Research Article

Personalized Online Education Learning Strategies Based on Transfer Learning Emotion Classification Model

Rongrong Wang1,2 and Zhengjie Shi1

1Huaiyin Institute of Technology, Huai'an 22301, Jiangsu, China
2Nanjing University of Aeronautics and Astronautics, Nanjing 210016, Jiangsu, China

Correspondence should be addressed to Rongrong Wang; jhwrr@hyit.edu.cn

Received 12 August 2021; Accepted 2 September 2021; Published 22 September 2021

Copyright © 2021 Rongrong Wang and Zhengjie Shi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the epidemic, online course learning has become a major learning method for students worldwide. Analyzing its massive data from the massive online education platforms becomes a challenge because most learners watch online instructional videos. Thus, analyzing learners’ learning behaviors is beneficial to implement personalized online learning strategies with sentiment classification models. To this end, we propose a context-aware network model based on transfer learning that aims to predict learner performance by solving learners’ problems and improving the educational process, contributing to a comprehensive analysis of such student behavior and exploring various learning models in MOOC video interactions. In addition, we visualize and analyze MOOC video interactions, enabling course instructors and education professionals to analyze clickstream data generated by learners interacting with course videos. The experimental results show that, in the process of “massive data mining,” personalized learning strategies of this model can efficiently enhance students’ interest in learning and enable different types of students to develop personalized online education learning strategies.

1. Introduction

In the last decades, technological advances have played an important and prominent role in the development of educational processes. Massive open online course (MOOC) is an outstanding innovation in the field of education, where more and more people are involved in online learning [1]. In particular, the impact of the epidemic has made learning through online course platforms (e.g., Coursera, edX, and Udacity [2]) the main way of learning for students worldwide, which offer mainly video-based course, quizzes, and forums [3]. The personalized design of the online video course plays an important role in the interest of the learners and is the main fulcrum to attract students to continue learning the course. Online learning platforms can store learner data in weblogs, which include their personal information and interactions with course content (e.g., videos, clickstreams events, forum discussions, and assessments). Video clickstreams (e.g., play, pause, search, and stop) are captured as they occur and then stored. Most learners spend most of their learning time watching video course, and as a result, many problems with learner-video interactions have gradually emerged [4].

Learner sentiment analysis can be performed by collecting, analyzing, and representing data related to learners’ interactions with the course that provides researchers and teachers with an opportunity to understand learners’ behavior and assess their performance through their interactions with the video content [5]. Several studies on learner engagement and explorations of patterns of learner behavior with video interactions have focused on analyzing data collected from learners’ interactions with different forms of course [6–8]. Related to this, many studies have provided many mechanisms to improve the quality of learning in MOOCs by exploring engagement behaviors and course content to predict learner performance and to reduce dropout rates [9, 10].
Educational environments face increasing complexity and diversity [11]; for example, students from different locations can take the same course. With regard to diversity, instructional designers are constantly challenged to adapt to the individual needs of students. Therefore, they must adopt teaching methods that are appropriate for different students [12]. This is the reason for the popularity of personalized learning. Personalized learning defined in the National Educational Technology Initiative supported by the U.S. Department of Education means that learning platforms can optimize learning paths and instructional methods based on the needs of each learner [13]. Such a learning platform allows students to pursue their personal learning goals at their own pace [14]. Thus, the main benefit of personalized learning is its ability to adapt to the needs of different students. This benefit is supported by empirical evidence, which suggests that personalized learning allows instructional designers to meet students’ needs and helps students clarify and improve their understanding of learning objectives [15, 16].

Based on a comprehensive observation of the personalization program, it should include personalized learning materials, personalized graphical user interface, personalized learning activities, personalized learning strategy assistance, and personalized collaboration [17]. They use intelligent agents to personalize the learning environment from three different perspectives: the system, the learner, and the teacher [18]. Many researchers have worked to create a personalized e-learning system to help learners improve their academic performance. However, the main personalized e-learning systems do not take into account the ability of the learners and the difficulty ranking of the recommended subjects. In addition, the continuity of the learning trajectory of the personalized course needs to be considered to make the learning process smoother for the learners. This is because inappropriate courseware may cause overload or learning without understanding, which may affect student performance [19].

Likewise, online learning requires attention to learning styles and individual differences. To this end, this paper provides an updated review of current research discussing a variety of individual differences, including learning styles, emotional states, gender differences, and prior knowledge. The study focused on video clickstream data in MOOC course. Analytic studies have focused on examining learner performance through video engagement behaviors using explicit features such as views and annotations [20, 21]. Therefore, analyzing learners’ learning behaviors is beneficial for implementing personalized online learning strategies with affective classification models. To this end, this paper aims to propose a context-aware network model based on transfer learning, which aims to predict learners’ performance by solving their problems and improving the educational process, contributing to a comprehensive analysis of such student behavior and exploring various learning models in MOOC video interactions [22]. We also visualize MOOC video interactions, enabling course instructors and education professionals to analyze clickstream data generated by learners interacting with course videos as implicit features rather than explicit features. In this way, it is possible to analyze how learners’ behavior in videos affects their performance in MOOC course, enabling the development of personalized online education learning strategies for different types of students.

As we all know, the application of deep learning and artificial intelligence technology in the field of education has always been the crown of online education and what many online education enterprises are doing. However, up to now, although there are many in-depth learning and artificial intelligence solutions, there are not many projects that can be applied and implemented in practice. Most solutions are still in the concept stage or data collection stage, which is still a distance from practical application. This paper will take the personalized question bank system we actually use in MOOC teaching as an example to introduce in detail the technical principle of the education system based on deep learning technology and the information architecture design technology in practical application. With the popularity of machine learning and deep learning, recommendation algorithms based on machine learning and deep learning have gradually become the mainstream application technology in recent years.

The contributions of this paper are as follows:

1. We designed a novel neural network model for online learning, and its learning method is transfer learning, which can effectively process the information of MOOC online course video.
2. In order to effectively extract important features, we design a feature extraction scheme based on free selection.
3. We have done a lot of experiments to show the superiority of the proposed scheme.

The rest of the paper is organized as follows. Section 2 describes related work that provides an overview of different personalization strategies for online learning. Section 3 presents the context-aware network model for transfer learning proposed in this paper. Section 4 shows the experimental setup and the introduction of MOOC data mining. Section 5 visualizes the MOOC data analysis and the performance evaluation of the model in this paper. Section 6 concludes the paper and discusses the limitations of this paper.

1.1. Related Work. E-learning helps the traditional learning process take a step forward by providing students with materials that can help them learn anytime and anywhere [23, 24]. However, many studies have shown that web-based online learning still lacks intelligence that may not be appropriate for each learner’s characteristics [25]. Wong et al. [7] proposed an autonomous approach to self-organization by creating learning objects (LO) that can provide learners with a good LO structure. Three common types of learning resource filtering are described in [10]: content-based (CB), collaborative filtering (CF), and hybrid filtering (HF). The use of CF in [26] will be analyzed based on the similarity between learners’ scores and then predict which substance is
more suitable for them. In contrast, the lack of rating defects in CF methods is discussed in [11], which occurs when users do not have sufficient rating documentation. The CF approach will face difficulties with high data sparsity. CB works by providing recommendations for each learning subject that fit the student’s learning goals and preferences. Therefore, CB will consider some learners’ factors such as their skills or talents, goals, attitudes, and psychological styles in the CB recommendation system [7, 16].

Albatayneh et al. [27] proposed that content-based e-learning systems provide suggestions to learners by matching their personalities with the learning outcomes/objectives of a particular course. Chen and Wang [28] implemented personalized learning on a handheld device to accommodate students with different cognitive styles. Tlili et al. [29] developed a personalized educational game based on learners’ personalities.

Based on the previous work, it can be concluded that personalized learning needs to be enhanced in many aspects: facilities, content-based recommendations, content filtering, and collaborative filtering to meet learners’ preferences. The purpose of this paper is to propose a personalized learning model based on transfer learning algorithms to screen learning methods that help reduce students’ failure factors.

2. Context-Aware Network Model Based on Transfer Learning

2.1. MOOC Data Acquisition Objectives. This section presents the idea of implementing deep learning in an existing IoT system architecture. The architecture is divided into two parts, one of which is the camera part that collects the video movements of students for action recognition [4, 14]. The collection of videos was broken down into 11 small video clips with four different actions for the students: entering the classroom, standing, sitting, and walking out of the classroom. The video clips are then classified into images with specific frames and the dataset for video-to-image classification is discussed. After the IoT system identifies student activity, it combines the results with the sensor dataset and determines whether the MOOC video content is the focus based on the context of the data as well as decides whether to inform students that they should focus on students. Every 10 minutes, the process takes data reading from context-aware sensors. Figure 1 depicts the data collection and the identification process of the sensor to control the student in-and-out of classroom. The video clips are then classified into images with specific frames and the dataset for video-to-image classification is discussed. After the IoT system identifies student activity, it combines the results with the sensor dataset and determines whether the MOOC video content is the focus based on the context of the data as well as decides whether to inform students that they should focus on students. Every 10 minutes, the process takes data reading from context-aware sensors. Figure 1 depicts the data collection and the identification process of the sensor to control the student in-and-out of classroom.

To collect data, a context-aware sensor classroom is created. The sensor data are collected through a Raspberry Pi board and the students’ motions are recorded by a video camera. Collect data from temperature, humidity, and luminance sensors using low-power context-aware sensors. Use MySQL [7] to manage the database. Video-to-image data require large datasets to obtain proper efficiency, so working with the C3D dataset makes the dataset large enough to be used for training and testing of action recognition experiments. Using passive infrared sensors (PIR) covering 360°, the largest area of motion can be detected. The same can also be used separately from different sensors to collect different data when students enter the classroom. Placement of sensors for data collection for sensors cover all possible space in the room, so errors are reduced. PIR sensors are placed to collect all the actions that are collected in the MySQL database provided by the server.

2.2. Transfer Learning Based on MOOC Video Data. Transfer learning is a common approach that trains small domain datasets into large domain datasets. In practical applications, most of the datasets are usually the largest in the domain where feature extraction can be effectively utilized. The C3D and transfer learning models proposed in this paper as shown in Figure 2 show how MOOC data can be combined with the classroom in human action data domain and transfer learning. Since the MOOC data are taught by a single person, while the classroom action dataset is composed of multiperson actions, the dataset faces a great challenge in transfer learning [31].

In Figure 2, we assume that the number of feature maps extracted by C3D is N and the size is M × M. The output c_n of the n-th (n = 1, 2, . . . , N) feature map after learning feature extraction by thousand is as follows:

$$c_n = \frac{1}{M \times M} \times \sum_{i=1}^{M} \sum_{j=1}^{M} w_{ij}^{(n)}$$

where w_{ij}^{(n)} represents the pixel value in the nth feature map.

The output y_l of the l-th (l = 1, 2, . . . , 1024) neuron in the first hidden layer of C3D is

![Figure 1: Flowchart of student identification in IoT system.](image-url)
where \( f \) represents the ReLU activation function, \( \theta_{l,n} \) represents the connection weights between the first implied layer and the flattened layer, and \( b_l \) represents the corresponding bias term.

Similarly, the output \( z_d \) of the \( d \)-th (\( d = 1, 2, \ldots, 1024 \)) neuron in the second hidden layer of C3D is

\[
y_d = f\left( \sum_{l=1}^{1024} \theta_{dl} \times y_l + b_d \right),
\]

where \( \theta_{dl} \) represents the connection weights between the second hidden layer and the first hidden layer neurons and \( b_d \) represents the corresponding bias term.

The input to the fully connected layer is \( g_1 \) and \( g_2 \), and the input to the Softmax classifier outputs the prediction in the form of a probability:

\[
\begin{align*}
    y'_1 &= \frac{e^{g_1}}{\sum_{k=1}^{400} e^{g_k}}, \\
    y'_2 &= \frac{e^{g_2}}{\sum_{k=1}^{400} e^{g_k}}.
\end{align*}
\]

When \( y'_1 \) is greater than 0.5 and \( y'_2 \) is less than 0.5, it means that the model identifies the result that the student is focused on learning and vice versa.

Transfer learning helps to build networks of knowledge-sharing concepts, which actually help to train datasets with learned concepts [32]. In this paper, we implement transfer learning to train our experimental action dataset using a large regional dataset of the MOOC video dataset and apply this transfer learning to C3D. The model in this paper successfully captures the feature vector for the first task and then redefines the convolution function with an additional fully connected layer and retraining the feature vector. Using 128 layers of convolution to improve programming efficiency, the image dataset added for transfer learning does not require large filtering layers. Finally using fully connected layers, classification can be performed. This makes it easy to transfer the knowledge to the network to perform another task. Finally, strategies for personalized learning inform the students.

2.3. Context-Aware Network. To further incorporate all the scenarios in the MOOC, we categorized these into effective categories to further enhance students’ strategy development for online personalized learning. The complete architecture is shown in Figure 3, which shows how the proposed context-aware network with a transfer learning architecture works. Two different architectures can be seen working in parallel to collect and predict sensor and action recognition outputs. Transfer learning is performed on the human action image dataset. The first part is action recognition, showing how the data are collected and used for feature extraction. C3D is used as a feature extraction tool to identify four different behaviors of students in the classroom, that is, entering, sitting, standing, and exiting the classroom. CNN is arguably the most widely used method in human behavior recognition. It consists of multiple hidden and pooling layers and fully connected layers.

The second part is to estimate the sensor reading for the future classroom using LSTM. A series of sensor readings, such as \((T_i, H_i, L_i)\), \((T_{i-1}, H_{i-1}, L_{i-1})\), and \((T_{i-n}, H_{i-n}, L_{i-n})\), were used to predict the sensor readings at time \((T_{i+m}, H_{i+m}, L_{i+m})\), where \( T_i \), \( H_i \), and \( L_i \) indicate the
students’ mental state, learning ability, and attention to online videos at time $i$, respectively. Identified student activity and predicted sensor readings $(T_{i,m}, H_{i,m}, L_{i,m})$ are fed into the ground truth data for individualized formulation of specific calculations. The output of action recognition is $(1, 0, 0, 0), (0, 1, 0), (0, 0, 1, 0), (0, 0, 1, 0)$, and $(0, 0, 0, 1)$, and each value indicates that the student is “recording,” “thinking,” “watching video,” and “daze.” Through the normalization layer, the identified action values and the “thinking,” “watching video,” and “daze.” Through the normalization layer, the identified action values and the predicted sensor values become 0 and 1, respectively, and the final output is the student’s learning effect at each moment.

The model in Figure 3 has its objective function set to $L_m$, which is a weighted contrast loss function. $L_m$ can be interpreted as a set of soft constraints that impose a significantly higher penalty for misclassifying a sample to any class belonging to another cluster compared to the penalty for misclassifying to a class belonging to the same cluster. In other words, minimizing the weighted contrast loss results in a smaller similarity measure for samples belonging to the same cluster and a larger one for samples spanning different clusters. The weighted contrast loss function is given by

$$L_m = \omega_{ij} \times \left( \frac{1 - Y}{2} \times D^2 + \frac{Y}{2} \times \max(0, m - D)^2 \right),$$

where $\omega_{ij} = \begin{cases} 0.1 & \text{if } i \in C_k \text{ and } j \in C_k \\ 1 & \text{otherwise} \end{cases}$ is the weight corresponding to the category labels $i$ and $j$ and $D$ is the L2 norm between a pair of samples. $Y$ is the label representing the degree of similarity between two samples, and $m$ is the margin. $C_k$ denotes the $k$-th cluster.

In order to avoid merging too many irrelevant details and noise in the sensor data fusion process and to minimize the effect of artifacts, we use the following steps to perform the fusion of the detail layer.

**Step 1.** The weight matrix $W^j$ is computed by the absolute value of the large rule.

$$W^j = \begin{cases} 1, & |d^j_i| > |d^j_9|; \\ 0, & j = 1, 2, \ldots, N. \end{cases}$$

**Step 2.** The weight matrix $W^j$, using Gaussian filtering, is processed in equation (6).

$$W^j_d = \text{Gaussian}(W^j, \sigma_s),$$

where $\sigma_s = 2$.

**Step 3.** The initial fusion of detail layers $d^j_i$ and $d^j_2$, by the weighted average rule, is obtained by $M^j$; that is, we have

$$M^j = W^j_d d^j_i + (1 - W^j_d) d^j_2.$$  

**Step 4.** The optimization strategy WLS is used to optimize $M^j$ to obtain the fused detail layer $D^j$. The procedure is as follows:

$$\sum_p \left( (D^j_p - M^j_p)^2 + a^j_p (D^j - (d^j_p))^2 \right).$$

Let the weight $a^j_p = (\sum_{q \neq p} |d^j_q| + \epsilon)^{-1}$ and $p$ denote the space position. The parameter $\epsilon$ is usually set to 0.0001 to prevent $\sum_{q \neq p} |d^j_q|$ from dropping to 0. It is a window centered on position $p$. A window that is too large causes poor fusion and consumes too much time, while a window that is too small does not eliminate the effect of noise. Minimization term $D^j_p - M^j_p$ aims to minimize the geometric distance between the detail layer $M^j$ and the fused detail layer $D^j$; minimization term $a^j_p (D^j_p - (d^j_p))^2$ aims to make the fused detail layer $D^j$ closer to the model detail.

![Figure 3: Transfer learning context-aware network combining video and sensors.](image)
layer $d^i_j$, so that the output data are more characteristic. $\lambda$ is a global parameter to control the weights of these two components. Equation (9) is converted into matrix form

$$\begin{align*}
(D^i - M^i)^T (D^i - M^i) + \lambda (D^i - d^i_j)^T A^i (D^i - d^i_j),
\end{align*}$$

where $D^i$, $M^i$, and $d^i_j$ are represented as vectors, and $A^i$ is a diagonal matrix with $a^i$.

Minimizing equation (10) yields the linear system of equations:

$$\begin{align*}
\left[ 2I + \lambda \left( A^i + (A^i)^T \right) \right] D^i &= 2M^i + \lambda \left( A^i + (A^i)^T \right) d^i_2.
\end{align*}$$

Since $A^i = (A^i)^T$, equation (11) can be simplified as

$$\begin{align*}
(I + \lambda A^i) D^i &= M^i + \lambda A^i d^i_2.
\end{align*}$$

The fused detail layer $D^i$ is obtained by using equation (12).

By analyzing the areas of sensor data containing noise or not related to the visual detail information or related to the visual detail information, it can be seen that the optimized strategy WLS can obtain a better-fused detail layer.

2.4. Experimental Setup. To demonstrate the effectiveness and practicality of the proposed model, a case study of two MOOC course was conducted. Firstly, data preprocessing was performed, including feature extraction from video clickstream data. As shown in Tables 1 and 2, we observe that the implementation of representing clickstream data in the (MMDS SELF-PACED) course takes longer than representing clickstream data in the (Automata SELF-PACED) course [18], which means that the time depends on the size of the clickstream data in order to consider more efficient and clearer execution of the visualization algorithm considering the running time [5, 13]. Data are generated separately each week for each participant in each course. This makes the model more flexible and efficient and requires little time to visualize the entire data. In the prediction phase, we considered unbalanced datasets, converted the datasets obtained during feature extraction into appropriate model input data, constructed shape-consistent padding vectors, and then marked them before feeding them into the model layer. The dataset was divided into 70/30 training/testing to determine the prediction of the learner’s performance [12].

The model uses Keras and TensorFlow as the modeling framework and Adam optimizer for training.

2.5. Visualization Results and Model Performance. In this section, we investigate the effectiveness of the proposed model for assessing learner behavior through video clickstreams and explain the possible relationship between learner behavior and their performance on the study dataset. The behavior of learners watching videos to complete the first task in a given week is classified as a community gathering within the network based on structural clustering generated based on structural identity, which is closely related to kindness. Each cluster was given a different color. As shown in Figure 4, the size of the video nodes is proportional to the number of associated learner nodes, which indicates the status of the learner’s utilization of the video, and this stage allows the teacher to monitor the learner’s behavior (e.g., the learner’s behavior and the most viewed videos) [15]. In addition, teachers can determine which learners are more likely to drop out, such as the red and orange notes.

In general, we focus on how the learner interacts with the viewed video. If learners take a long time to interact with the video (reflecting a high level of interest), which implies that they make an effort while watching the current video (e.g., most pause/backward search events), this can be explored in real-time video utilization in order to more precisely analyze specific video utilization. Thus, instructors can directly select videos of interest, such as videos of most events. For example, we selected video (2) from the "Theory of Automata" course and video (14) from the "MMD Self-Paced" course, which are the most popular videos. As shown in Figure 5, for play and pause events, the $x$-axis indicates the number of learners watching the video, and the $y$-axis indicates the time of the event and the actual time of the video. Depending on the video real time and viewing time, we can interpret the seek mode, especially the location with dense seek lines. From Figure 5, we observe positive search events above the $x$-axis when learners interact with the video, which can be used as an indicator to assess whether learners are bored or the video content is not interesting, especially when searching for positive events multiple times. In contrast, videos may be more interesting or have more relevance in quizzes when "backward search" events occur, especially in the most viewed positions. By exploring data visualization to explain the second task, we can expect that learners will not interact with future videos when they do not interact with the course videos and when they find the course difficult to follow, uninteresting, or boring. Conversely, when learners become interested in incomplete video content, there is a lot of rewatching, indicating that they will continue until the end of the course.

Four different types of actions were added to the IoT system to classify students’ actions during daily classroom activities. With 44 video action clips from four different actions in the MOOC, the classroom action dataset provides the diversity of multiplayer actions. The dataset in Figure 6 is used for action recognition in real action videos collected in the smart classroom, providing a variety of perspectives on the diversity of video data, which makes it a good dataset in terms of context awareness and action recognition. In Figure 6, a sample video action dataset for classroom action recognition can be seen.

Online learning of students in the classroom is known from Figure 6, and the results shown in Figure 7 are reported here in order to compare the efficiency of the personalized recommendation scheme in this paper. Accuracy and subject operating characteristics were considered to evaluate model performance on the test dataset over several weeks. We observed that, in the "Massive Dataset Mining" course, the online video viewing of this paper was relatively stable, as
Table 1: Basic information of the two datasets.

| Course name | Self-matching speed | MMDS          | Automatic control |
|-------------|---------------------|---------------|-------------------|
| Number of weeks | Video | Student | Handling events | Video | Student | Handling events |
| 1           | 11     | 4,578   | 258,899         | 5      | 2,784   | 155,439        |
| 2           | 12     | 1,446   | 170,913         | 5      | 716     | 54,608          |
| 3           | 17     | 747     | 695,839         | 3      | 498     | 29,707          |
| 4           | 17     | 881     | 87,532          | 5      | 368     | 21,589          |
| 5           | 9      | 579     | 34,536          | 4      | 362     | 26,780          |
| 6           | 15     | 565     | 52,192          | 4      | 199     | 9,625           |
| 7           | 15     | 313     | 17,433          | -      | -       | -               |

Table 2: Video click event information.

| Course name | Self-matching speed | MMDS          | Automatic control |
|-------------|---------------------|---------------|-------------------|
| Event name  | Counting Average    | Counting Average |
| Play        | 234,346 0.178       | 108,975 0.366 |
| Suspension  | 194,159 0.147       | 92,835 0.312  |
| Find        | 139,807 0.106       | 61,302 0.206  |
| Exchange rate changes | 6,628 0.005 | 3,997 0.013 |
| Load        | 69,626 0.053        | 23,900 0.08   |
| Stop        | 672,778 0.511       | 6,739 0.023   |

Figure 4: Default views of the learner interacting with the video.

Figure 5: Video utilization of events (play, pause, and search).
shown in Figure 7(a), and the students did not pause until the end of the learning process. And as shown in Figure 7(b), without personalized recommendations, this student may pause the learning video at any time, which is not interesting for them all to learn the content. One possible reason is that nearly three-quarters of the learners did not have video interaction at the end of the course. On the other hand, as we expected, the model in this paper plays an important role in dealing with longer memory in massive amounts of data.

As clearly seen from the loss drop of the plot in Figure 8, during the model training process, our method converges faster than the baseline scheme, the training process is very stable, and the model training is completed starting at 27k steps, while the latest DenseNet requires 32k steps. The accuracy of our scheme increases almost linearly during the training process, as shown in the right panel in Figure 8, and the training accuracy is always the highest. In general, it can be seen from Figure 8 that the method in this paper can converge quickly and stably, which can be attributed to the C3D-based migratory learning feature extraction and automatic multisource sensing data fusion. The unstable training accuracy and poor convergence of ResNet indicate that the design idea of our framework has the effect of mitigating overfitting, while our framework has better generalization ability compared to ResNet.

The collected course data were visualized by classifying it on the model of this paper, as shown in Figure 9. Each point corresponds to a node on the graph, and its color corresponds to its node class. Clustering of certain course classes was observed, while others were separated. For example, magenta and green belong to the same cluster, so they are close to each other, but far from the other classes. This is based on the fusion of dynamic features from different sensor data. The points of different colors are also well classified, indicating that our model can learn better the correlation of sensor data from different sources.

We choose different courses of MOOC to demonstrate the iterative process of our model. As we can see in Figure 10, at the beginning of the algorithm, since we are following the
Figure 8: Loss and accuracy during model training.

Figure 9: Visualization of course classification.
data fusion algorithm to assign weights to each course, most of the irrelevant courses will have larger weights, and after 100 iterations, the scores tend to stabilize and the scores of relevant courses will gradually grow. The task improves to 0.3, and the performance of the model task in this paper improves from 0.066 to 0.384, which is a relatively significant improvement. In fact, if we only consider the similarity between courses, most students study courses without relevant information such as gender and age, which appear in MOOC, so the calculation of course similarity and the classification of age appear to be very important, which can make the similarity calculation between courses more accurate and thus improve the ranking effect of course recommendation on students.

3. Conclusions

This paper proposes a context-aware network model based on transfer learning, which aims to predict learners’ performance by solving their problems and improving the educational process, contributing to a comprehensive analysis of such student behavior and exploring various learning models in MOOC video interactions. The experimental results show that, in the process of “massive data mining,” the accuracy of this model is 90.30% better than the baseline, and it can realize different types of students to develop personalized online education learning strategies. The scheme in this paper achieves a certain effect of personalized learning strategy, but the model is too large, and the model structure can be optimized in the future. On the other hand, scenario analysis can be done directly from the MOOC video content without focusing too much on the students, who change too much, resulting in inaccurate predictions.

Data Availability

The datasets used in this paper are available from the corresponding author on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] K. S. Hone and G. R. El Said, “Exploring the factors affecting MOOC retention: a survey study,” Computers & Education, vol. 98, pp. 157–168, 2016.
[2] C. M. Stracke and A. Bozkurt, “Evolution of MOOC designs, providers and learners and the related MOOC research and publications from 2008 to 2018,” in Proceedings of the International Open & Distance Learning Conference (IODL19), Eskisehir, Turkey, January 2019.
[3] P. V. Murugan and V. M. Queen, “MOOCs as a digital learning platform,” International Journal of Multidisciplinary Research Review, vol. 3, no. 1, pp. 28–37, 2020.
[4] T. Zhou, S. Huang, J. Cheng, and Y. Xiao, “The distance teaching practice of combined mode of massive open online
course micro-video for interns in emergency department during the COVID-19 epidemic period," *Telemedicine and e-Health*, vol. 26, no. 5, pp. 584–588, 2020.

[5] A. Moattarian, M. H. Tahririan, and A. Alibabaei, “Task complexity manipulation and EFL learners’ interactions in the process of collaborative pre-planning.” *Applied Research on English Language*, vol. 8, no. 1, pp. 51–78, 2019.

[6] L. Y. Li, “Effect of prior knowledge on attitudes, behavior, and learning performance in video lecture viewing,” *International Journal of Human-Computer Interaction*, vol. 35, no. 4-5, pp. 415–426, 2019.

[7] J. Wong, M. Khalil, M. Baars, B. B. de Koning, and F. Paas, “Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course,” *Computers & Education*, vol. 140, Article ID 103595, 2019.

[8] D. J. Lemay and T. Doleck, “Grade prediction of weekly assignments in MOOCS: mining video-viewing behavior,” *Education and Information Technologies*, vol. 25, no. 2, pp. 1333–1342, 2020.

[9] K. Limsathitwong, K. Tiwatthanont, and T. Yatsungnoen, “Application of artificial intelligence in dropout prediction and analysis through educational data mining,” in *Proceedings of the 2018 5th International Conference on Business and Industrial Research (ICBIR)*, pp. 110–114, IEEE, Bangkok, Thailand, May 2018.

[10] V. Hegde and P. P. Prageeth, “Higher education student dropout prediction and analysis through educational data mining,” in *Proceedings of the 2018 2nd International Conference on Inventive Systems and Control (ICISC)*, pp. 694–699, IEEE, Coimbatore, India, June 2018.

[11] G. Nagarajan and R. I. Minu, “Multimodal fuzzy ontology creation and knowledge information retrieval,” in *Proceedings of the International Conference on Soft Computing Systems*, pp. 697–706, Springer, Dubai, UAE, November 2016.

[12] T. Soffer and A. Cohen, “Students’ engagement characteristics predict success and completion of online courses,” *Journal of Computer Assisted Learning*, vol. 35, no. 3, pp. 378–389, 2019.

[13] D. Ifenthaler and J. Y.-K. Yau, “Utilising learning analytics to support study success in higher education: a systematic review,” *Educational Technology Research & Development*, vol. 68, no. 4, pp. 1961–1990, 2020.

[14] M. Liz-Dominguez, M. Caicedo-Rodríguez, M. Llamas-Nistal, and F. A. Mikic-Fonte, “Systematic literature review of predictive analysis tools in higher education,” *Applied Sciences*, vol. 9, no. 24, 5569 pages, 2019.

[15] B. Li and X. Xu, “Application of artificial intelligence in basketball sport,” *Journal of Education, Health and Sport*, vol. 11, no. 7, pp. 54–67, 2021.

[16] R. A. Minhas, A. Javed, A. Irzata, M. T. Mahmood, and Y. B. Joo, “Shot classification of field sports videos using AlexNet convolutional neural network,” *Applied Sciences*, vol. 9, no. 3, 483 pages, 2019.

[17] K. S. McCarthy, M. Watanabe, J. Dai, and D. S. McNamara, “Personalized learning in istart: past modifications and future design,” *Journal of Research on Technology in Education*, vol. 52, no. 3, pp. 301–321, 2020.

[18] J. D. Basham, T. E. Hall, R. A. Carter, and W. M. Stahl, “An operationalized understanding of personalized learning,” *Journal of Special Education Technology*, vol. 31, no. 3, pp. 126–136, 2016.

[19] P. A. Bishop, J. M. Downes, S. Netcho et al., “Teacher roles in personalized learning environments,” *The Elementary School Journal*, vol. 121, no. 2, pp. 311–336, 2020.