those results. Furthermore, the model must enable teachers, counselors, learning mentors, and other personnel to intervene effectively in addressing the learning challenges faced by students.

Therefore, a Transformer model for information attenuation perception was proposed based on the inherent characteristics of student information temporal data and the objective need for interpretability, targeting the characteristics of multimodal low-density temporal data of students. The framework structure is shown in Figure 2.

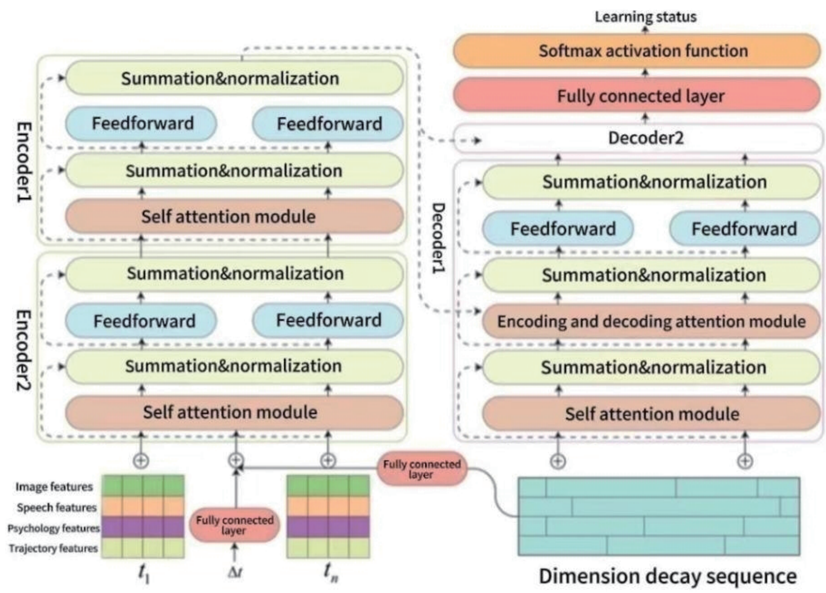

FIGURE 2. Model framework based on information attenuation perception Transformer.

3. Training of transformer model based on information attenuation perception

Model Training

Training is used to ensure that the model can learn features from the training data and has good generalization ability on new data. The key steps in the model training process include the selection of loss function and optimizer, gradient pruning, and model iteration.

(1) Definition of loss function

Choosing a loss function is one of the key decisions in machine learning because it directly affects the objectives of optimization algorithms. The cross-entropy loss function is widely used in classification tasks to evaluate the difference between the model output and the actual label. Cross-entropy can directly quantify the inconsistency between the predicted probability distribution and the true distribution, making the goal of model optimization straightforward. When the prediction of the model deviates from the true label, the gradient size provided by cross-entropy loss is appropriate, which helps quickly correct incorrect predictions and effectively improve the accuracy of predicting student learning status. The expression of the cross-entropy loss function is:

$$L = -\sum_{i=1}^{C} t_i \log(p_i) \tag{3.1}$$

where C is the total number of categories, ti is the true label of the j-th category, and p is the probability of the model predicting the j-th category. The student learning status examined in this study belonged to a classification task for which the cross-entropy loss function was more suitable.

Considering the large gradients provided by the cross-entropy loss function during misclassification, its use can help the model learn faster and accelerate the convergence process. Additionally, cross-entropy loss was less prone to gradient vanishing than with squared error loss due to the use of the softmax function in this study, making it particularly effective for training deep networks..

(2) Selection of optimizers

After careful evaluation, this study used Adam (adaptive moment estimation) as the optimizer. The data on student learning status are characterized by complexity and imbalance. The Adam optimizer combines the core advantages of Momentum and RMSprop optimizers and has the function of adaptive learning rate adjustment. It automatically adjusts the learning rate for each parameter based on the recent gradient changes of the parameters and uses this information to accelerate the optimization process of parameters while reducing directional oscillations. The applicability of the Adam optimizer was also verified in this study through comparative analysis.

The uniqueness of the Adam optimizer lies in its ability to automatically adjust the learning rate for each parameter based on the size of the nearest gradient change of the parameter, making it suitable for dealing with the problem of uneven data updates [29]. This was particularly crucial for analyzing the complexity of student learning status in this study. Also, these features of the Adam optimizer enable it to perform well in various data scenarios, especially in applications that predict student learning status, effectively handling various challenges.

The adaptive learning rate mechanism of the Adam optimizer significantly optimizes the initial speed of model training, enabling it to quickly approach the optimal solution in the early stages. It also automatically reduces the learning step size when approaching the optimal solution, thus enhancing the stability of the entire training process. The Adam optimizer helps alleviate the problem of overfitting by dynamically adjusting the learning rate, which is particularly crucial when dealing with unknown datasets. It ensures that the model performs more reliably and effectively on new data. In addition, the tuning mechanism of the Adam optimizer ensures that the model maintains good performance under various data conditions, especially in learning status prediction involving complex data structures. Its application can significantly enhance the accuracy of predictions and improve the model' s generalization ability.

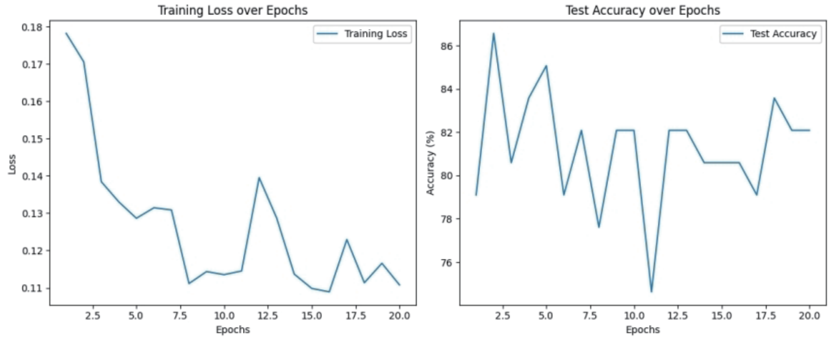

FIGURE 3. Application effect of the cross-entropy loss function and Adam optimizer.

As shown in Figure 3, the average training loss obtained by using the cross-entropy loss function and Adam optimizer in the 15th to 20th cycles after stable training was 0.19, and the average accuracy obtained was 81.92

(3) Gradient cropping definition

Gradient cropping is used to ensure the stability of the model training process. It avoids excessive gradient updates by limiting the maximum norm of gradients, thus preventing numerical instability and gradient explosion problems during training.

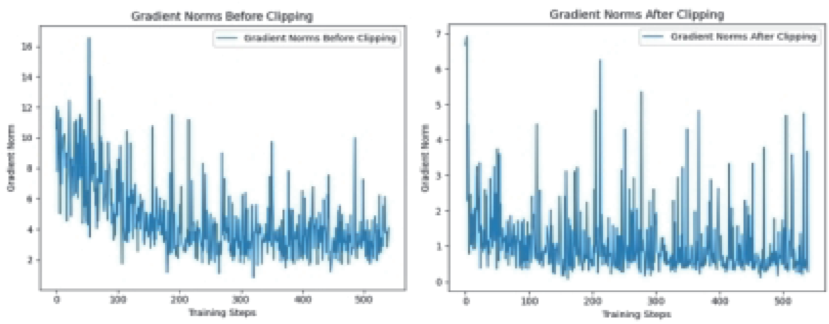

FIGURE 4. Changes in gradient norm with training before and after gradient cropping.

Gradient cropping has a significant impact on the gradient norm during training. This study applied gradient clipping to reduce the gradient norm. The specific parameter settings are shown in Figure 4. A comparison revealed that the gradient norm used to be as high as 16 in the first 100 training steps before gradient pruning, and the final gradient norm was between 1 and 11. The gradient norm fluctuated significantly and could potentially explode. The gradient norm was limited to a reasonable range and remained within 7 during the initial training phase by continuously adjusting the clipping parameters after gradient cropping, eventually stabilizing between 0 and 5, making training more stable.

(4) Training

Training cycle

The training usually includes multiple cycles. This study used 40 cycles for training. The model traversed the entire training dataset once in each cycle and learned through multiple iterations.

The steps in each cycle were as follows:

a. Model training mode settings

At the beginning of each training cycle, the model enters the training mode by calling "model. train()" to ensure that all settings are correctly applied to the learning process. Specific layers such as Dropout and BatchNorm exhibit different behavioral characteristics in this mode than during evaluation, which helps prevent overfitting and adjust normalization parameters based on each batch of data. In addition, the weights and biases of the model are updated based on the loss function feedback and backpropagation results exclusively during the training mode.

b. Batch iteration:

The DataLoader class is used in the code to create a batch iterator, which divides the dataset into multiple batches, each containing a specified number of data samples. The specific implementation code is as follows:

train_loader = DataLoader(train_dataset, batch_size=10, shuffle=True)

Batch size is an important hyperparameter because it affects the efficiency and effectiveness of model training. In this study, the batch size was set to 10, representing the processing of 10 data samples per batch, which ensured its effectiveness while accelerating training efficiency.

Shuffle is set to True in the training data loader, implying that the data is randomly shuffled at the beginning of each training cycle. This helps the model avoid learning the order of data, thereby improving its generalization ability.

4. Verification of transformer model based on information attenuation perception

We selected second-year students from a university as the experimental subjects after completing the model building and training, aiming to deeply explore the accuracy and effectiveness of the Transformer model in predicting the learning status of sports major students,. A total of 30 participants with stable academic performance and good mental health were strictly selected from these students to ensure the rigor and reliability of the experiment. The participants were evenly assigned to two groups: experimental and control groups. Prior to starting the experiment, we completed the construction and training of the Transformer model, which was designed to receive and process multidimensional core parameters including sex, attendance rate, competition, failing grades, test scores, and personality trait scores. The Transformer model could quickly and comprehensively assess the learning status of each participant using these parameters.

During the experiment, we strictly controlled all external factors that could influence the results, such as additional exposure to learning materials and unsystematic information input, to ensure the integrity of the experimental environment and the accuracy of the data. At the same time, we established a fixed testing schedule, conducting tests every other day within a consistent time frame and recorded the results of each test in detail. Statistical tools were used to analyze and interpret the collected data in detail after 4 weeks of continuous testing and data accumulation. The results of the experiment are shown in Table 2. We believe that the Transformer model is highly accurate in assessing the effectiveness of personalized learning resources for students.

TABLE 2. Data processing results[1]

Method	Excellent Students			Average Students			Poor Performing Students			rsum
	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	
SCAN*	67.4	90.3	95.8	58.0	84.0	90.5	48.6	77.7	85.2	697.5
BFAN*	68.1	91.4	-	59.5	84.9	-	50.8	78.4	-	-
DPRNN	70.2	91.6	95.8	62.9	86.5	92.0	55.5	81.3	88.2	724.0
CAAN	70.1	91.6	97.2	61.5	85.3	92.6	52.8	79.0	87.9	718.0
GSMN*	76.4	94.3	97.3	66.9	88.3	93.2	57.4	82.3	89.0	745.1
SHAN*	74.6	93.5	96.9	65.0	87.4	92.7	55.3	81.3	88.4	735.1
SGRAF*	77.8	94.1	97.4	68.2	88.6	93.1	58.5	83.0	88.8	749.5
IAPTM*	**77.2**	**94.1**	**96.0**	**68.0**	**89.7**	**93.9**	**58.8**	**85.3**	**91.7**	**754.7**

Seven typical technical solutions were selected for comparative validation to demonstrate the superiority of the technical solution proposed in this study: SCAN, BFAN, DPRNN, CAAN, GSMN, SHAN, and SGRAF. Core indicators were calculated separately for three types of students based on the student experimental samples provided in this study: excellent students, average students, and poor-performing students. Then, the data obtained using the technical scheme proposed in this study were compared with the data results from the seven typical technical schemes. As shown in Table 2, the technical scheme IAPTM proposed in this study demonstrated better recognition and matching effects in terms of the three indicators R@1, R@5, and R@10 for identifying the three types of students, validating the technical advancement of the proposed IAPTM scheme.

Model Definition and Loading

This study used the Transformer encoder layer provided by PyTorch to process input features and output the original score of the category through linear classification. Simultaneously, we prepared a temperature parameter for calibrating the confidence level of softmax output. We loaded the previously trained model parameters and switched the evaluation mode using model.eval().

Data Preprocessing

We used Pandas to read the data from the CSV file and StandardScaler to standardize the feature data to a mean of 0 and a standard deviation of 1. Then, we split the dataset into training and testing sets, with the testing set accounting for 20%. We customized the PyTorch Dataset class streamline data loading and processing. Further, we added a dimension through unsqueeze (0) so that each feature tensor had a batch dimension.

Model Prediction

[1]SCAN(Semantic Compositional Attention Networks), BFAN(Bidirectional Focal Attention Network), DPRNN(Dual-Path RNN), CAAN(Context-Aware Attention Network), SHAN(Sequential Hierarchiacl Attention Network), SGRAF(Semantic Graph Attention Networks), IAPTM(Information Attenuation Perception Transformer Model)

We normalized the data and converted it into a PyTorch tensor. We used the softmax function to convert the original score output by the model into a probability distribution and finally obtained the category with the highest probability and its corresponding probability (i.e., confidence) torch.max.

User Input and Prediction

We prompted the user to input feature data and read the data, used the model for prediction, and finally output the prediction result and confidence.

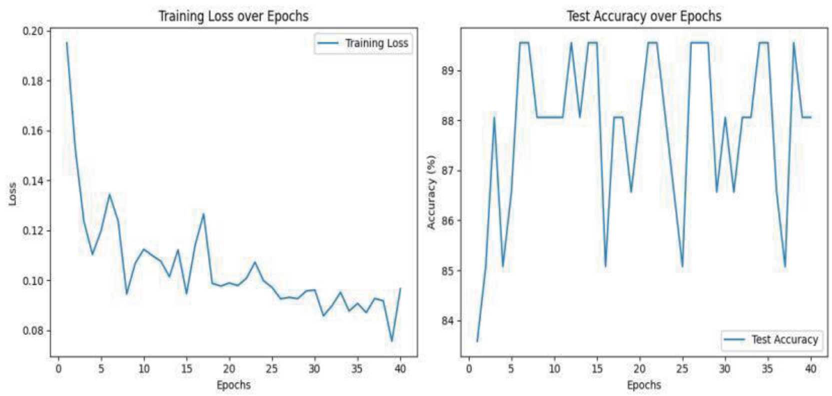

FIGURE 5. Training loss and accuracy of the final code with complete data.

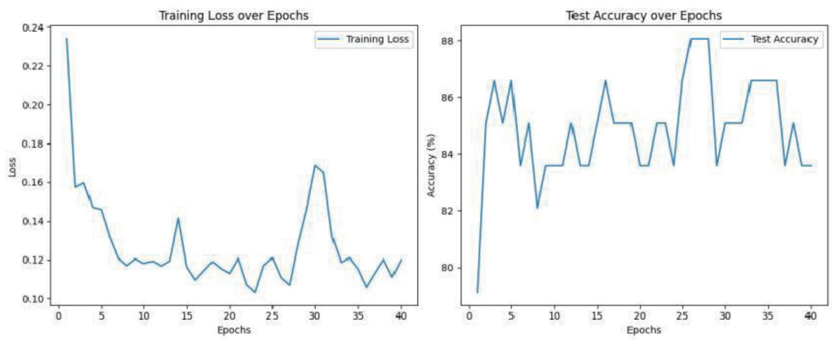

FIGURE 6. Training loss and accuracy of the final code with 10% missing data.

Comparing Figures 5 and 6, the left side shows the training results with complete data, whereas the right side shows the results with 10% missing data, indicating a significant improvement compared with the results tested in the previous section. When using this model for prediction in practice, the "torch.max" function is used to obtain the maximum probability as the confidence level. This model maintains a high confidence level, whether facing extreme or moderate data, indicating its high reliability.

As shown in Figure 7, a basic student learning status prediction system was established using the Transformer model established in this study. In the input

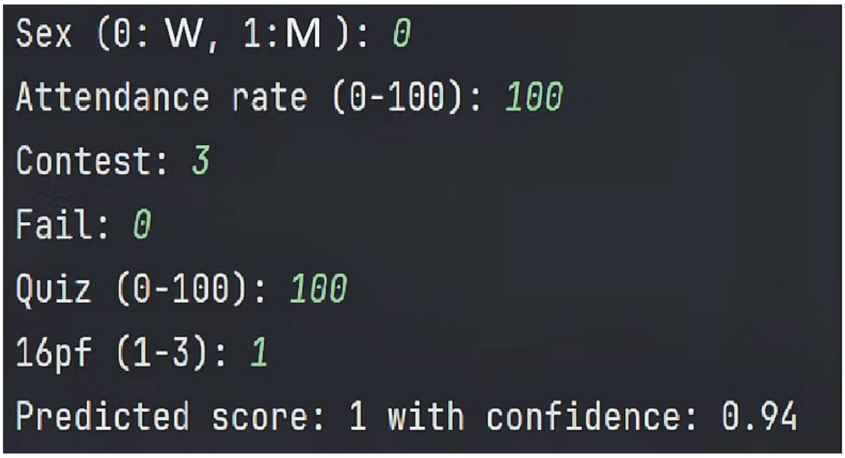

FIGURE 7. Test results of excellent students predicted using the model established in this study.

information, sex represented sex, with input 0 representing female and input 1 representing male. Attendance represented the attendance rate, with input values ranging from 0 to 100, reflecting the percentage of students attending classes. Contest represented the number of students participating in subject-related competitions, with the input value being a natural number. Fail represented the number of failed subjects in the student's previous related subjects, with the input value being a natural number. Quiz represented the average score achieved by students in the test, with input values ranging from 0 to 100. The 16PF was the score obtained from the personality trait test of the student, which was replaced by 1, 2, and 3 to suit learning, neutral impact, and potential negative impact, respectively. As demonstrated in this study, a female student with a 100% attendance rate, who participated in three related competitions, did not fail the course, scored full marks in daily tests, and had a good personality suitable for learning, was inputted into the system. The predicted result was 1 and the confidence rate reached 94%, indicating that the student's learning status was at a relatively excellent level and the model prediction was relatively accurate.

As shown in Figure 8, a poor-performing female student with an attendance rate of 12%, who did not participate in relevant competitions, passed three related subjects, and had an average score of 15 in daily tests, was inputted into the system. Personality may also have a negative impact on learning. The data input of the student with overall level deficiency into the system yielded a prediction result of 0, with a confidence level of 88%, indicating that the student's learning status was at a level needing improvement and the model prediction was also relatively accurate.

As shown in Figure 9, an average female student with an attendance rate of 87%, who participated in a related competition, passed a related subject, and had an average score of 78 in daily tests, was inputted into the system. The predicted result was 1 with a confidence level of 94%, indicating that the student's learning status was excellent and the model prediction was also accurate.

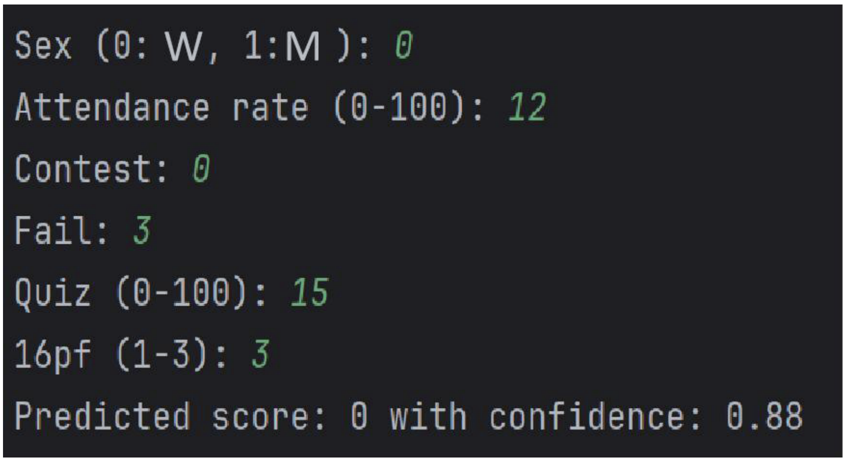

FIGURE 8. Test results of poor-performing students predicted by applying the model established in this study.

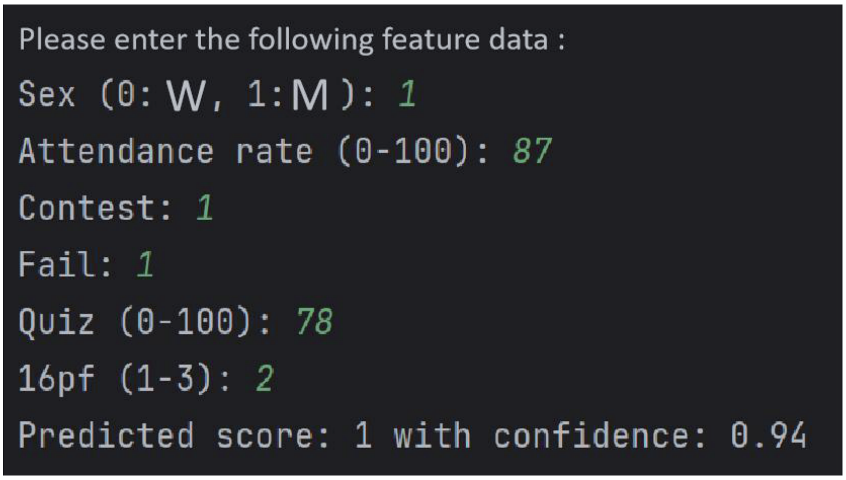

FIGURE 9. Test results of average students predicted by applying the model established in this study.

The verification and application of learning status prediction based on the information attenuation perception Transformer model were completed through the aforementioned specific examples, demonstrating the entire process from strict evaluation of the model to practical application. Also, the accuracy and robustness of the model constructed in this study were verified through rigorous evaluation on an independent test set, ensuring that the model could provide reliable predictions in practical applications.

In terms of practical applications, a simple and intuitive user interface was also presented, allowing users to input relevant educational feature data and quickly obtain predicted learning status results. The model demonstrated a high accuracy of 88.01% after rigorous testing when the data were complete and after 40 training cycles, with a loss function reduced to 0.0940. The model could still maintain an accuracy of 84.19% with a loss function of 0.1221 even in the case of 10% missing data, demonstrating its superior data fault tolerance and ensuring high accuracy in subsequent practical testing.

Model Performance Evaluation

Validation was performed on the Flicker30k dataset to demonstrate the superiority of the student learning status prediction based on the information attenuation perception Transformer model in mitigating the interference of noisy features during the matching process, as shown in Table 3..

TABLE 3. Results of data matching

Method	Excellent students			Average students			Poor performing students			rsum
	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	
MSCOCO 1k test										
SCAN*	72.7	94.8	98.4	65.8	91.6	96.6	58.8	88.4	94.8	761.9
BFAN*	74.9	95.2	-	67.2	91.8	-	59.4	88.4	-	-
DPRNN	75.3	95.8	98.6	68.9	92.8	96.9	62.5	89.7	95.1	775.6
CAAN	75.5	95.4	98.5	68.4	92.6	96.9	61.3	89.7	95.2	773.5
GSMN*	78.4	96.4	98.6	70.9	93.3	97.2	63.3	90.1	95.7	783.9
SHAN*	76.8	96.3	98.7	68.7	93.0	97.3	62.3	89.6	95.8	778.8
SGRAF*	79.6	96.2	98.5	71.4	93.5	97.3	63.2	90.7	96.1	786.5
IAPTM*	**79.8**	**96.3**	**98.7**	**72.0**	**93.5**	**97.3**	**64.2**	**90.7**	**95.9**	**788.4**
MSCOCO 5k test										
SCAN*	50.4	82.2	90.0	44.5	75.8	85.2	38.6	69.3	80.4	616.4
CAAN	52.5	83.3	90.9	46.9	76.8	86.9	41.2	70.3	82.9	631.7
SGRAF*	58.8	84.8	92.1	50.2	77.9	86.8	41.6	70.9	81.5	644.6
IAPTM*	**59.4**	**86.0**	**92.3**	**50.9**	**78.9**	**87.0**	**42.3**	**71.8**	**81.7**	**650.3**

IAPTM is the abbreviation of the methodology proposed in this study; R@1, R@5, and R@10 in the table are Recall@k, which are divided into three task settings of excellent students, average students, and poor-performing students, and contain a total of nine metrics data. The rsum in the last column of the table represents the sum of the recall values from the first nine columns, offering a more comprehensive assessment of the performance of the matching model.

Compared with previous methods, the IAPTM proposed in this study achieved some improvement across all indexes. Specifically, compared with the baseline model SGRAF, the improvement in the task setting of excellent students, was 2.0%, 0.2%, and 0.3% in R@1, R@5, and R@10, respectively; the most important metric, R@1, demonstrated significant improvement. In the task setting for poor-performing students, the most important metric R@1 improved by 1.6%, R@5 by 1.5%, and R@10 by 1.4%. Finally, the composite metric rsum significantly improved by 7%. Thus, IAPTM achieved a more significant performance improvement on the Flicker30k dataset, validating its effectiveness in mitigating the noise feature interference during the matching process.

5. CONCLUSIONS

The application of AI technology in predicting sports major student learning status is of great significance. It can not only improve teaching efficiency and quality and optimize resource allocation but also provide personalized learning experiences and support for students. This approach stimulates their learning motivation and promotes educational equity. Precision teaching for sports major students can tailor personalized learning plans based on the physical fitness levels, interests, and learning paces of students. Each student differs in terms of physical condition and learning progress. Teachers can provide each student with the most suitable learning content by accurately assessing students' learning status, enabling them to make continuous progress at an appropriate pace. Furthermore, teachers understand students' interests and then design course content according to their preferences. This encourages more active participation in sports activities and stimulates students' learning motivation and interests, thereby enhancing their intrinsic motivation and engagement. Additionally, feedback from students helps teachers promptly identify students encountering difficulties or progressing slowly in certain projects. Teachers can then adjust teaching methods accordingly, thus offering personalized guidance and support to ensure every student has appropriate opportunities for development. Regarding the physical health of sports major students, which is of great concern, teachers can reasonably arrange exercise intensity by understanding the physical conditions and changes in the athletic abilities of students. This prevents accidental injuries caused by overtraining or unsuitable training methods and safeguards students' physical and mental well-being.

This study proposed a sports major student learning status prediction method based on the Transformer model. The temporal relationship of student learning information data was obtained through temporal encoding, and the long-range dependency relationship of student status data was extracted using scaled dot product attention to obtain remote sequence-related information. The multi-head attention mechanism was used to learn the feature information of different positions in student data, capture the global dependency relationship, and then fuse the feature information through the fully connected layer to output the predicted result of the student learning status. The experimental results showed that the established model achieved better detection performance on real datasets.

The relevant research results of this study can provide technical support for not only personalized teaching but also the development of optimizing educational resource allocation, thus preventing learning obstacles and enhancing learning motivation. Building on our existing research, we plan to continuously advance the development of the Artificial Intelligence-Classroom Quality Evaluation system. This educational assessment system leverages AI and Big Data analysis. Using vast amounts of classroom teaching video data, the system employs deep learning techniques in AI training to intelligently recognize classroom behaviors of both sports teachers and students, encompassing face recognition, expression analysis, and so on. The results are presented to students, teachers, parents, and other educators in a more intuitive graphical format through Big Data analysis of the behaviors of teachers and students, including students' movement trajectories, classroom performance, concentration levels, and other relevant data. This provides scientific decision-making support for educational policymakers. Future studies should explore more feature integration technologies and optimize network architecture to further enhance the performance and practicality of models, bringing more reform and innovation to the education field.

References

[1] L. Alarabi, M. F. Musleh and M. Musleh, *ST-Hadoop: a MapReduce framework for spatio-temporal data*, GeoInformatica **22** (2018), 785–813.

[2] E. A. Alasadi and C. R. Baiz, *Generative AI in education and research: Opportunities, concerns, and solutions*, Journal of Chemical Education **100** (2023), 2965–2971.

[3] P. Anand, *Big data is a big deal*, Journal of Petroleum Technology **65** (2013), 18–21.

[4] J. Bernard, T. Chang, E. Popescu and S. Graf, *Learning style Identifier: Improving the precision of learning style identification through computational intelligence algorithms*, Expert Systems with Applications **75** (2017), 94–108.

[5] A. Hanif, F. Q. Jamal and M. Imran, *Extending the technology acceptance model for use of e-learning systems by digital learners*, IEEE Access **6** (2018), 73395–73404.

[6] J. Hu, H. Liu, Y. C. Chen and J. Qin, *Strategic planning and the stratification of Chinese higher education institutions*, International Journal of Educational Development **63** (2018), 36–43.

[7] X. Hu, J. Su, Y. Dai, J. Xiong, X. Yu and M. Zhang, *Research on the Application of Internet of Things and Artificial Intelligence Technology in Smart Classroom*, in: Proc. - Int. Conf. Appl. Phys. Comput., ICAPC, Institute of Electrical and Electronics Engineers Inc., Apr. 2023, pp. 418–425.

[8] E. Jando, Meyliana, A. N. Hidayanto, H. Prabowo, H. L. H. S. Warnars and Sasmoko, *Personalized E-learning Model: A systematic literature review*, in: Proc. Int. Conf. Inf. Manag. Technol., ICIMTech, Institute of Electrical and Electronics Engineers Inc., United States, Jul. 2017, pp. 238–243.

[9] J. Jia, *Design, implementation and evaluation of blended learning for the undergraduate course "Education and Artificial Intelligence"*, in: Commun. Comput. Info. Sci., Springer Verlag, Apr. 2018, pp. 211–222.

[10] D. Jiang, *Howslearning: A learning state classification approach in intelligent education*, in: Adv. Intell. Sys. Comput., Springer Verlag, Aug. 2018, pp. 369–375.

[11] L. Judijanto, M. R. Atsani and S. Chadijah, *Trends in the development of artificial intelligence-based technology in education*, International Journal of Teaching and Learning **2** (2024), 1722–1733.

[12] S. S. Kukeneh, A. Shahbahrami and M. Mahdavi, *Personalized virtual university: Applying personalization in virtual university*, in: Int. Conf. Artif. Intell., Manage. Sci. Electron. Commer., AIMSEC - Proc., IEEE Computer Society, Aug. 2011, pp. 6704–6706.

[13] S. Lai, B. Sun, F. Wu and R. Xiao, *Automatic Personality Identification Using Students' Online Learning Behavior*, IEEE Transactions on Learning Technologies **13** (2020), 26–37.

[14] G. Li, *Big data related technologies, challenges and future prospects*, Information Technology & Tourism **15** (2015), 283–285.

[15] N. Lin, G. Chen, K. Zheng and Y. Tang, *CluSoAF: A cluster-based semantic oriented analyzing framework for user behaviors in mobile learning environment*, in: Lect. Notes Comput. Sci., Springer Verlag, Jan. 2015, pp. 340–351.

[16] K. Liu, Y. Ni, Z. Li and B. Duan, *Data Mining and Feature Analysis of College Students' Campus Network Behavior*, in: IEEE Int. Conf. Big Data Anal., ICBDA, Institute of Electrical and Electronics Engineers Inc., United States, May. 2020, pp. 231–237.

[17] Q. Liu, B. Liu and Y. Lin, *The influence of prior knowledge and collaborative online learning environment on students' argumentation in descriptive and theoretical scientific concept*, International Journal of Science Education **41** (2018), 165–187.

[18] N. Luo, M. Zhang and D. Qi, *Effects of different interactions on students' sense of community in e-learning environment*, Computers and Education **115** (2017), 153–160.

[19] T. Ma, W. Wang and Y. Chen, *Attention is all you need: An interpretable transformer-based asset allocation approach*, International Review of Financial Analysis **90** (2023): 102876.

[20] X. Meng, J. Tao and H. Chang, *A conditional joint modeling approach for locally dependent item responses and response times*, Journal of Educational Measurement **52** (2015), 1–27.

[21] Z. Ming, H. Wang, M. Xu and D. Pan, *Evaluation of path stretch in scalable routing system*, International Journal of Machine Learning and Cybernetics, **6** (2015), 339–345.

[22] Y. Shen, J. Zhang and Y. Li, *Behavior Recognition of Teachers and Students in the Smart Classroom Based on Deep Learning*, in: Int. Conf. Inf. Sci. Educ., ICISE-IE, Institute of Electrical and Electronics Engineers Inc., Mar. 2024, pp. 345–349.

[23] K. Shringare, *Apache hadoop goes realtime at facebook*, National Journal of Management and Technology **3** (2015), 205–211.

[24] C. Shuying and C. Lei, *Reform and research of computer professional curriculum system based on systematic capability training under the background of emerging engineering education*, in: Int. Conf. Comput. Sci. Educ., ICCSE, Institute of Electrical and Electronics Engineers Inc., United States, Sep. 2018, pp. 1–5.

[25] J. Singh, G. Singh and A. Verma, *The anatomy of big data: concepts, principles and challenges*, in: Int. Conf. Adv. Comput. Commun. Syst., ICACCS, Institute of Electrical and Electronics Engineers Inc., Jun. 2022, pp. 986–990.

[26] V. L. Uskov, J. P. Bakken, K. Gayke, D. Jose, M. F. Uskova and S. S. Devaguptapu, *Smart university: A validation of "smartness features–Main Components" Matrix by real-world examples and best practices from universities worldwide*, in: Smart Innov. Syst. Technol., Springer Science and Business Media Deutschland GmbH, Jun. 2019, pp. 3–17.

[27] B. Wang, K. Deng, W. Wei, S. Zhang, W. Zhou and S. Yu, *Full cycle campus life of college students: A big data case in China*, in: Proc. - IEEE Int. Conf. Big Data Smart Comput., BigComp, Institute of Electrical and Electronics Engineers Inc., United States, May. 2018, pp. 507–512.

[28] D. Wang, H. Han, Z. Zhan, J. Xu, Q. Liu and G. Ren, *A problem solving oriented intelligent tutoring system to improve students' acquisition of basic computer skills*, Computers and Education **81** (2015), 102–112.

[29] Z. Wang, M. Wang, C. Zeng and L. Li, *Multi-Scale Deformable Transformers for Student Learning Behavior Detection in Smart Classroom*, ArXiv, **abs/2410.07834** (2024).

[30] E. P. Xing, Q. Ho, W. Dai, J. K. Kim, J. Wei, S. Lee, X. Zheng, P. Xie, A. Kumar and Y. Yu, *Petuum: A new platform for distributed machine learning on big data*, in: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., Association for Computing Machinery, 2 Penn Plaza, Suite 701, New York, NY 10121-0701, United States, Aug. 2015, pp. 1335–1344.

[31] P. Xu, X. Zhu and D. A. Clifton, *Multimodal learning with transformers: A survey*, IEEE Transactions on Pattern Analysis and Machine Intelligence **45** (2023), 12113–12132.

[32] J. Yi, H. Ni, Z. Wen, B. Liu and J. Tao, *CTC regularized model adaptation for improving LSTM RNN based multi-accent Mandarin speech recognition*, in: Proc. Int. Symp. Chin. Spok. Lang. Process., ISCSLP, Institute of Electrical and Electronics Engineers Inc., United States, Oct. 2016, pp. 985–997.

[33] J. Yu, Z. Yao, X. Li, H. Zheng, M. Lu, S. Tu, M. Li, J. Li, Q. Zhong, Z. Liao, L. Hou and J. Tang, *MoocRadar: A Fine-grained and Multi-aspect Knowledge Repository for Improving Cognitive Student Modeling in MOOCs*, in: SIGIR - Proc. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., Association for Computing Machinery, Inc, Jul. 2023, pp. 2924–2934.

[34] X. Zhai, S. Wei, Y. Guo and M. Zhang, *Smart classroom: an evaluation of its implementations and impacts – based on the longitude data of physics learning in a high school*, China Educational Technology **356** (2016), 121–127.

[35] D. Zhao, X. Ma and S. Qiao, *What aspects should be evaluated when evaluating graduate curriculum: Analysis based on student interview*, Studies in Educational Evaluation **54** (2017), 50–57.

Manuscript received July 1, 2024
revised December 18, 2025

G. Yang
School of Marxism, Xi'an Jiaotong University, Xi'an, China;
Department of Physical Education, Xidian University, Xi'an, China
E-mail address: gbyang@xidian.edu.cn

W. Yang
Department of Physical Education, Xidian University, Xi'an, China
E-mail address: 23221215234@stu.xidian.edu.cn

Y. Wang
School of Marxism, Xi'an Jiaotong University, Xi'an, China
E-mail address: yxwang2015@xjtu.edu.cn