A Bloom Cognitive Hierarchical Classification Model for Chinese Exercises Based on Improved Chinese-RoBERTa-wwm and BiLSTM

Zhaoyu Shou[1], Yipeng Liu[2], Dongxu Li[3]*, Jianwen Mo[4], Huibing Zhang[5]
Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory,
Guilin University of Electronic Technology, Guilin, 541004, China[1]
School of Information and Communication, Guilin University of Electronic Technology, Guilin, 541004, China[2, 3, 4]
School of Computer and Information Security, Guilin University of Electronic Technology, Guilin, 541004, China[5]

***Abstract*—Assessing students' cognitive ability is one of the most important prerequisites for improving learning effectiveness, and the process involves aspects such as exercises, students' answers and teaching cases. In order to effectively assess students' cognitive ability, this paper proposes a Chinese text classification model that can automatically and accurately classify Bloom's cognitive hierarchy of exercises, starting from the exercises. Firstly, FreeLB perturbation is added to the input Embedding to enhance the generalization performance of the model, and Chinese-RoBERTa-wwm is used to obtain the pooler information and sequence information of the text; secondly, LSTM is used to extract the deep-associative features in the sequence information and combine with the pooler information to construct the semantically informative word vectors; lastly, the word vectors are fed into BiLSTM to learn the sequence bi-directional dependency information to obtain more comprehensive semantic features to achieve the accurate classification of the exercises. Experiments show that the model proposed in this paper significantly outperforms the baseline model on three Chinese public datasets, achieving 94.8%, 94.09% and 94.71% accuracies respectively, and also effectively performs the Bloom cognitive hierarchy classification task on two Chinese exercise datasets with less data.**

Keywords*—Chinese Text Classification; Chinese-RoBERTa-wwm; BiLSTM; Bloom Cognitive Hierarchy

I. Introduction

Assessing students' cognitive abilities is an essential component of the teaching process [1], as they are an important factor in determining the quality of learning activities and are indispensable in the learning process. Testing is an essential method for assessing students' cognitive abilities, and test scores correspond to learning outcomes at different cognitive levels, which in turn reflect students' cognitive abilities [2]. Therefore, it is necessary to develop test questions for courses according to a standard that meets different cognitive levels [3], helping teachers better grasp the cognitive abilities of students and achieve the teaching goals of the course, such as Bloom Taxonomy. The cognitive domain of Bloom taxonomy covers different cognitive levels from simple to complex [4, 5], categorizing exercises into six cognitive levels from high to low based on the criteria of different cognitive levels involved in students' learning processes [6]. Typically, teachers manually categorize exercises into the corresponding Bloom cognitive levels based on their understanding of the domain being taught, a process that is not only time-consuming but also highly subjective. Therefore, automating this process is a major task in pedagogical research.

Mohammed [7] and Setyaningsih [8] combined machine learning methods with natural language processing techniques to achieve good results with Bloom's Cognitive Hierarchical Taxonomy for test exercises in different courses. However, all of the above studies used traditional machine learning methods with low model accuracy and poor generalization [9]. Therefore, text classification methods need to be further improved to apply to the Bloom cognitive hierarchy domain. Compared with traditional machine learning algorithms, deep learning algorithms have emerged in the field of text classification [10]. Applying deep learning methods to Bloom's cognitive hierarchy of exercise classification has better results than machine learning methods [11]. In addition, the use of sequence learning models such as BiLSTM to capture the bidirectional dependency information in word vector sequences can improve the accuracy of classification models more effectively [12].

Inspired by the above research, this paper proposes a Chinese exercise Bloom cognitive hierarchy classification model based on improved Chinese-RoBERTa-wwm and BiLSTM (Chinese-FRLB). The model incorporates FreeLB antiperturbation into the input Embedding, combines sequence information and pooler information of the Chinese-RoBERTa-wwm output to generate word vectors, and learns their bi-directional sequence dependency information through BiLSTM to accurately classify the Bloom cognitive hierarchy of Chinese exercises.

The main contributions of this study can be summarized as follows:

1) A dataset of Chinese exercises was constructed based on Bloom classification hierarchy, providing a database for modeling the use of Chinese exercises to assess students' cognitive abilities in the learning process.

2) Proposing a word vector representation method that integrates both text sequence information and pooler

*Corresponding Author.

information. It utilizes LSTM to extract deep-level associative features from the sequence information outputted by Chinese-RoBERTa-wwm, enhancing the model's semantic comprehension ability. Additionally, it combines semantic information to characterize word vectors containing both deep and shallow semantic features, thereby avoiding the loss of con-textual global information and sentence structure information.

3) Proposing a Bloom cognitive hierarchical classification model for Chinese exercises. The model adds FreeLB adversarial perturbations to the input Embedding to train the model to distinguish between real samples and adversarial samples, enhance the generalization ability of the model, obtain semantically rich word vectors by using improved Chinese-RoBERTa-wwm, and inputs them into BiLSTM to learn bidirectional dependency information between word vector sequences, mine the implied dimensional correlation between words from the spatial level, and improve the classification accuracy of the text of the exercises more effectively.

The rest of the paper is organised as follows: Section II briefly describes the related work of this paper. Section III describes the proposed method and related techniques in detail, illustrating the composition of the experimental dataset. Section IV shows the comparison experimental results with other benchmark models and reliability analysis of the algorithm of this paper. Section V and Section VI concludes the work and provides future work with an outlook.

II. Related Works

The existing text classification methods are mainly categorized into ML method and DL method. The machine learning models ignore semantic information in texts and require manual labeling, which is time-consuming and laborious. In contrast, DL methods embed the text feature encoding process into model training, effectively extracting semantic information from texts and improving text classification accuracy.

Ashish et al. [13] proposed the Transformer architecture based on the attention mechanism, which provides a new deep-learning approach for text classification models. Under the Transformer architecture, large-scale pre-trained models have achieved tremendous success in the field of text classification. Devlin [14] proposed the pre-training model BERT, which adopts bidirectional training of Transformer and combines MLM (Masked LM) and NSP (Next Sentence Prediction) for pre-training on large-scale unlabeled corpora, fully learning the contextual implicit semantic information. Yang et al. [15] proposed the XLNet model, which integrates the advantages of both self-encoding and self-regression based on the autoregressive model Transformer-XL by adding the BERT model idea, combining the advantages of both autoencoding and autoregressive pre-training models. However, the above BERT and its improved model are applied to the Chinese classification task by segmenting according to a single text, which loses the semantic nature of Chinese words, and the ability to extract semantic information is weaker compared to the model in this paper.Liu et al. [16] proposed RoBERTa based on BERT, which removed the NSP task from BERT and expanded the training scale and training data, enabling RoBERTa to generalize better to downstream tasks than BERT. However, the BERT model only uses pooler information in text classification and does not fully consider the use of other feature information, whereas the combination of sequence information and pooler information used in this paper can learn more semantic features compared to the model. Cui et al [17] improved the Chinese version of BERT by proposing the whole word masking (wwm) strategy, i.e., masking the Chinese word, which improves the performance of the BERT model in the field of Chinese text classification. However, the BERT-base model only uses textual pooler information in text classification and does not fully consider utilizing other feature information. Xu [18] proposed a Chinese text classification method that synthesizes semantic and structural information, which uses cross-entropy and hinge loss to effectively combine Chinese-BERTology-wwm with the GCN method, demonstrating good performance in both long and short text corpora. However, in contrast to the model in this paper, the method does not take into account the interdependence information between sequences of word vectors.

Nowadays, DL demonstrates significant advantages in the field of text classification, and with the assistance of deep learning methods, various Bloom cognitive hierarchical classification models have also achieved good results. Shaikh et al. [19] used the Word2vec word embedding model to acquire word vectors with textual semantic information, which were then inputted into an LSTM model for performing Bloom cognitive hierarchical classification of Course Learning Outcomes (CLOs). However, the word vectors generated based on Word2vec do not encompass the contextual information of the input text. Mathiasen et al. [20] discarded the traditional word embedding technique and used the Transformer model to extract word vectors containing contextual information in job advertisements. Experimental results demonstrate that this model outperforms models using traditional word embedding techniques. Gani et al. [21] employed the RoBERTa model to extract word vectors from exercise questions' texts, which were then classified by a CNN network into different Bloom cognitive levels and the experimental results indicate that this method can more accurately categorize exercises into different Bloom cognitive levels. However, the above methods are all based on English exercises, and when faced with sparse data, the model may overfit prematurely, resulting in poorer robustness compared to the model in this paper.

In summary, there is a shortage of research on Bloom cognitive classification of Chinese text, and datasets are scarce. Existing deep learning-based Chinese text classification models do not fully consider the deep associative features in the output sequences of pre-trained models, and ignore the bidirectional semantic features of word vector sequences. In addition, most Bloom cognitive hierarchical classification models are only applicable to individual courses with poor generalization. Therefore, this paper constructs an exercise dataset containing multiple courses labeled with Bloom cognitive hierarchies and proposes a Bloom cognitive hierarchy classification model applicable to Chinese exercise texts. The model enhances its generalization ability by adding FreeLB adversarial perturbations to the input Embedding, achieving Bloom cognitive hierarchical classification of exercises from different courses. It utilizes LSTM to extract deep associative features from the output sequence information of Chinese-RoBERTa-wwm, and integrates semantic information to construct word vectors rich in semantic content. By employing BiLSTM, it learns the bidirectional dependency information of word vector sequences, accurately categorizing exercises into their corresponding Bloom cognitive levels.

III. Materials and Methods

The steps for processing Chinese exercise text using the Chinese-FRLB model are as follows: Firstly, preprocess the Chinese exercise text dataset to obtain Attention_mask and Input_ids vectors. Randomly initialize Inputs_embeds vectors based on the Input_ids vectors, and input Inputs_embeds and Attention_mask into Chinese-RoBERTa-wwm to obtain sequence information T and pooler information C. Then, use LSTM to extract deep sequential correlation features from sequence information T and combine it with pooler information C to represent word vectors Y. Finally, use BiLSTM to learn the bidirectional dependency information of the word vector sequence Y and combine it with Softmax to output the exercise classification results. The gradient parameters of the model are returned to the FreeLB module to calculate the perturbation values δ_t and added to the Inputs_embeds to participate in the training process of the model to increase the generalization performance of the model. The model architecture is shown in Fig. 1.

Algorithm 1 demonstrates the basic steps of the Chinese-FRLB model in classifying Chinese exercise texts into their corresponding Bloom cognitive levels from the perspective of data parameter variation.

Algorithm 1: Chinese-FRLB Model

Initialize: Traning dataset *X*, Ascent steps *K*, Perturbations bound ε, Ascent steps size α

Compute: Bloom Taxonomy Level *P*

According to *X*, obtain *Attention_mask* and *Token_types_ids*, and then according to *Inputs_embeds*, obtain *Token_types_ids*

For epoch = 1…*N* do

  For minibatch $B \subset X$ do

    Initialize FreeLB perturbations δ_0

    For t = 1…*K* do

      Sequence_output T, *Pooler_output C* ← RoBERTA-wwm(*Inputs_embeds* + δ_0)

      Obtain the deep associated features of text

      $H \leftarrow \text{LSTM}(T)$

      Obtain word vectors *Y* from contact *C* and *H*

      Obtain the sequence associated features of word vectors

      $H_{BiLSTM} \leftarrow \text{BiLSTM}(Y)$

      Reduce the dimensionality of H through the linear layer to obtain the vector *L*

      $P \leftarrow \text{Softmax}(L)$

      Update the perturbations δ_t

      $\delta_t \leftarrow \Pi_{\|\delta\|_{F\le\varepsilon}}(\delta_0 + \alpha \cdot g_{adv} / \|g_{adv}\|_F)$

    End

  End

End

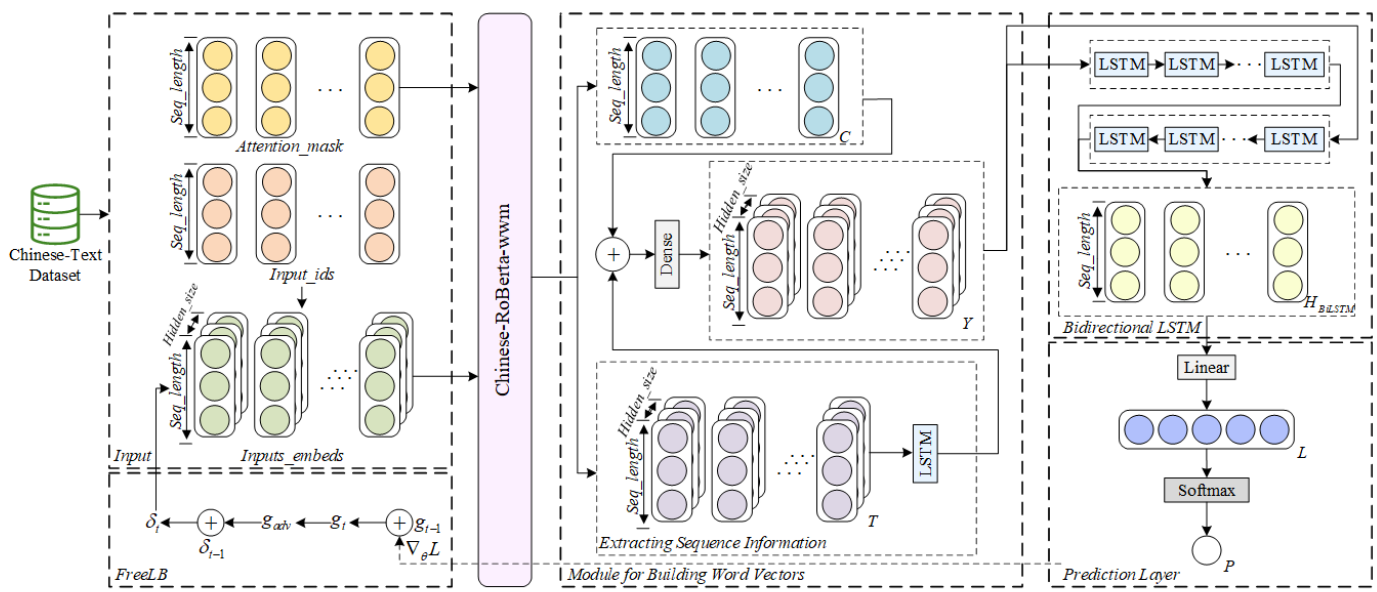


Fig. 1. Chinese-FRLB model diagram.

A. *Chinese-RL-wwm*

The study by Waheed et al [22] showed that combining pooler and sequential information to construct word vectors can enable the model to learn more profound semantic information, thereby enhancing classification accuracy. Inspired by this research, this paper proposes the Chinese-RL-wwm network framework, as illustrated in Fig. 2. Chinese-Roberta-wwm [17] is utilized to obtain the sequence information and the pooler information of a single sentence of Chinese text, and the LSTM model [23] is used to extract the deep associative features of the sequence information, and it is spliced with the semantic information as the word vector of the input text. Compared with the word vectors constructed by the Chinese-Roberta-wwm model, the word vectors output by the proposed network framework not only encompass all its features, but also consider the deep dependency information within the text sequence.

Using the statement "下列 C 语言常量中，错误的是 (Among the following C language constants, the incorrect one is)" as an example, firstly, the special tokens [CLS] and [SEP] are used to mark the beginning and end of the sentence, as shown in Fig. 2. Then, using the WordPiece splitter to split and construct the Embedding data (word embedding $\{E_{[CLS]},...\}$, text embedding $\{E_A,...\}$, and positional embedding $\{E_0,...\}$) as input to the RoBERTa-wwm model to obtain pooler information $C \in \mathbf{R}^H$ and sequential information $T_i \in \mathbf{R}^H$, where H is the number of hidden layers of the model, and i is the input data except [CLS]. Secondly, C and T_i are concatenated as the input of the LSTM model, which is used to mine the sequential dependency information of the preceding and following texts in the utterances, and characterize the deep-level association feature vector h_t of the utterances, the output of which corresponds to the input data including [CLS]. Finally, the semantic information of the sentence is represented by the word vector y_s obtained through Eq. (1), where s represents the input data including [CLS].

$$y_s = W(h_t \oplus C) + b \tag{1}$$

In the equation, W represents the weight parameters of the fully connected layer, and b denotes the bias parameters..

1) Chinese-RoBERTa-wwm: The embedding vectors required for BERT are obtained through Eq. (2).

$$E_{BERT} = E_{Token} + E_{Segment} + E_{Position} \tag{2}$$

The Chinese-RoBERTa-wwm model is built based on the bidirectional Transformer [13], and the network framework consists of a stack of 12 Encoder layers. The model processes the data as follows:

a) First, E_{BERT} is input into the multi-head attention mechanism as Q, K, and V in the attention mechanism, and the residuals of E_{BERT} and ATT_{OUT} are connected using the LayerNormalisation layer as shown in Eq. (3) and Eq. (4).

$$ATT_{OUT} = Attention(E_{BERT}, E_{BERT}, E_{BERT}, mask) \tag{3}$$

$$output_1 = LayerNorm(ATT_{OUT} + E_{BERT}) \tag{4}$$

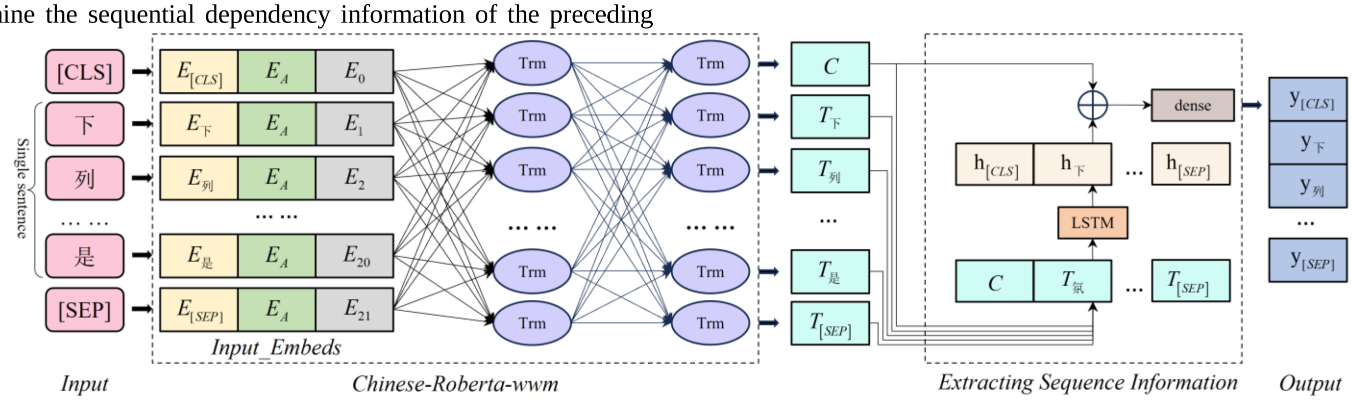


Fig. 2. Framework diagram of Chinese-RL-wwm module.

b) Next, $output_1$ is input into the Point-wise Feed Forward Neural Network layer. Then, the LayerNormalization layer is applied again to residually connect $output_1$ with the output results of the Point-wise Feed Forward Neural Network layer, yielding the output of the Encoder layer, as depicted in Eq. (5) and Eq. (6).

$$Feed\_Forward_{OUT} = FeedForward(output_1) \tag{5}$$

$$output = LayerNorm(output_1 + Feed\_Forward_{OUT}) \tag{6}$$

c) Finally, the output is fed into the next layer of Encoder network, and the final pooler information $C \in \mathbf{R}^H$ as well as the sequence information $T_i \in \mathbf{R}^H$ is output after 12 layers of stacked Encoder networks in turn.

2) LSTM: In this paper, the LSTM model [23] is utilized to characterize the deep-level associative features of the sequence information $T_i \in \mathbf{R}^H$ output from Chinese-RoBERTa. Before inputting into the LSTM model, the sequence information T_i is integrated into a sequence vector $T = \{T_1, T_2, ..., T_i\}$ according to the order of sentences. As shown in Fig. 3, the LSTM achieves the functions of selectively forgetting the information of the previous moment, selectively updating the information of the current moment, and selecting specific information as the output of the current moment through three gating units, namely, the forgetting gate f_t, the input gate i_t, and the output gate o_t, as shown in Eq. (7), Eq. (8) and Eq. (9).

$$f_t = Sigmoid(W_f \cdot [h_{t-1}, T_i] + b_f) \tag{7}$$

$$i_t = Sigmoid(W_i \cdot [h_{t-1}, T_i] + b_i) \tag{8}$$

$$o_t = Sigmoid(W_o [h_{t-1}, T_i] + b_o) \tag{9}$$

In the equation, W_f, W_i, W_C, b_f, b_i and b_C are trainable parameters. After passing through the three gates, the new candidate value vector $\tilde{C}_t$ and output vector h_t are computed as shown in Eq. (10), Eq. (11) and Eq. (12).

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, T_i] + b_C) \tag{10}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{11}$$

$$h_t = o_t * \tanh(C_t) \tag{12}$$

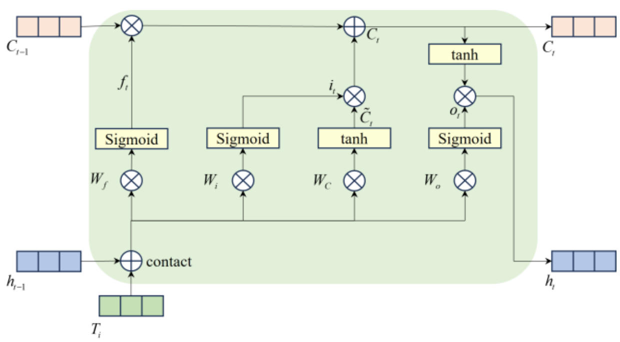


Fig. 3. Structure of LSTM gating mechanism.

In this case, the vector h_t represents deep-level associative features in the text sequence. And then, the word vector y_s which contains the semantic information of the text will be constructed by combining Eq. (1) with h_t and the pooler information *C*.

B. BiLSTM

Inspired by the [12], after obtaining word vectors y_s that combine the deep associative feature vectors of the utterances with the pooler information, the BiLSTM model is used to learn the bi-directional dependency information between the word vector sequences for the subsequent text categorization.

Pass the output $Y = \{y_{[cls]}, y_{下}, ..., y_{[sep]}\}$ of Chinese-RL-wwm from the first moment to moment *t* into the forward LSTM and save the output at each time point as shown in Eq. (13). Reversing the sequence of *Y* and then once time fed into the backward LSTM, while also saving the output at each time point, as shown in Eq. (14).

$$\vec{h}_t = Sigmoid(\vec{W}[\vec{h}_{t-1}, Y]) \tag{13}$$

$$\overleftarrow{h}_t = Sigmoid(\overleftarrow{W}[\overleftarrow{h}_{t-1}, Y]) \tag{14}$$

In the equation, $\vec{h}_{t-1}$ and $\overleftarrow{h}_{t-1}$ are the previous output states of the forward LSTM and backward LSTM at time *t*-1, respectively. $\vec{W}$ and $\overleftarrow{W}$ are the weights matrices of the forward and backward propagation, respectively.

Concatenate the output vectors corresponding to the forward LSTM and the backward LSTM to obtain the final required vector, as shown in Eq. (15).

$$H'_t = W_1 \vec{h}_t + W_2 \overleftarrow{h}_t + c_t \tag{15}$$

In the equation, W_1 and W_2 are the forward and backward output weights, respectively, and c_t is the bias optimization parameter.

C. FreeLB

In order to improve the generalization of the model, FreeLB (Free Large-Batch) [24] adversarial perturbations are added to the input embedding layer of the text data. During the model training process, the gradient parameters $\nabla_\theta L$ accumulated by each perturbation are calculated firstly, as shown in Eq. (16).

$$g_t = g_{t-1} + \frac{1}{K} \mathrm{E}_{(Z,y)\in B}[\nabla_\theta L(f_\theta(\boldsymbol{X} + \delta_{t-1}), y)] \tag{16}$$

In the equation, g_{t-1} and δ_{t-1} are the gradient and perturbation at the previous moment, and $X + \delta_{t-1}$ is the approximation of the local maximum at the intersection of two spherical neighborhoods $\mathrm{L}_t = \mathrm{B}_{\boldsymbol{X}+\delta_o}(\alpha t) \cap \mathrm{B}_{\boldsymbol{X}}(\varepsilon)$.

Then, the perturbation δ_t is updated by gradient ascent as shown in Eq. (17).

$$\delta_t = \Pi_{\|\delta\|_{F \le \varepsilon}} (\delta_{t-1} + \alpha \cdot g_{adv} / \|g_{adv}\|_F) \tag{17}$$

Eventually, the gradient parameter g_K obtained after *K* iterations is used for updating the model parameters θ, as shown in Eq. (18).

$$\theta = \theta - \tau g_K \tag{18}$$

D. Datasets

To verify the advancement and effectiveness of the proposed Chinese-FRLB model, experiments were conducted on three sets of Chinese news headline datasets and two sets of Chinese Bloom cognitive level exercise datasets classified by course instructors, namely Toutiao-S[1], THUCNews #1[1], THUCNews #2[1], Bloom-5classes, and Bloom-6classes datasets.

The Toutiao-S dataset is a subset of the Toutiao Chinese news headline classification corpus, containing five categories with 17,500 news headlines. Among them, 15,000 headlines are used for training, and 2,500 headlines are used for testing.

Based on the original Sina news classification system, two sets of datasets with different classification categories were redefined: THUCNews #1 and THUCNews #2. THUCNews #1 contains 10 categories and 55,315 news headlines, with 45,315 headlines used for training and 10,000 headlines used for testing. THUCNews #2 contains 14 categories and 54,599 news headlines, with 41,999 headlines used for training and 12,600 headlines used for testing.

TABLE I. STATISTICAL DATA FOR DIFFERENT DATA SETS

Datasets	#Proportion of Different Categories of Training and Testing Dataset	#Average length
Toutiao-S	All 20%	25
THUCNews #1	All 10%	19
THUCNews #2	All 7.1%	18
Bloom-5classes	30%/25%/23%/11%/11%	50
Bloom-6classes	All 16.7%	52

The Bloom-5classes dataset was selected from the final exam papers and textbook exercises of the "C Programming Language" [25] course. After discussion with the teacher of the course area, it was concluded that Bloom's cognitive level of analysis is suitable for program questions, but not for textual descriptions such as multiple-choice, fill-in-the-blanks, and programming questions. However, Chinese-RoBERTa-wwm could not handle program questions, so the dataset had only five categories of Bloom cognitive hierarchy: memorization, comprehension, application, evaluation, and creativity, and 1011 exercise questions, of which 807 were training and 204 were testing.

[1]https://github.com/anglgn/Chinese-Text-Classification-Dataset

The Bloom-6classes dataset is based on the Bloom-5classes dataset and includes exam papers, textbook exercises, and MOOC question banks from the "Introduction to Computer Science" [26] course, as well as subjective questions related to Bloom's cognitive levels and exercises from the textbook "Computer Science: An Interdisciplinary Approach" [27] by Princeton University Press. It contains six categories of Bloom's cognitive levels, with a total of 2,824 exercise questions, of which 2,122 are for training and 702 for testing. Table I summarizes the category distribution and other statistics of the five datasets.

IV. RESULT AND DISCUSSION

A. Baseline Model and Assessment Indicators

In order to evaluate the performance of the proposed Chinese-FRLB model, ACC and F1-Score are used as evaluation metrics. Five baseline models are compared with the Chinese-FRLB model on five datasets, as shown in Table II.

B. Experimental Environment and Experimental Hyperparameter Settings

The experimental environment of this paper is shown in Table III.

In the model, the maximum sequence lengths for input text sequences are set to 60, 32, 32, 512, and 512 for the Toutiao-S, THUCNews #1, THUCNews #2, Bloom-5classes, and Bloom-6classes datasets, respectively. After multiple experimental comparisons, the learning rate for the Chinese-RoBERTa-wwm module was set to 8e-5. As for the LSTM and BiLSTM modules, due to their smaller model parameters, the learning rate was set to 10 times that of the Chinese-RoBERTa-wwm module, which is 8e-4. The weight decay coefficients for all three modules are set to 1e-5. FreeLB set the learning rate to 4.5e-2 as per the original papers' reference, and the initialization delta is set to 5e-2. For text data, which features sparse characteristics, the performance of LSTM and BiLSTM is optimal when the number of layers is set to 1. The hidden layer embedding dimensions for the Chinese-RoBERTa-wwm module are set to the original standard of 768， the hidden layer embedding dimensions for LSTM are set to 512. In order to ensure effective handling of word vectors output by the Chinese-RL-wwm module, the hidden layer embedding dimensions for BiLSTM are set to 768 to match those of the Chinese-RoBERTa-wwm module. The Dropout regularization parameter is set to 0 according to the requirements of the FreeLB adversarial training, the iteration number of stochastic gradient descent is set to 50, and the batch size of the model is set to 32, which is optimized using the Adam optimizer. Gradient descent optimization using Adam optimizer.

In order to save the computational resources of the proposed model in this paper and realize the lightweight deployment in realistic scenarios, this paper adjusts the number of hidden layers of RoBERTa base in Chinese-RoBERTa-wwm module downward from 12 to 6. The effectiveness of this method is validated on the Bloom-5classes and Bloom-6classes datasets.

TABLE II. BASELINE MODEL

Baselines	Description
XLNet[19]	Using the Attention mask inside the Transformer and combining it with the dual-stream attention mechanism.
Chinese-BERT-wwm[21]	Pre-trained model using bidirectional Transformer and wwm task and NSP task on large Chinese datasets.
Chinese-RoBERTa-wwm[21]	Pre-trained model using bidirectional Transformer and enhanced dynamic wwm task on larger Chinese datasets.
Chinese-BERT-wwm-GCN-LP[22]	Combining the Chinese-BERT-wwm model with a text-constructed heterogeneous graph GCN model using cross-entropy and hinge loss.
Chinese-RoBERTa-wwm-GCN-LP[22]	Combining the Chinese-RoBERTa-wwm model with a text-constructed heterogeneous graph GCN model using cross-entropy and hinge loss.

TABLE III. EXPERIMENTAL ENVIRONMENT

Experimental environment	Environment configuration
Operating systems	Linux
CPU	Intel(R) Xeon(R) Gold 6330H
Video Cards	GeForce RTX 3090
RAM	32GB
ROM	1T SSD
Programming Languages	Python 3.8
Framework	Pytorch

TABLE IV. EXPERIMENTAL RESULTS OF CHINESE-ROBERTA-WWM MODULE WITH DIFFERENT NUMBER OF HIDDEN LAYER LAYERS

Dataset	Num of Hidden Layers	#Acc	#F1	#Model Parameters
Bloom-5classes	Six	**0.7255**	**0.7242**	**72998134**
	Twelve	0.7206	0.7191	115525366
Bloom-6classes	Six	**0.6937**	**0.6974**	**72998134**
	Twelve	0.6795	0.6813	115525366

As shown in Table IV, for the Bloom cognitive hierarchical dataset with less data and categories, reducing the number of hidden layers not only drastically reduces the number of parameters of the model and shrinks the training time of the model, but also optimizes the structure of the model, leading to improved classification accuracy.

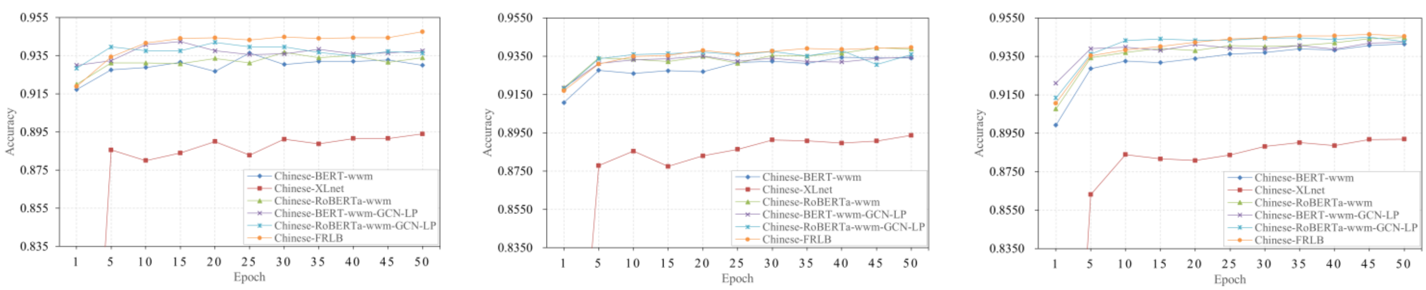
C. Experiments and Analysis of Results

1) Analysis of experimental results based on three sets of Chinese news headline datasets: The ACC curves of each model on the Toutiao-S, THUCNews #1 and THUCNews #2 datasets are shown in Fig. 4 as (a), (b) and (c).

Fig. 4(a) corresponds to a dataset with five categories, and it can be seen that the ACC value of this model starts to lead the baseline models after 10 epochs, and shows a clear advantage over the baseline models after 15 epochs; Fig. 4(b) corresponds to a dataset of 10 categories, and due to the more parameters of this model, the ACC value of this model starts to lead the baseline models only after 20 epochs, but it still has an advantage over the baseline models; Fig. 4(c) corresponds to a dataset of 14 categories, and due to the increase of the categories and the difficulty of the classification, the ACC value of this model starts to lead the baseline models only after 25 epochs, but it has a better classification performance and better robustness compared to the baseline models. In summary, the proposed model has an advantage over other baseline models when the dataset categories are reduced, which side by side indicates that it can be better utilized for the less-category Bloom cognitive hierarchical classification task.

As can be seen from Fig. 4, the model in this paper is not effective at the beginning of training in each dataset, which is due to the fact that the model adds FreeLB antagonistic perturbation at the input and has more modules, with a slightly larger number of parameters than that of the other baseline models, resulting in a more complex model structure and slower convergence in the early stage. However, unlike Chinese-RoBERTa-wwm-GCN-LP, which incorporates the use of GCN modules, the model in this paper does not fluctuate due to the change in the performance of the baseline model, which proves that the addition of FreeLB perturbation not only enhances the model's generalization ability, but also improves the robustness of the model.

The performance metrics of the Chinese-FRLB model on THUCNews #1, THUCNews #2, and Toutiao-S datasets versus other baseline models are shown in Table V and analyzed as follows:



(a) (b) (c)

Fig. 4. ACC curves of Chinese-FRLB model and baseline models on Chinese news headlines dataset: (a) Comparison between models based on the Toutiao-S dataset; (b) Comparison be-tween models based on the THUCNews #1 dataset; (c) Comparison between models based on the THUCNews #2 dataset.

TABLE V. PERFORMANCE METRICS OF ALL MODELS ON THE CHINESE NEWS HEADLINES DATASET

Model	Datasets					
	Toutiao-S		*THUCNews #1*		*THUCNews #2*	
	#ACC	*#F1*	*#ACC*	*#F1*	*#ACC*	*#F1*
Chinese-BERT-wwm(12 of hidden layers)	0.9368	0.9368	0.9350	0.9351	0.9417	0.9417
Chinese-XLnet	0.8972	0.8973	0.8937	0.8938	0.8926	0.8924
Chinese-RoBERTa-wwm(12 of hidden layers)	0.9376	0.9377	0.9396	0.9397	0.9450	0.9450
Chinese-BERT-wwm-GCN-LP	0.9424	0.9424	0.9356	0.9357	0.9426	0.9425
Chinese-RoBERTa-wwm-GCN-LP	0.9432	0.9432	0.9385	0.9385	0.9462	0.9461
Chinese-FRLB(12 of hidden layers)	**0.9480**	**0.9480**	**0.9409**	**0.9409**	**0.9471**	**0.9472**

a) Compared with the baseline models, the proposed model in this paper achieves the best performance in terms of ACC and F1-Score on all three public datasets.This indicates the state-of-the-art of the proposed model in this paper in the field of Chinese classification.

b) The Chinese-BERT-wwm-GCN-LP and Chinese-RoBERTa-wwm-GCN-LP models combine the Chinese-BERTology-wwm and GCN modules to combine semantic and structural information of the text. The method proposed in this paper, which comprehensively utilizes the Chinese-RoBERTa-wmm module and LSTM module to combine semantic and sequential information, outperforms Chinese-BERT-wwm-GCN-LP and Chinese-RoBERTa-wwm-GCN-LP models. This proves that the method in this paper is more effective in extracting the implicit deep and shallow semantic features of Chinese text.

c) The model proposed in this paper combines Chinese-RoBERTa-wwm and BiLSTM. Compared to Chinese-BERT-wwm and Chinese-RoBERTa-wwm, BiLSTM is employed to extract bidirectional semantic features between word vector sequences before using fully connected layers for text classification. The results show that the performance of the model proposed in this paper is superior, indicating that the combined use of sequence learning models effectively enhances the overall Chinese text classification capability of the model.

d) The model proposed in this paper is constructed based on the Chinese-RoBERTa-wwm model, and compared to the Chinese-XLnet model, the Chinese-RoBERTa-wwm model performs better. It shows that the Chinese-RoBERTa-wwm model can learn more bi-directional contextual information..

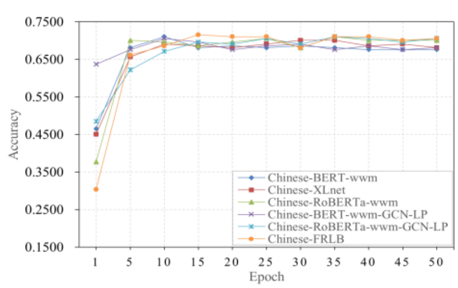
2) Analysis of Experimental Results Based on Two Sets of Bloom Cognitive Hierarchy Exercise Datasets: Fig. 5 and Fig.6 show the ACC curves of each model on the Bloom-5classes and Bloom-6classes datasets, respectively. The experiments in this section aim to verify the effectiveness of the proposed model in reducing the number of RoBERTa base hidden layers in the Chinese-RoBERTa-wwm base model. The number of hidden layers in the BERT base and RoBERTa base of the Chinese-BERT-wwm, Chinese-RoBERTa-wwm and Chinese-FRLB structures are set to be 12 and 6 respectively, while the number of hidden layers in the Chinese-BERT-wwm-GCN-LP and Chinese-RoBERTa-wwm-GCN-LP remains unchanged. Fig. 5 shows the experimental results for 12 layers in the hidden layer, while Fig. 6 shows the results for 6 layers.

The following conclusions can be drawn from Fig. 5 and Fig. 6:

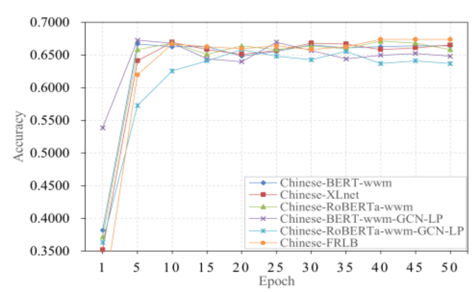
a) The results in Fig. 5(a) show that this paper's model slightly outperforms each baseline model after 10 epochs. In addition, Fig. 5(b) shows that this paper's model has a clear advantage over the baseline model for the Bloom 6-classes dataset, which contains different course exercises. Comparing Fig. 5 and Fig. 6, it is evident that the Chinese-FRLB model with six of RoBERTa base hidden layers is more stable and performs better. This suggests that for the two small Bloom Chinese exercise datasets, reducing the RoBERTa base hidden layers to a 6-layer model is more effective.

b) As shown in Fig. 6(a) and Fig. 6(b), the individual baseline models may show too much variation within different datasets due to the small size of the dataset and the large span of the domain, which manifests itself as a problem of poor generalization ability, resulting in the inability to be widely applied to the classification of exercises in different courses. Compared with the model proposed in this paper, it is demonstrated that the use of FreeLB adversarial training can effectively stabilize the gradient update of the model and improve the robustness and generalization of the model.

c) At the start of training, the model proposed in this paper was ineffective due to the sparse dataset and inefficient learning in the early stages. However, the model began to show advantages in the middle of training. Fig. 6(a) shows that after 15 epochs, this paper's model outperforms the baseline models in terms of ACC value. Fig. 6(b) further demonstrates that after 25 epochs, this paper's model significantly outperforms the baseline models. These results suggest that the Chinese-RL-wwm module utilized by this paper's model is capable of extracting deep semantic information even with a smaller dataset size.



(a)



(b)

Fig. 5. ACC curves of Chinese-FRLB model(12 of hidden layers) with the baseline model on two Bloom cognitive hierarchy exercise datasets: (a) Comparison between models based on the Bloom-5classes dataset; (b) Comparison between models based on the Bloom-6classes dataset.

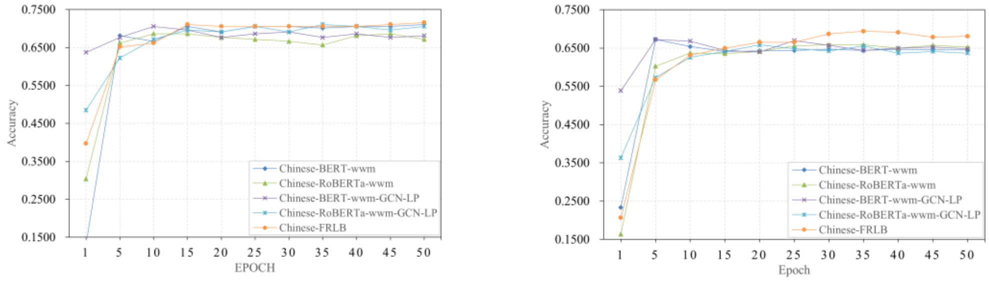


Fig. 6. ACC curves of Chinese-FRLB model(6 of hidden layers) with the baseline model on two Bloom cognitive hierarchy exercise datasets: (a) Comparison between models based on the Bloom-5classes dataset; (b) Comparison between models based on the Bloom-6classes dataset.

Since the structural parameters of the XLnet model are different from those of the BERT model, it is not used as a baseline model for comparison in this Fig. 6

Table VI presents the performance metrics of the Chinese-FRLB model proposed in this paper on Bloom-5classes and Bloom-6classes datasets with other baseline models. The analysis is as follows:

a) The Chinese FRLB model proposed in this article outperforms the baseline model in terms of ACC and F1 Score evaluation metrics on two small-scale Chinese Bloom cognitive level exercise datasets, using a 12 layer BERT base hidden layer Chinese RoBERTa wwm basic model. This indicates the effectiveness of the proposed model in the Chinese Bloom cognitive level classification task.

b) The model proposed in this paper, Chinese-FRLB, improves performance when the number of BERT base hidden layers is reduced from 12 to 6. This is due to the reduction of the BERT base hidden layers of the Chinese-FRLB base model Chinese-RoBERTa-wmm effectively optimizes the structure of the model, which is helpful to prevent the model from focusing too much on global information and ignoring important local information when extracting text features with a small number of data sets, which indicates that for a small number of data sets, simplifying the structure of the text feature extraction model can help to improve the text classification ability.

D. Ablation experiment

1) Impact of individual modules on model performance: In order to verify the effects of FreeLB against perturbations, Chinese-RL-wwm module and BiLSTM module on the Chinese-FRLB model, ablation experiments are conducted on five datasets in this paper. The experimental results are shown in Table VII, and the following conclusions can be obtained:

a) From the results of the public datasets, it can be seen that the method of using LSTM to extract deep associative features and combining pooler information to represent word vectors can effectively obtain rich semantic information features. Learning bidirectional dependency information of word vector sequences using BiLSTM can further improve the classification performance of the model. Adding FreeLB adversarial training can enhance the robustness of the models composed of different modules. Their contributions to the overall model are not identical, but removing any one of them would result in performance degradation, indicating that the introduction of these three modules is effective on public datasets, and their functions in the model are complementary.

b) From the results of the Bloom cognitive hierarchy exercise datasets, it can be observed that for this kind of dataset with low data volume and domain spanning, the deep association features extracted by combining LSTM can focus on more sequence association features than the word vectors output by Chinese-RoBERTa-wwm, thereby obtaining more semantic information for the model and compensating for the feature sparsity due to the low data volume; using BiLSTM can extract the implicit deep semantic features between word vector sequences on the basis of the former, enhancing the classification ability of the model under small data volume conditions; adding FreeLB adversarial training significantly improves classification performance and generalization ability of Bloom's cognitive hierarchy exercises with small amounts of data and covering different courses.

TABLE VI. PERFORMANCE METRICS FOR ALL MODELS ON BLOOM'S COGNITIVE HIERARCHY EXERCISE DATASET

Model	**Datasets**			
	Bloom-5classes		***Bloom-6classes***	
	#ACC	***#F1***	***#ACC***	***#F1***
Chinese-BERT-wwm(12 of hidden layers)	0.7108	0.6929	0.6709	0.6732
Chinese-BERT-wwm(6 of hidden layers)	0.7206	0.7157	0.6724	0.6742
Chinese-XLnet	0.7059	0.6903	0.6781	0.6759
Chinese-RoBERTa-wwm(12 of hidden layers)	0.7157	0.7019	0.6752	0.6761
Chinese-RoBERTa-wwm(6 of hidden layers)	0.7010	0.6993	0.6624	0.6629
Chinese-BERT-wwm-GCN-LP	0.7108	0.7092	0.6738	0.6754
Chinese-RoBERTa-wwm-GCN-LP	0.7206	0.7189	0.6610	0.6637
Chinese-FRLB (12 of hidden layers)	0.7206	0.7191	0.6795	0.6813
Chinese-FRLB(6 of hidden layers**)**	**0.7255**	**0.7242**	**0.6937**	**0.6974**

TABLE VII. CHINESE-FRLB MODEL ABLATION EXPERIMENT RESULTS

Model	Datasets									
	Toutiao-S		*THUCNews #1*		*THUCNews #2*		*Bloom-5classes*		*Bloom-6classes*	
	#ACC	*#F1*	*#ACC*	*#F1*	*#ACC*	*#F1*	*#ACC*	*#F1*	*#ACC*	*#F1*
Chinese-RLB	0.9464	0.9464	0.9392	0.9393	0.9433	0.9433	0.7108	0.7109	0.6610	0.6627
Chinese-FRL	0.9460	0.9460	0.9405	0.9405	0.9422	0.9423	0.7010	0.7013	0.6738	0.6782
Chinese-FRB	0.9460	0.9460	0.9406	0.9405	0.9448	0.9449	0.7059	0.7063	0.6724	0.6775
Chinese-FRLB	**0.9480**	**0.9480**	**0.9409**	**0.9409**	**0.9471**	**0.9472**	**0.7255**	**0.7242**	**0.6937**	**0.6974**

In each ablation experiment, one module is removed from the Chinese-FRLB model to evaluate the effectiveness of each module. Specifically, Chinese-RLB refers to removing the FreeLB perturbation, Chinese-FRL refers to removing the BiLSTM module, and Chinese-FRB refers to using the original Chinese-RoBERTa-wwm module. In the datasets of Bloom-5classes and Bloom-6classes, using the RoBERTa base with 6 of hidden layer layers.

2) The effect of the number of hidden layers in the RoBERTa base: In order to verify the sophistication of the Chinese-FRLB model in setting the number of hidden layers of RoBERTa base in the Chinese-RoBERTa-wwm structure to 6, this paper carries out the ablation experiments with different numbers of hidden layer layers on two datasets, and the experimental results are shown in Table VIII.

From Table VIII and the analysis of the experimental process, it can be seen that reducing the number of hidden layers of RoBERTa base will result in the model being unable to effectively extract the text feature information, resulting in a significant decrease in its performance. When the number of hidden layers in RoBERTa base is too high, then the training time and the space complexity of model increase, leading to increased memory consumption. Moreover, it may cause the model to overlook locally important information, resulting in suboptimal classification performance. When the number of hidden layers of RoBERTa base is set to 6, all the metrics are the best.

TABLE VIII. PERFORMANCE METRICS OF CHINESE-FRLB MODEL AT DIFFERENT NUMBER OF ROBERTA BASE HIDDEN LAYER LAYERS

Num of Hidden Layers	Datasets			
	Bloom-5classes		*Bloom-6classes*	
	#ACC	*#F1-Score*	*#ACC*	*#F1-Score*
Four	0.6814	0.6801	0.6425	0.6414
Six	**0.7255**	**0.7242**	**0.6937**	**0.6974**
Eight	0.7010	0.6991	0.6524	0.6500
Ten	0.7010	0.6993	0.6752	0.6753
Twelve	0.7206	0.7191	0.6795	0.6813

V. CONCLUSIONS

To assist teachers in accurately categorizing Chinese exercises into the corresponding Bloom levels and accurately assessing students' cognitive abilities, this paper proposes a Bloom cognitive level classification model for Chinese exercises based on the Bloom classification method. Specifically, the model utilizes sequence information and pooler information to model word vectors and combines a BiLSTM sequence learning model. The introduced FreeLB adversarial perturbation exhibits better stability in two small-scale Chinese Bloom cognitive level exercise datasets. In the word vector representation stage, LSTM extracts deep associative features combined with pooler information to effectively construct word vectors with rich semantic feature information. During classification, the semantic features of word vectors extracted by BiLSTM further improve the accuracy of the model in classifying different datasets. Experimental results on three Chinese public datasets and two sets of Chinese Bloom cognitive level exercise datasets demonstrate that the proposed model accurately classifies the Bloom levels of Chinese exercises and also performs well in other text classification tasks. In addition, this paper conducts ablation experiments on three sub-modules in the model, and the results show that all three modules can effectively improve the overall performance of the model.

VI. FUTURE WORK

The Chinese-FRLB model proposed in this paper can effectively classify the Bloom cognitive level of exercises, but there are still problems such as semantic ambiguity and data sparsity that need further refinement and improvement. Therefore, in our future research, we will work on enriching the text feature representation by using glyph and pinyin information as well as deeper lexical extraction techniques to better capture the semantic information of Chinese text. In addition, we will explore more advanced word segmentation models and methods combining multiple techniques to optimize the text pre-processing and feature extraction processes, and consider lightweight models to further improve the running speed of the models.

ACKNOWLEDGMENT

The National Natural Science Foundation of China (62177012, 61967005, 62267003).

The Project of Guangxi Wireless Broadband Communication and Signal Processing Key Laboratory (GXKL06240107).

References

[1] Ullah Z , Lajis A , "Jamjoom M ,et al.A Rule-Based Method for Cognitive Competency Assessment in Computer Programming Using Bloom's Taxonomy," IEEE Access, 2019, pp(99):1-1.

[2] Mulatsih B. "Implementation of Revised Bloom Taxonomy in Develope Chemistry Questions in the Domain of Knowledge," Ideguru: Jurnal Karya Ilmiah Guru, vol. 6(1), pp. 1-10, 2021.

[3] Haris S S, Omar N. "Bloom's taxonomy question categorization using rules and N-gram approach," Journal of Theoretical & Applied Information Technology, vol. 76(3), 2015.

[4] Subiyantoro E , Ashari A , "Suprapto.Cognitive Classification Based on Revised Bloom's Taxonomy Using Learning Vector Quantization," 2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM), 2020, pp: 1-8.

[5] Kennedy D, Writing and using learning outcomes: a practical guide, University College Cork, 2006.

[6] Lajis A, Nasir H M, Aziz N A, "Proposed assessment framework based on Bloom taxonomy cognitive competency: Introduction to programming," Proceedings of the 2018 7th International Conference on Software and Computer Applications, 2018, pp. 97-101.

[7] Mohammed M , Omar N , "Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec," PLoS ONE, 2020, 15(3):e0230442.

[8] Setyaningsih E R, Listiowarni I, "Categorization of exam questions based on Bloom taxonomy using naïve bayes and laplace smoothing," 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT). 2021: IEEE, pp. 330-333.

[9] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," IEEE Comput. Intell. Mag.,vol. 13, no. 3, pp. 55–75, Aug. 2018.

[10] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, "Deep learning for hate speech detection in tweets," International World Wide Web Conferences Steering Committee (IW3C2), 2017, pp. 759–760.

[11] Romadhony A, Abdurohman R, "Primary and High School Question Classification based on Bloom's Taxonomy," 2022 10th International Conference on Information and Communication Technology (ICoICT). 2022: IEEE, pp. 234-239.

[12] Zhang Y, Liu J, "Research on Chinese Sentiment Classification Based on BERT-wwm-BiLSTM-SVM," 2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT), 2023: IEEE, pp. 832-837.

[13] Vaswani A , Shazeer N , Parmar N ,et al. "Attention Is All You Need," Advances in neural information processing systems, 2017, pp: 1-9.

[14] Devlin J, Chang M W, Lee K, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[15] Z. Yang, Z. Dai, Y. Yang, J.G. Carbonell, R, Salakhutdinov and Q.V. Le, "XLnet: Generalized autoregressive pretraining for language understanding," Advances in neural information processing systems, 2019, 32.

[16] Liu Y , Ott M , Goyal N ,et al. "RoBERTa: A Robustly Optimized BERT Pretraining Approach," 2019, pp: 1-10.

[17] Cui Y, Che W, Liu T, Qin B, Yang Z, "Pre-training with whole word masking for ChineseBERT," IEEE-ACM Transactions on Audio Speech and Language Processing , vol. 29, pp. 3504–3514, 2021.

[18] Xu X, Chang Y, An J, et al. "Chinese text classification by combining Chinese-BERTology-wwm and GCN," PeerJ Computer Science, vol. 9, pp. e1544, 2023.

[19] Shaikh S, Daudpotta S M, Imran A S, "Bloom's learning outcomes' automatic classification using lstm and pretrained word embeddings," IEEE Access, vol. 9, pp. 117887-117909, 2021.

[20] Mathiasen M, Nielsen J, Laub S, "A Transformer Based Semantic Analysis of (non-English) Danish Jobads," Proceedings of the 15th International Conference on Computer Supported Education (CSEDU), 2023: IEEE, Volume 1, pp. 359-366.

[21] Gani M O, Ayyasamy R K, Sangodiah A, et al. "Bloom's Taxonomy-based exam question classification: The outcome of CNN and optimal pre-trained word embedding technique," Education and Information Technologies, 2023, pp. 1-22.

[22] Waheed A , Goyal M , Mittal N ,et al. "BloomNet: A Robust Transformer based model for Bloom's Learning Outcome Classification," 2021, pp: 1-8.

[23] Mousa A , Schuller B , "Contextual Bidirectional Long Short-Term Memory Recurrent Neural Network Language Models: A Generative Approach to Sentiment Analysis," Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Volume 1, 2017.

[24] Zhu C , Cheng Y , Gan Z ,et al. "FreeLB: Enhanced Adversarial Training for Natural Language Understanding," arixv preprint arixv:1909.11764, 2019, pp: 1-9.

[25] Qiang F ,et al. Programming in C, Guangxi: Guangxi Normal University Press, 2021.

[26] Dong R S, Intrduction to Computer Science: Thinking and Methods(Third Edition), Beijing: Higher Education Press, 2015, pp. 1-335.

[27] Sedgewick R, Computer Science: An Interdisciplinary Approach, Gong X L, et al，translate. Being: China Machine Press, 2020, pp. 1-636.