Flipped Classroom Design of College Ideological and Political Courses Based on Long Short-Term Memory Networks

Fei Su¹ and Zhe Fan²

¹College of Sport and Health Science, Xi’an Physical Education University, Xi’an 710000, Shaanxi, China
²Shijiazhuang Vocational College of Scientific and Technical Engineering, Zhengding, Shijiazhuang 050800, Hebei, China

Correspondence should be addressed to Zhe Fan; fanzhezhengding@163.com

Received 9 April 2021; Accepted 17 June 2021; Published 13 July 2021

1. Introduction

Flipped classroom [1–3] benefits from the expansion of information technology. It reverses the transfer of knowledge in the classroom and the deepening of knowledge under the classroom in time and space and realizes the transfer of knowledge and knowledge from the two levels of teaching structure and teaching process. Before class, students complete the cognition of basic knowledge through online self-learning [4, 5]. Teachers in the classroom organize classroom activities to further consolidate and sublimate knowledge and achieve a teaching form of deep learning. Some universities’ realistic teaching and related study findings provide some reference value for the use of flipped classrooms in ideological and political theory courses. The interpretation and perception of basic theoretical knowledge is not the only way to enhance the teaching impact of ideological and political theory courses [6, 7], but more importantly, the contradiction transformation of internalization and externalization in behavior. This is a realistic problem faced by teachers of ideological and political theory. It is also a frequently discussed and important topic in the field of ideological and political theory teaching and research [8, 9].

Take online teaching [10, 11] as a useful supplement to traditional teaching methods, continuously innovate online teaching methods and combine online teaching with traditional teaching methods. The characteristic of flipped classroom teaching is to transfer knowledge in the classroom in the traditional teaching model and to reverse the time and space of knowledge deepening in the two parts outside the class, to realize the flip of the structure and process of knowledge transfer and knowledge internalization. In recent years, flipped classroom teaching has gradually become a focus of education and teaching reforms at all stages. Research on flipped classrooms in ideological and political theory
courses in colleges and universities has increased, but there is not much research on teaching reform in this field [12].

This paper begins by examining the empirical connotation of flipped classrooms and then examines the relevance and viability of using flipped classrooms in college ideological and political theory classes [13], as well as the theoretical foundation and core concepts of their use, before using "basic" courses as an example to present ideological and political theory in a concrete manner. A long short-term memory neural network [14–17] prediction model optimized by hybrid genetic algorithm is built in the design and implementation of the flipped classroom teaching process of theoretical courses and is based on neural network technology [18–21]. It can predict and evaluate the teaching quality of flipped classrooms in college ideological and political courses. The key offerings of the proposed study are as follows:

1. To employ neural network technology to develop a long- and short-term memory neural network prediction model focused on hybrid genetic algorithm optimization that can predict and assess the teaching content of flipped college ideological and political courses.

2. The root mean square error between the true value and the expected value of the teaching output predicted by the long- and short-term memory neural network is used as the fitness function in this paper to automatically find the optimal dropout probability and the number of hidden layer units of the neural network.

3. To use a sequential quadratic programming algorithm to advance local search, rapidly and precisely optimize dropout probability and the number of hidden layer units, and input the obtained optimal parameters into a long- and short-term memory network to predict the teaching output of college ideological and political courses flipped classroom and evaluation.

2. Background

2.1. The Connotation of Flipped Classroom. The partnership between education and teaching and information technology is becoming closer as network information technology develops at a rapid pace. The flipped classroom was born in the development of multimedia technology. Before the term “flipped classroom” officially appeared, the similar term “Inverted Classroom” existed, but the two were not completely equivalent. The flipped classroom really appeared in 2007 when two chemistry teachers at Woodland Park High School in the United States reorganized the teaching classroom to help students who were absent. Teachers make teaching videos and teaching materials by themselves and arrange for students to watch them at home. The assignments will be completed under the guidance of teachers within the specified time in class. The period from 2008 to 2012 is the stage of practice and development and innovation of flipped classrooms. In 13 years, it has moved from primary and secondary schools to colleges and universities. Teachers in American colleges and universities began to use flipped courses for teaching [22] and achieved some good results.

With the application and development of flipped classrooms in university teaching, related research on flipped classrooms in ideological and political theory courses continues to emerge. The analysis of its meaning still needs to be traced from the existing research results of the academic community. In the existing research results, there are different understandings and interpretations of flipped classrooms:

1. A flipped classroom is a teaching model that consists of a structure with set procedures and predictable classroom teaching activities implemented using modern information technology and driven by the “student-centered” educational theory. Lai Huiming believes that “flipped classroom is an innovative teaching model held by recent information technology that is, the conventional teaching method is inverted in the classroom. Independent learning, a teaching style of collaborative learning between teachers and students in the classroom, is carried out using the provided teaching tools.

2. Flipped classroom is a teaching method. It emphasizes deep learning through classroom communication and learning, not just on the cognitive level.

3. Flipped classroom is a teaching structure. Scholars regard the operation process as the core foothold of a flipped classroom and emphasize the external form as the biggest feature to highlight the characteristics of the flipped classroom.

4. A flipped classroom is a type of instruction. It focuses on a state of interactive contact created during the teaching process by teachers and students. The flipped classroom is a method of instruction in which students use interactive learning materials developed and generated by teachers for self-directed learning prior to class, then actively participate in interactive activities between classmates and teachers, and complete classroom exercises in the classroom.

2.2. Features of Flipped Classroom. The flipped classroom is characterized by the reversal of teacher and student positions, as well as the transformation of teaching concepts, the alteration of classroom layout, and the creation of new classroom activities:

1. A shift in teaching philosophy from “teacher-centered” to “student-centered” is needed before flipped classroom teaching can be implemented.

2. The guarantee for the implementation of flipped classroom teaching is the transformation of the teaching system from “teaching first and then learning” to “learning first and then teaching, teaching through learning”
2.3. The Dilemma of Flipped Classroom in Ideological and Political Theory Course. Educators, instructional objects, and teaching facilities are the three main components of the teaching process, and they all play a role in ensuring that teaching methods are used effectively. The use of any teaching method must consider the coordination between them. Whether the three are coordinated directly affects the teaching effect of the teaching method. The ideological and political theory class flipped classroom has many influences during the implementation process. The use of classrooms in ideological and political theory courses has not been widely recognized and promoted. The most important influencing factor is the insufficient ability to use information technology.

3. Methodology

This section explains the hybrid genetic algorithm-optimized LSTM [23] prediction model, as well as the concepts of the LSTM and the flipped classroom assessment model of college ideological and political courses.

3.1. LSTM

3.1.1. Basic Principles of LSTM. The hidden layer of Simple Recurrent Neural Network has only one state $h_t$ which is suitable for short-term input, but it does not work well for long-term input. To improve this situation, LSTM RNN is used to save a longer-term state by adding a state $c_t$. The newly added state $c_t$ is called the unit state. At time $t$, LSTM RNN has three inputs: the input value $x_t$ of the network at the current time, the output value $h_{t-1}$ of the LSTM RNN at the previous time, and the unit state $c_{t-1}$ at the previous time. The LSTM RNN has two outputs: the output value of the LSTM RNN at the current time $h_t$ and the current unit state $c_t$. The time dimension diagram of LSTM RNN is shown in Figure 1.

The calculation equations of the input gate, forget gate, and output gate of LSTM are as follows:

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i),$$

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f),$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o),$$

where $x_t \in \mathbb{R}^d$ is the input vector of the LSTM unit; $i_t \in \mathbb{R}^h$, $f_t \in \mathbb{R}^h$, and $o_t \in \mathbb{R}^h$ are the activation vectors of the input gate, forget gate, and output gate, respectively; and $h_{t-1} \in \mathbb{R}^h$ represents the hidden state vector, which is also called the output vector of the LSTM unit. $W \in \mathbb{R}^{dh \times d}$, $U \in \mathbb{R}^{dh \times h}$, and $b \in \mathbb{R}^h$ are the weight matrix and bias vector that need to be learned during training.

Calculate the activation vector $c_t'$ used to describe the current unit input based on the previous output and this input. The calculation equation is as follows:

$$c_t' = \sigma_h(W_c x_t + U_c h_{t-1} + b_c),$$

where $\sigma_g$ represents the hyperbolic tangent tanh activation function. The current cell state $c_t$ can be obtained by inputting the activation vector $c_t'$ of the current cell and the cell state $c_{t-1}$ at the previous moment:

$$c_t = f_t \circ c_{t-1} + i_t \circ c_t'. \tag{3}$$

The operator $\circ$ represents the Hadamard product (elementwise product). Finally, the output of the LSTM unit can be obtained as

$$h_t = o_t \circ \sigma_h(c_t). \tag{4}$$

LSTM can be trained on a set of sequence data in a supervised manner, through time backpropagation to calculate the gradient required in the optimization process, to change each weight of the LSTM network to make the error of the corresponding weight (in the output layer of the LSTM network); the derivative is proportional.

The sigmoid activation function and its derivative need to be used in the backpropagation process, and the hyperbolic tangent activation function and its derivative are used at the same time, which are defined as follows:

$$\sigma_g'(z) = y(1 - y),$$

$$\sigma_h(z) = y = \frac{e^z - e^{-z}}{e^z + e^{-z}}, $$

$$\sigma_h'(z) = 1 - y^2. \tag{5}$$

3.1.2. Dropout. Generally, if the algorithm wants to have a good performance on the training set and the test set, overfitting should be avoided. For this reason, a regularization method needs to be added to the algorithm. Dropout is a regularization method with a better effect. Dropout refers to a regularization method that randomly eliminates or retains a unit in the neural network during neural network training.

When using a trained model to make predictions [24], it is necessary to input test data and perform forward propagation.
At this time, the output value of the layer that has undergone dropout processing needs to be multiplied by the dropout probability during training on the basis of the original output. Although the network discards units with a certain probability through dropout during training, all units are still used during recognition. Therefore, the number of units with output values will increase the inverse times of the dropout probability. Since the units are connected by weight coefficients, they need to be multiplied by the probability. The comparison chart before and after the neural network application dropout is shown in Figures 2 and 3.

Standard backpropagation learning establishes a fragile collaborative adaptation mechanism. This mechanism is suitable for training datasets, but not for data that is not involved in training. Dropout breaks this collaborative adaptation mechanism by randomly hiding any hidden layer unit. It makes the neural network not excessively dependent on a specific input feature, to achieve the purpose of suppressing overfitting.

3.1.3. Adam. The optimization algorithm of deep learning is crucial to the training of the model. It adjusts the learning rate of each weight in the neural network through the moment estimation of the gradient. The algorithm uses the powerful function of the adaptive learning rate method to design a separate learning rate for different parameters.

In order to estimate the momentum, Adam uses the exponential moving average to calculate the gradient of the current minibatch:

\[
\begin{align*}
m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t, \\
v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2,
\end{align*}
\]

where \(m\) and \(v\) are moving averages, \(g\) represents the gradient of the current minibatch, \(\beta_1\) and \(\beta_2\) are the newly introduced hyperparameters of the algorithm, and their default values are 0.9 and 0.999, respectively.

Due to the influence of the moving average, the estimate obtained above is biased, and the revised estimate is

\[
\begin{align*}
m_t' &= \frac{m_t}{1 - \beta_1^n}, \\
v_t' &= \frac{v_t}{1 - \beta_2^n}, \\
\omega_t &= \omega_{t-1} - \eta \frac{m_t'}{\sqrt{v_t'} + \epsilon}
\end{align*}
\]

3.2. Hybrid Genetic Algorithm

3.2.1. Genetic Algorithm. Choosing the optimal parameters for deep learning tasks is very challenging. Improper selection of the initial values of the learning parameters may result in noise in the data or weak learning ability of the learning algorithm used. Therefore, genetic algorithms can be used to automatically find the optimal learning parameters. The genetic algorithm diagram is shown in Figure 4.

To select the best individual, a fitness function needs to be used. The result of the fitness function represents the quality of the solution, that is, the fitness of the individual. The higher the fitness, the better the quality of the solution.

3.2.2. Sequential Quadratic Programming. SQP combines two basic algorithms for solving nonlinear optimization problems: the active set method and the Newton method. It has a solid theoretical foundation and is designed to solve large-scale technical-related problems. It provides powerful algorithm tools. The constrained nonlinear optimization problem can be written as

\[
\begin{align*}
\min & \quad f(x), \\
h(x) &= 0, \\
g(x) &\leq 0,
\end{align*}
\]

where \(x\) is the vector of design variables, and \(f(x)\) is the objective function.
3.2.3. Hybrid Genetic. The genetic algorithm tends to converge to the local optimal solution rather than the global optimal solution. It can be improved by using a larger population, but if a larger initial population is used, the calculation amount of the algorithm will be greatly increased; if a smaller number is used population, the algorithm may not be able to find the optimal solution. Although the genetic algorithm can quickly converge to the region near the optimal solution, it still requires a huge amount of calculation to achieve the final convergence in the nearby region. The sequential quadratic programming algorithm has the advantages of faster calculation speed and good boundary search. Therefore, a hybrid genetic algorithm that combines genetic algorithm and sequential quadratic programming algorithm is selected. The genetic algorithm is first used to run near the optimal solution, and then the SQP optimization algorithm is used to perform a more efficient and fast local search to find the global optimal solution.

3.3. Our Algorithm. To convert the teaching quality prediction problem of the flipped classroom of college ideological and political courses into regression modeling problems, this research proposes a hybrid genetic algorithm-optimized LSTM network method HGA-LSTM to find the best hidden layer unit for teaching quality prediction quantity and dropout probability. Generally, the larger the dataset is, the more the hidden layers and neurons can be used for modeling without overfitting. Since the experimental dataset is large enough, higher accuracy can be achieved. If there are too few neurons in each layer, the predictive model will be difficult to adapt during training. Using more hidden layer units can better update the weights, but it also means more calculations and longer training time, so the number of hidden layer units is not as good as possible but has to be related to the number of datasets adapted. Since LSTM is easier to learn long-term dependence, in order to prevent the gradient from disappearing and exploding, choosing an appropriate dropout probability can effectively avoid the problem of data overfitting. The principle of the algorithm in this paper is shown in Figure 5.

4. Experiments and Results

4.1. Experimental Environment. In the process of neural network training, the setting of hyperparameters will greatly affect the performance of the network model. This article uses a da as the optimizer of our algorithm. We use GTX 1060 to accelerate the entire training process, the running platform is win 10, and the running memory is 8 GB. The entire model is built using the PyTorch framework. In the data processing stage, we extract 10% of the data from each category as the test set, 63% as the training set, and 27% as the validation set. The data in the test set is stored separately. During the training and verification process, the model cannot touch the test set data. Only in the testing phase, the test set data is read.

4.2. Dataset. This paper selects the historical data of the ideological and political courses in colleges and universities in a certain city and produces a training set and a test set.

4.3. Evaluation Index. To characterize the accuracy of the prediction model, error evaluation indicators are needed. Commonly used error evaluation indicators include average absolute error (MAE), root mean square error, and average absolute percentage error (MAPE). The calculation equation is as follows:

\[
\text{RMSE} = \sqrt{\frac{\sum_{n=1}^{N} |x_{(n)} - \bar{x}_{(n)}|^2}{N-1}}
\]

\[
\text{MAE} = \frac{\sum_{n=1}^{N} |x_{(n)} - \bar{x}_{(n)}|}{N}
\]

\[
\text{MAPE} = \left( \frac{\sum_{n=1}^{N} |x_{(n)} - \bar{x}_{(n)}|}{\sum_{n=1}^{N} x_{(n)}} \right) \times 100\
\]

where \(x_{(n)}\) represents the actual value and \(\bar{x}_{(n)}\) represents the predicted value.

4.4. Comparative Experiment. The performance comparisons of the four approaches are represented in Table 1. In the training phase, the RMSE errors of the method in this paper and GRU are 0.0194 and 0.0366, respectively. The RUL prediction based on HGA-LSTM is 0.0172 lower than the RMSE error of GRU, indicating that the prediction accuracy of HGA-LSTM is higher. In the test phase, the RMSE error and RZ based on HGA-LSTM are 0.0270 and 0.99_51, respectively. The RMSE error is 0.0193 less than using GRU and RZ is 0.0300 higher than using GRU, indicating that LSTM RNN has high prediction accuracy and good stability. These results show that the method based on HGA-LSTM has better performance than GRU on both the training set and the test set. SIM RNN cannot learn long-term dependence. The training set RMSE
is 0.0261 higher than HGA-LSTM, and the test set RMSE is 0.1099 higher. From the perspective of various errors, the method in this paper provides the best performance for predicting the quality of flipped classroom teaching in college ideological and political courses (as shown in Figure 6).

Table 1: Comparative experiment results.

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>MAPE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIM RNN</td>
<td>0.074</td>
<td>0.619</td>
<td>0.057</td>
</tr>
<tr>
<td>GRU</td>
<td>0.388</td>
<td>0.350</td>
<td>0.033</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.026</td>
<td>0.195</td>
<td>0.018</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>0.021</strong></td>
<td><strong>0.192</strong></td>
<td><strong>0.016</strong></td>
</tr>
</tbody>
</table>

Bold values represent the smallest values, which means the method achieves better performance than other methods.

Figure 5: Prediction and analysis of flipped classroom teaching quality.

Figure 6: Experimental results of our method.

Figure 7: Comparative experiment results of dropout. (a) No dropout. (b) Dropout.
### 4.5. Comparison Prediction for Dropout

From Figure 7, it is shown that the prediction effect of LSTM RNN using dropout is significantly better. The specific error is shown in Table 2. The RMSE of the training set and the test set without any antioverfitting algorithm are 0.0402 and 0.0992, respectively, and the error of the test set is 2.47 times of the RMSE of the training set, indicating that there is an obvious overfitting phenomenon. At the same time, we found that adding dropout can increase the model accuracy.

### 5. Conclusion

The rising of information technology and advancements have encouraged the flipped classroom in an operative and effective way. It flips knowledge transfer and knowledge internalization from two levels of teaching structure and teaching process, reversing the traditional teaching knowledge transfer in class and knowledge deepening after class from time and space. In this paper, we proposed a novel recurrent neural network method with a channel attention mechanism for music feature classification. Because the music classification method based on the convolutional neural network ignores the timing characteristics of the audio itself. In this regard, this article combines the proposed convolution structure with the bidirectional recurrent neural network, proposes a music classification model based on the convolution recurrent neural network, and uses the attention mechanism to assign different attention to the output of the recurrent neural network at different times. The classification accuracy of the model on the GTZAN dataset has increased to 93.1%, the AUC on the multilabel labeling dataset MagnaTagATune has reached 92.3%, surpassing other comparison methods, and the labeling of different music labels has been analyzed. This method has good labeling ability for most of the labels of music genres and also has good performance on some labels of musical instruments, singing, and emotion categories. The effectiveness of the proposed approach is shown from the experimental results.

### Data Availability

The data used to support the findings of this study are included within the supplementary information file(s).

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### References


