

Research Article

Research on Multilevel Classification of High-Speed Railway Signal Equipment Fault Based on Text Mining

Fan Gao,¹ Fan Li,² Zhifei Wang,² Wenqi Ge ,³ and Xinqin Li²

¹Postgraduate Department, China Academy of Railway Science, Beijing 100081, China

²China Academy of Railway Sciences Corporation Limited, Beijing 100081, China

³School of Control and Mechanical Engineering, Tianjin Chengjian University, Tianjin 300384, China

Correspondence should be addressed to Wenqi Ge; gewenqi77@126.com

Received 11 June 2020; Revised 31 March 2021; Accepted 3 May 2021; Published 19 May 2021

Academic Editor: Jit S. Mandeep

Copyright © 2021 Fan Gao et al. 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.

In this paper, the multilevel classification model of high-speed railway signal equipment fault based on text mining technology is proposed for the data of high-speed railway signal fault. An improved feature representation method of TF-IDF is proposed to extract the feature of fault text data of signal equipment. In the multilevel classification model, the single-layer classification model was designed based on stacking integrated learning idea; the recurrent neural network BiGRU and BiLSTM were used as primary learners, and the weight combination calculation method was designed for secondary learners, and k -fold cross verification was used to train the stacking model. The multitask cooperative voting decision tree was designed to correct the membership relationship of classification results of each layer. Ten years of signal switch machine fault data of high-speed railway are used for experimental analysis; the experiment shows that the multilevel classification model can effectively improve the classification of signal equipment fault multilevel classification task evaluation index and can ensure the correctness of the subordinate relations' classification results.

1. Introduction

High-speed railway signal equipment is an important infrastructure to ensure the safety of high-speed railway operation [1]. With the accumulation of high-speed railway operation mileage, a large number of fault data of signal equipment are generated. Most of these fault data are stored in the form of unstructured text. This data contains important information on high-speed railway safety. For a long time, business personnel has been carrying out fault equipment for the data according to experience diagnosis and classification because there are many different types of signal equipment and equipment faults of different principles in high-speed railway, and the subordinate relationship between equipment and fault causes is strict; to carry out in-depth fault data analysis of high-speed railway, multilevel classification of fault data is required, while the manual multilevel classification work is easy to cause classification inaccuracy. In smart railway and railway under the

construction of big data, it is urgent to study the machine learning algorithm based on text mining to realize the multilevel classification of high-speed railway signal fault equipment.

Multilevel classification methods include top-down classification, global classification, and shrinkage classification [2, 3]. High-speed railway signal equipment fault multilevel classification applies the strategy of “divide and rule” in the top-down classification method, decomposes the large-scale fault equipment multilevel classification problem into a single-level classification problem, obtains the classification results of each level through the design of a single-level classification model, and finally collects and affiliates the classification results of all levels through the voting strategy of multitask collaborative decision tree correction to realize multilevel classification of signal fault equipment.

The single-layer classification model is a typical text classification model. Generally, the feature vector extraction methods such as BOW (Bag of Words) [4], TF-IDF (term

frequency-inverse document frequency) [5], TM (Topic Model), and Word2Vec [6] based on deep learning are used to quantify the text data. Based on the text feature vector, the machine learning model is used to learn and classify the text features. The text classification machine learning model includes a single classifier, integrated learning classifier, and deep learning model. The single classifier model includes DT (Decision Tree), SVM (Support Vector Machine), and NBC (Naive Bayes Classifier). The integrated classifier improves the classification performance by combining multiple single classifiers, mainly including bagging and boosting. The integrated classifier based on the idea of stacking [7] improves the classification and generalization performance by stacking different types of single classifiers. The two most common models of the deep learning classification model are RNN (Recurrent Neural Network) [8] and CNN (Convolutional Neural Network) [9], as well as BiGRU (Bidirectional Gated Recurrent Unit) [10] and BiLSTM (Bidirectional Long Short Term Memory) [11].

In this paper, the multilevel fault classification model of high-speed railway signal equipment is designed based on the research methods of feature extraction and single-layer classification model of railway safety text. Firstly, according to the characteristics of fault data of high-speed railway signal equipment, an improved feature extraction method based on TF-IDF is proposed. To avoid the problem of overfitting caused by a single classifier, the k -fold cross validation + stacking classification model is used to implement the single-layer classification model. In the stacking model, the weight allocation mechanism, which combines the whole and the category weight as the secondary learner, is proposed to improve the single-layer classification performance of the model. Based on the stacking model, tasks of different levels are classified, and a multitask cooperative voting decision tree is designed to correct the membership relationship of the classification results of different levels and improve the classification performance of the whole multilevel classification model. Finally, the experiment is carried out with the signal equipment fault data of high-speed railway from 2009 to 2018 to verify the validity and correctness of the multilevel classification model.

2. High-Speed Railway Signal Fault Text Feature Extraction

The fault data of high-speed railway signal equipment comes from the railway traction power supply management information system (EMIS). The fault data records the detailed information of the fault in a structured form. The reason information of the fault is stored in the form of natural language text. Based on the analysis of text data of high-speed railway fault causes, this paper classifies the fault causes and parts. The longest character length of fault record in the dataset is 100. A total of 2596 fault cases are selected, as shown in Table 1.

The text data of high-speed railway signal equipment fault cause analysis contains keywords with features such as a switch, red light band, and paster. TF-IDF is used to extract the features of fault text data [12]. The principle of the

TF-IDF method is that if a word appears more frequently in the sample and there are fewer samples with the word in the full-text file, it means that the word has a higher impact on the sample identification, with a good ability to distinguish. Because of the large number of fault documents of high-speed railway signal equipment, but each fault document is a short text, the TF-IDF method is used to extract features directly, which is easy to cause redundancy and sparseness of feature vector and lack of data specificity, so this paper improves the feature extraction method of high-speed railway signal equipment fault data based on the TF-IDF method.

The improved TF-IDF high-speed railway signal equipment fault text data feature extraction method is shown in Figure 1. First of all, the Chinese text content should be segmented. In this paper, the Jieba word segmentation tool based on the professional corpus and the common corpus is used to segment the signal fault text [13], and the auxiliary words such as “de,” “Le,” and other words that cannot represent the document features are cleaned up, and then, the TF-IDF is used to calculate the weight of the vocabulary set to form a vocabulary weight matrix, and the number of each vocabulary is counted to form a vocabulary dictionary. The TF-IDF weight matrix is the number of documents and the vocabulary of all documents, so it has a large dimension. The TF-IDF weight matrix has a serious sparsity. According to the TF-IDF value, the words in each sample are sorted, allowing words to repeat, selecting the first 100 words with the most sample features, reducing the dimension of the feature vector, replacing the word frequency with the corresponding word ID, forming the feature dictionary matrix, and inputting the text feature vector and the labels of each level after one hot coding into the text classification model.

3. Single-Layer Text Classification Model of Signal Equipment Fault Based on Stacking

The fault text feature dataset of high-speed railway signal equipment is divided into the training set, verification set, and test set, which are input into the stacking single-layer classification model. Based on the fault single-layer text classification of high-speed railway signaling equipment based on stacking, by using the cyclic neural network BiGRU and BiLSTM as the first learning device of stacking and using the prediction results of the two neural networks as features to train the combined weighted secondary learning device, the prediction results of the primary learning device are integrated by the secondary learning device. To avoid the overfitting problem caused by the model trained from the training set and to train multiple single-layer classification models and achieve the purpose of producing multiple prediction results from the same test set, the k -fold cross-validation method is adopted, as shown in Figure 2.

3.1. Principles of BiLSTM and BiGRU Primary Learners. With the deepening of neural network layers, the “gradient vanishing” phenomenon is more serious. In order to overcome gradient disappearance and realize deep learning

TABLE 1: Sample data of high-speed railway signal turnout fault.

Serial number	Communication signals	Cause analysis	Fault location	Cause of failure
1	Signal	Troubleshooting for J3 hidden danger of No. 6 turnout	Paste checker	Member adjustment
2	Signal	276#-1 switch safety contact is an open circuit, and it is recovered after adjustment	Switch machine	Other
3	Signal	13# switch J3 is in a reverse position, and ice is formed at the hook head and locking lever head (switch type: S700 K)	External locking and installation device	Natural disaster
4	Signal	147#-X1 reverse position checker is at fault, and it will be recovered after the following key points are replaced	Closely fitting checker	Contact group
5	Signal	The No.1 (1/42, S700 K) switch point 1 slide bed plate is jammed by ice	External locking and installation device	Natural disaster

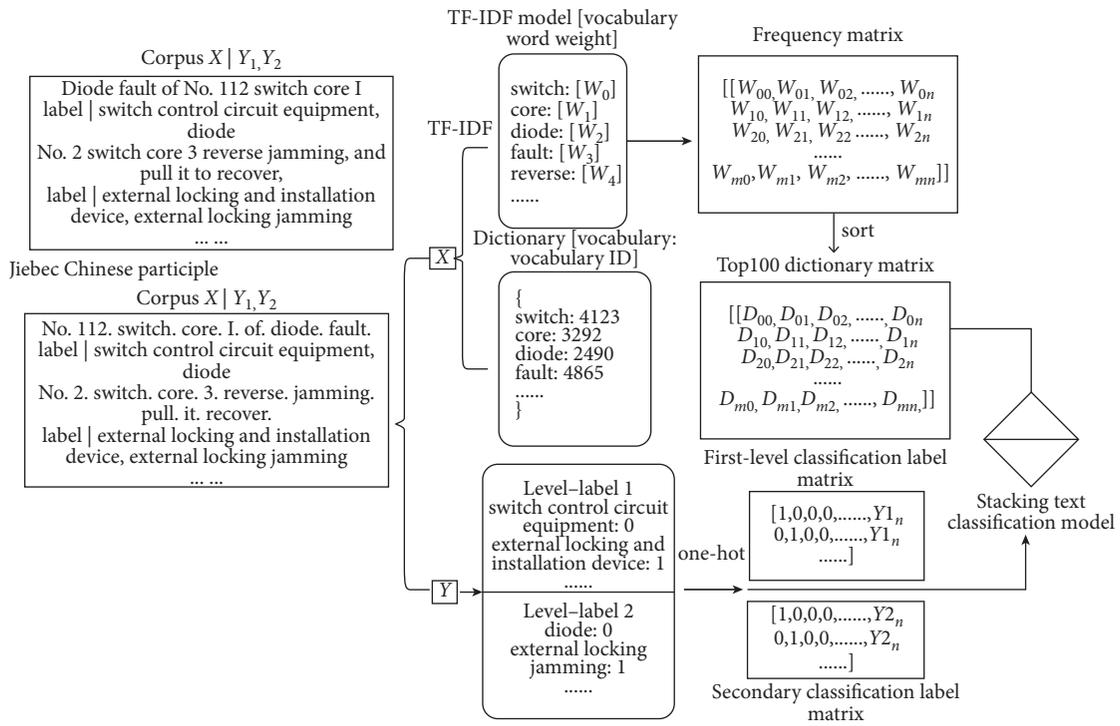


FIGURE 1: The feature extraction process of the high-speed railway signal fault text.

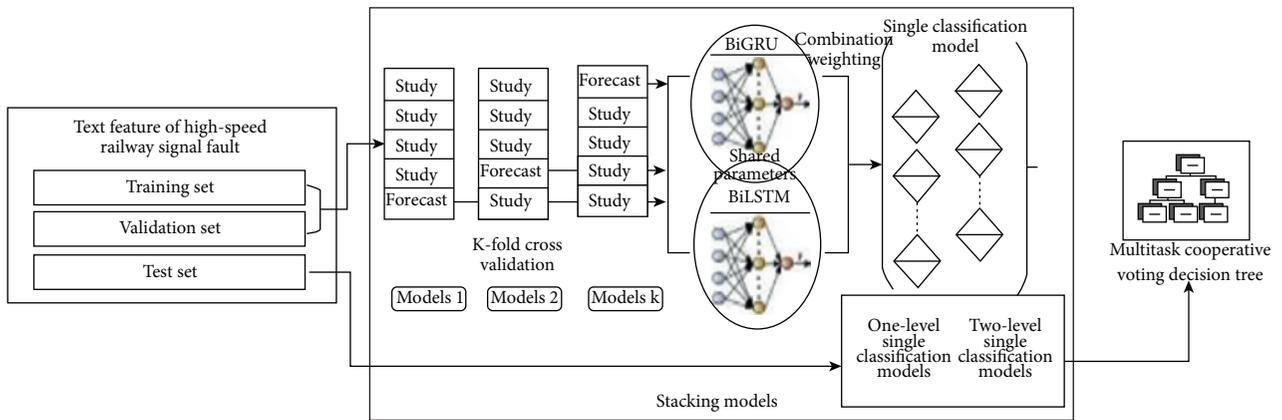


FIGURE 2: Fault text classification model of stacking signal equipment.

of the neural network, DNN (deep neural networks) is formed. The potential problem is the expansion of the number of parameters. So the convolutional neural network CNN is formed. CNN connects all neurons in the upper and lower layers by using “convolutional kernels” as intermediaries. However, it is impossible to model the changes in time series. To meet this demand, the recurrent neural network (RNN) appears. RNN is a kind of neural network for processing sequence information. Because of its pre-existing and postdependent structure, RNN has been widely used in natural language applications. The particularity of RNN lies in its output s_t at time t , which not only depends on the input layer x_t , but also depends on the output s_{t-1} of the previous node. Its learning process is a process of predicting the next word, such as x_{t-1} , x_t , and x_{t+1} , which is an input “switch positioning none.” Then, o_{t-1} and o_t correspond to “positioning” and “no” these two items, predicting what is the most likely next time. Through the training of signal fault corpus, o_{t+1} is most likely to be “representation.” h_t represents the state of the hidden layer at time t , x_t represents the input at time t , o_t represents the output at time t , and s_t represents the memory unit at time t and the linear parameter matrix of U , W models. As shown in Figure 3, its learning process is a process of predicting the next word.

$$\begin{aligned} \sigma(z) &= y = \frac{1}{1 + e^{-z}}, \\ \tanh(z) &= y = \frac{e^z - e^{-z}}{e^z + e^{-z}}, \\ s_t &= f(U * x_t + W * s_{t-1}), \\ o_t &= \text{softmax}(Vs_t), \\ h_t &= \left[\vec{h}_t, \overleftarrow{h}_t \right]. \end{aligned} \quad (1)$$

If the sequence is too long, RNN will lead to gradient dissipation. LSTM can solve this problem by learning to rely on information for a long time. LSTM is a variant of RNN. LSTM and GRU make information selectively affect the state of every moment in RNN by designing the gate structure. The structure of “gate” is an operation that uses sigmoid and a bit-by-bit multiplication. The combination of these two operations is a “gate” structure. Since the output of the sigmoid is a value between 0 and 1, it is helpful to select and forget the information. 0 means to discard all the information, and 1 means to retain all the information. Generally, the sigmoid function is selected as the activation function and $\tanh(z)$ function as the output function.

LSTM is divided into three gates, input gate, forgetting gate, and output gate, as shown in Figure 3. The LSTM relies on a number of “gates” that allow information to selectively influence the state of the RNN at each moment. The forgetting gate f_t decides to discard information, and input gate i_t consists of two layers. First, the sigmoid layer is used as the input layer to determine the value to be updated. Then, the $\tanh(z)$ layer generates a new vector \overline{C}_t into the cell state,

and C_t replaces the new input information with the information to be forgotten. Finally, the sigmoid layer determines which cell states need to be output in the output layer o_t :

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ \overline{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \\ C_t &= f_t * C_{t-1} + i * \overline{C}_t, \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ h_t &= o_t * \tanh(C_t). \end{aligned} \quad (2)$$

GRU is used to change the forgetting gate, input gate, and output gate of LSTM into update gate and reset gate and combine the unit state and output into one state, as shown in Figure 3:

$$\begin{aligned} z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]), \\ r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]), \\ \overline{h}_t &= \tanh(W \cdot [r_t * h_{t-1}, x_t]), \\ h_t &= (1 - z_t) * h_{t-1} + z_t * \overline{h}_t. \end{aligned} \quad (3)$$

In the classical recurrent neural network, the transmission of the state is one-way from front to back. In some problems, the output of the current time is not only related to the previous state but also related to the subsequent state. At this time, BiRNN (Bidirection Recurrent Neural Network) is needed to solve this kind of problem. The two-way RNN considers both the above information and the following information of the prediction word and retains the important information of the word from the front to the back and from the back to the front, which can more effectively predict. BiRNN is composed of two RNNs superimposed together. The output is determined by the states of the two RNNs. By replacing the RNN in BiRNN with LSTM or GRU structure, BiGRU and BiLSTM are formed.

3.2. Principle of Combined Weighted Secondary Learner. Combined weighted secondary learners not only consider the overall learning ability of neural networks but also consider the performance of the neural network in different categories. According to the learning results of a single neural network on the same input, the weight of a single neural network is assigned. The higher the accuracy is, the greater the weight of the neural network is. This method can effectively suppress the influence of few values and extreme values in the learning process of the neural network. The weight of the neural network in each category is calculated according to formulas (4) and (5). The weight of a neural network in a category is calculated by calculating the logarithm of an error proportion of the classification neural network in a category. The better the performance is, the greater the weight of the small error proportion is. When the error proportion exceeds 0.5, the weight is calculated as 0. Finally, according to formula (6), the weight of the whole

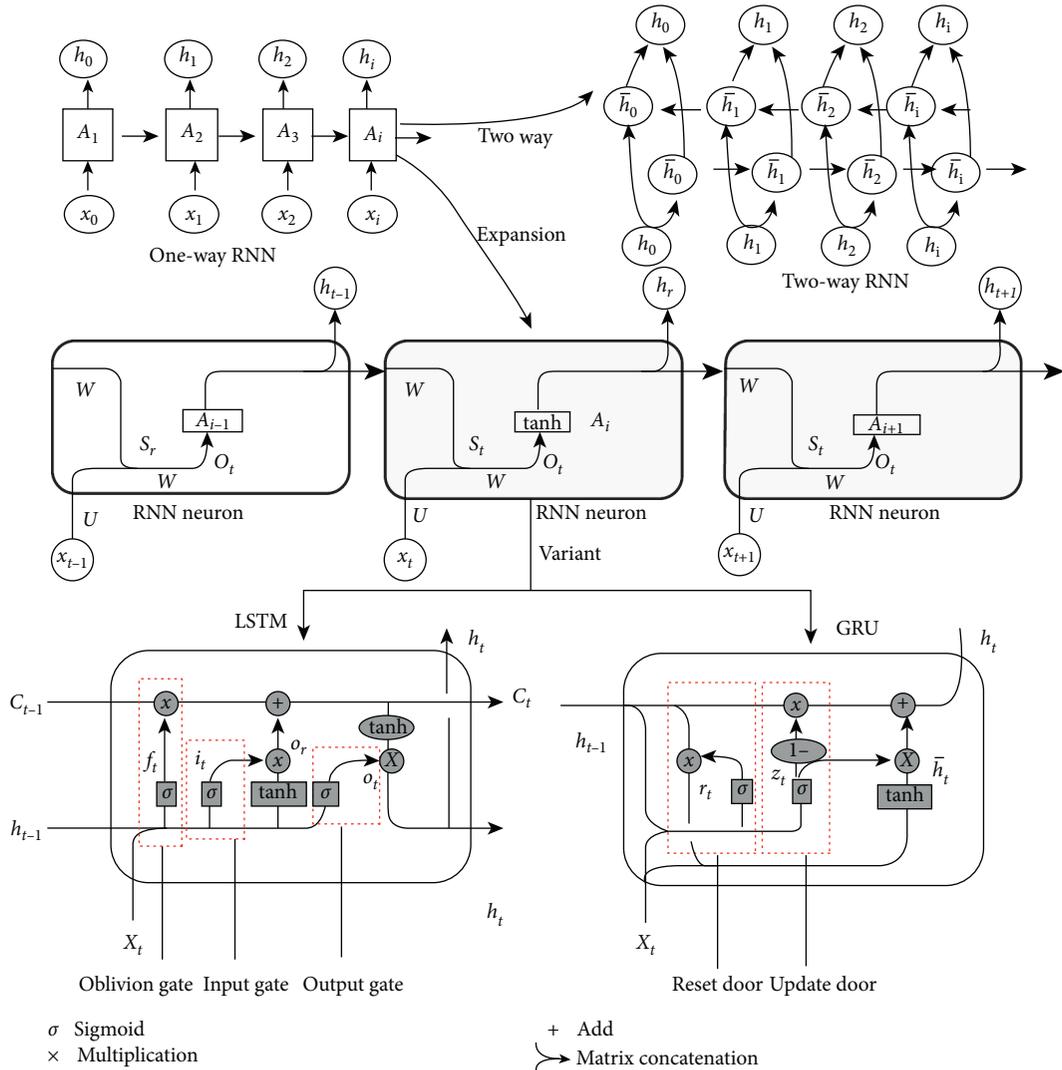


FIGURE 3: RNN and its variant neuron structural unit.

neural network and the weight of the category are added, and the prediction value of the model to the label is recalculated:

$$\varepsilon_{ij} = \frac{\text{errorNum}_{ij}}{\text{textNum}_i}, \quad (4)$$

$$\begin{cases} \alpha_{ij} = \ln\left(\frac{1 - \varepsilon_{ij}}{\varepsilon_{ij}}\right), & \varepsilon_{ij} < 0.5, \\ 0, & \varepsilon_{ij} \geq 0.5, \end{cases} \quad (5)$$

$$p_i = \sum_{j=1}^n (w_j + \alpha_{ij}) \cdot p_{ij}. \quad (6)$$

In the above formula, ε_{ij} represents the prediction error ratio of the j th neural network in the i th label, α_{ij} represents the weight of the j th neural network in the i th label, w_j represents the weight value of the j th neural network as a whole, and $\sum_{j=1}^n w_j = 1$.

3.3. *Stacking Single-Layer Classification Model.* In the stacking model, two neural networks are used as primary learners. In the data preprocessing layer, character feature vectors are reduced to dimension and then transformed into embedding, which is input into two bidirectional neural networks BiGRU and BiLSTM, respectively. After learning, two neural networks output the prediction probability of each classification label in the Softmax layer, respectively, and pass the combined weight classifier. The prediction results of the two primary learners are integrated and calculated, and finally, the classification of input data by the stacking model is output, as shown in Figure 4.

4. The Principle of Multitask Cooperative Voting Based on the Decision Tree

After k -fold cross learning, the stacking model generates k single-layer classification models, and the test data use k -fold cross learning to generate k prediction results. For different

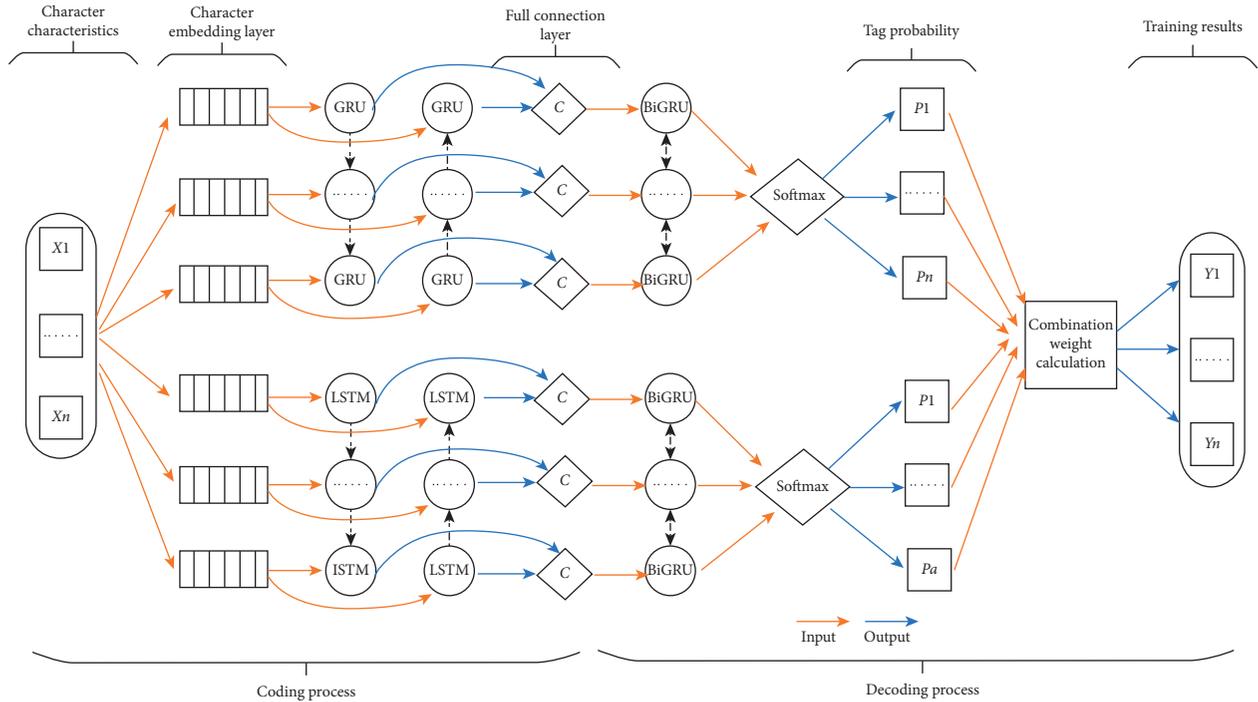


FIGURE 4: The network structure of the stacking single-layer classification model.

prediction results of the same data, the secondary and above prediction results have subordination. A multitask collaborative voting decision tree is designed to further improve the accuracy of the stacking model and the final prediction reliability.

A decision tree is mainly used for classification and regression tasks. Through ID3, C4.5, and C5.0 algorithm, train the decision tree and carry out appropriate pruning operation; then, it can effectively classify the unknown data. According to the subordination relationship between different levels of signal fault, this paper introduces the idea of multitask cooperative voting; after k -fold cross validation, the k -mode produces multiple prediction results for the same data to vote, and different levels of voting strategies are adopted. Represented by a decision tree, a multitask cooperative voting decision tree is formed. As shown in Figure 5, the multitask data is reclassified according to the voting decision tree.

This paper designs a three-level voting decision tree, which can be expanded according to the third-level voting principle. Set the level relation Dictionary of 3 layers as $D = \{D_1, D_2, D_3\}$. After k -fold cross validation, each prediction task has K prediction results at each layer, which are expressed as $r_l = \{r_{l1}, r_{l2}, \dots, r_{ln}\}$, n as the value of K , $v_l = \text{vote}(r_l)$ as voting for multitask r_l , v_l means voting results including $\{0: \text{valid}, 1: \text{abandoned}, 2: \text{all members abandoned}\}$, $v_{l'} = \text{vote}(r_{l'})$ is the residual vote, i.e., task r_l removes the members selected in the last voting and votes the result $v_{l'}$ again, and $\chi D(v_l)$ indicates the membership relationship between the voting result of v_l and set D . Because this relationship is clear, it will be described from the two values of 0 and 1; 0 indicates membership and 1 indicates nonmembership. $\delta(D)$ means to randomly select a

subset from set D as the prediction result, and the voting decision tree of three-tier multitasking is shown in Figure 5.

5. Experimental Verification and Result Analysis

In this paper, the 10-year data of switch equipment in high-speed railway signal basic equipment from 2009 to 2018 are used for verification, of which 70% are training set samples, 20% are verification set samples, and 10% are testing set samples. The data includes 7 categories of first-class labels and 64 categories of second-class labels. The comprehensive evaluation model of the $F1$ value is constructed by precision and recall.

The precision calculation formula is

$$\text{Precision} = \frac{1}{|C|} \sum_{i \in c} \frac{(TP_i + TN_i) \times TP_i}{TP_i + FP_i} \quad (7)$$

The recall calculation formula is

$$\text{Recall} = \frac{1}{|C|} \sum_{i \in c} \frac{(TP_i + TN_i) \times TP_i}{TP_i + FN_i} \quad (8)$$

The F -score calculation formula is

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (9)$$

where C is the total number of all samples, c is the total number of all categories, TP_i is the number of samples correctly classified into this category, TN_i is the number of samples correctly identified not in this category, FP_i is the number of samples mistakenly classified into this category,

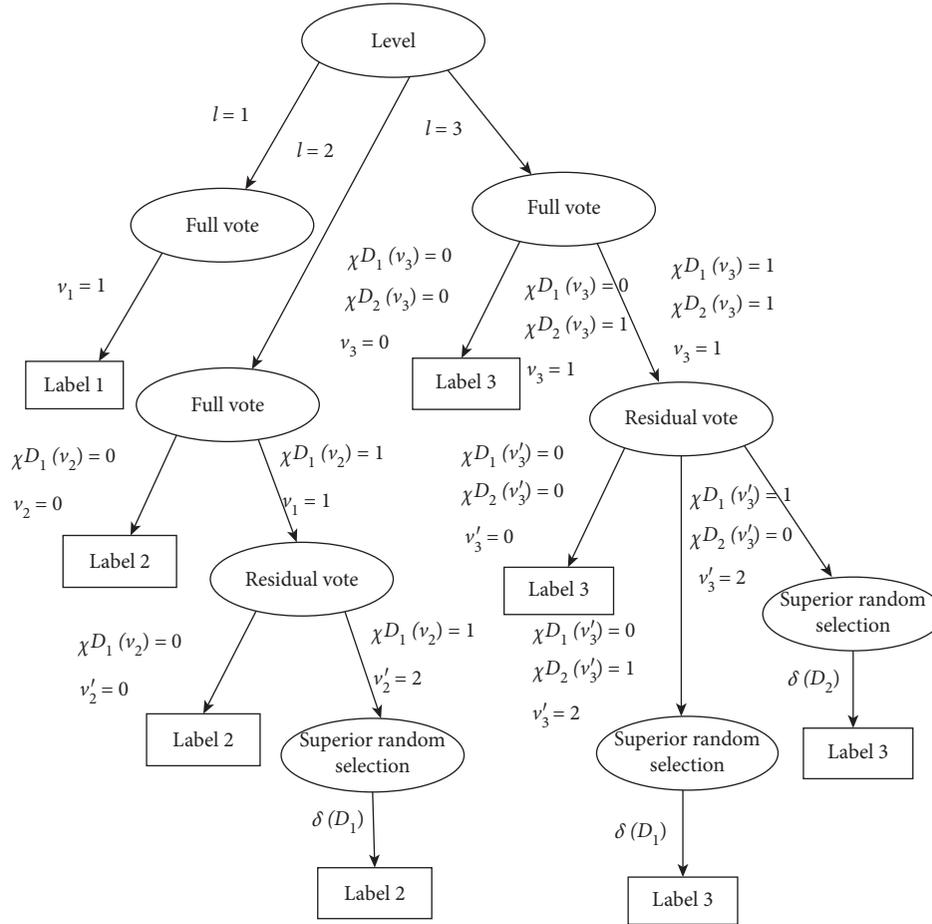


FIGURE 5: Three-level multitask cooperative voting decision tree.

and FN_i is the number of samples belonging to this category but mistakenly classified into other categories.

5.1. Experimental Analysis of Fault Text Single-Layer Classification of the Signal Switch Machine Based on the Stacking Model

5.1.1. Overall Weight Distribution of BiGRU and BiLSTM. The overall weight of BiGRU and BiLSTM is based on the learning results of the same eigenvector of a single neural network. In this paper, BiGRU and BiLSTM are designed to have the same network parameters. k -fold cross validation is set as $k=5$, the number of iteration rounds of the neural network is 50, the network input batch size is 256, the embedded layer dimension is 100, and the hidden layer dimension is 512. The loss function values for the BiGRU and BiLSTM primary and secondary training processes are shown in Figure 6, from which it can be seen that the number of iterative rounds is between 30–50, and the loss function loss value is close to the minimum and is largely stable. Compared with BiLSTM, BiGRU has a smaller loss function and better classification performance, and the second classification is less variable than the primary classification loss function. As the k value increases, the loss

value gradually decreased and tended to be stable during each iteration. The commonly used k values are 3, 5, 6, 10, etc. As shown in Figure 6, when $k=5$, the loss value tends to be consistent with $k=4$, so $k=5$ is chosen to avoid the occurrence of the overlearning and underlearning state.

After $k=5$ times of training, average the results of each training and get the training results of the two neural networks, as shown in Table 2.

It can be seen from Table 2 that, under the same parameters of the two neural networks, each evaluation index of BiGRU is higher than that of BiLSTM, and a higher overall weight can be obtained.

5.1.2. Weight Calculation of BiGRU and BiLSTM Categories. The number and weight of each category in the primary category of high-speed railway signal equipment fault are shown in Table 3. Due to the large number of secondary categories, considering the length of the article, only the analysis results of the category weight of the primary category are listed. From Table 3, it can be seen that the failure data of the paste checker, external locking and installation device, and switch machine are less. When the database is small and the number of errors is slightly larger, the category weight is small. On the contrary, when the number of

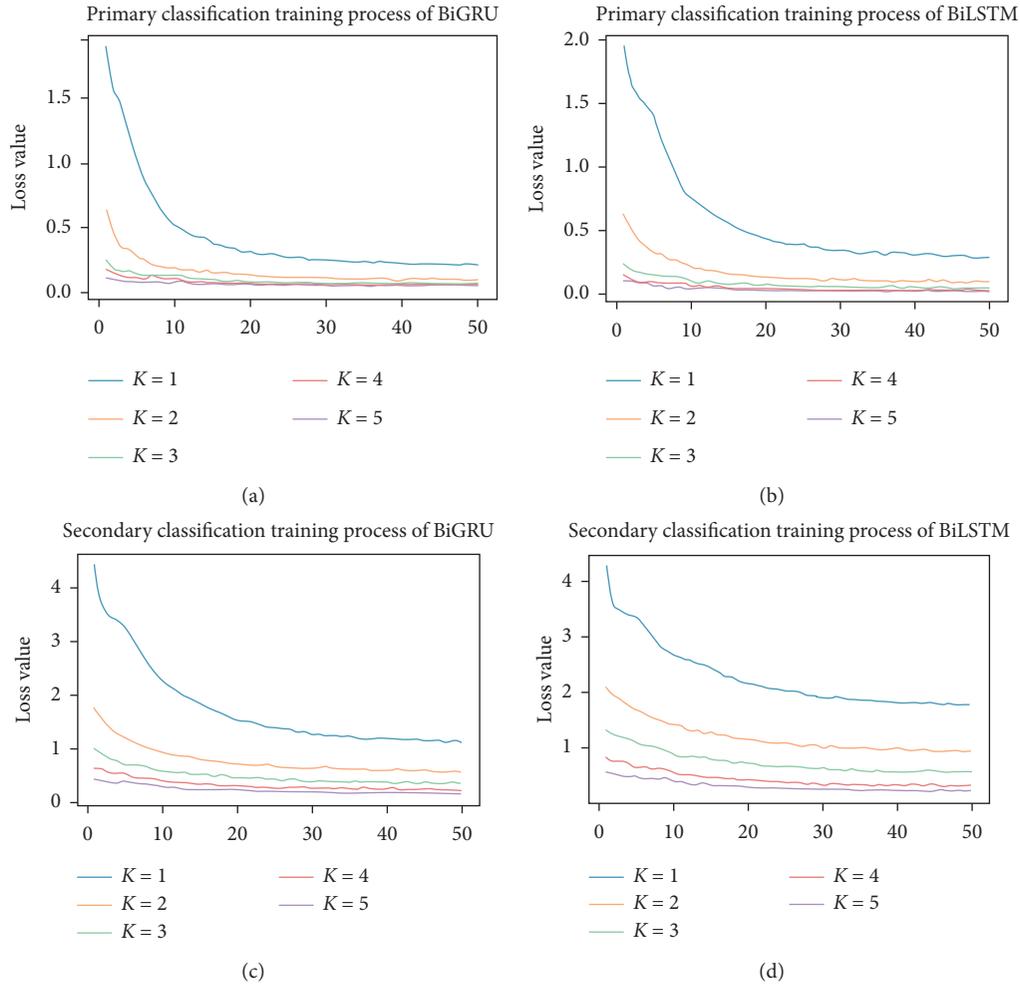
FIGURE 6: k -cross loss function of BiGRU and BiLSTM networks.

TABLE 2: Training results of BiGRU and BiLSTM neural networks.

Method	Level	Accuracy	Recall	F1 value
BiGRU	First-level classification	0.8745	0.8579	0.8661
	Second-level classification	0.7456	0.6809	0.7117
BiLSTM	First-level classification	0.8671	0.8368	0.8517
	Second-level classification	0.6713	0.6025	0.6350

supporting equipment and unexplained failures is large, the network learning effect is good and the category weight is also large.

Through the above experiments, the weight of each category of the neural network is obtained. Under different overall weight allocation, the evaluation indexes of the first-level fault classification and the second-level fault classification of the deep learning integration model are shown in Figure 7. It can be seen from the figure that when the weight of BiGRU is 0.7 and the weight of BiLSTM is 0.3, the evaluation index of the deep learning integration model is the highest.

5.1.3. Stacking Model Classification. By recomputing the output of the two networks with combined weighting, the

common classification prediction results are obtained. The final classification results are shown in Table 4. It can be seen that the value of each classification evaluation index has been improved. Experiments show that the stacking single-layer classification model is a model that can effectively improve the fault text classification index of high-speed railway signal equipment.

5.2. Experimental Analysis of the Voting Mechanism Based on the Decision Tree. After $K = 5$ cross validation, the stacking model has 5 classification models at each level, and 5 classification results are generated for the same test dataset. Based on the multitask collaborative voting decision tree designed in this paper, the voting results at each level are shown in Table 5. The first level has no affiliation

TABLE 3: Weight calculation results of the signal fault category of the first-level classification.

Category	Classified network	Number of classification errors/total number of categories	ϵ	Category weight
Switch machine	BiGRU	5/28	0.1786	1.5259
	BiLSTM	8/28	0.2857	0.9164
External locking and installation device	BiGRU	7/13	0.5386	0
	BiLSTM	8/13	0.6154	0
Paste checker	BiGRU	2/3	0.6667	0
	BiLSTM	1/3	0.3333	0.6933
Switch control circuit equipment	BiGRU	23/66	0.3485	0.6256
	BiLSTM	27/66	0.4091	0.3679
Maintenance equipment	BiGRU	24/100	0.2400	1.1527
	BiLSTM	31/100	0.3100	0.8001
Supporting equipment	BiGRU	193/1038	0.1859	1.4769
	BiLSTM	216/1038	0.2081	1.3364
Unknown cause	BiGRU	87/436	0.1995	1.3894
	BiLSTM	93/436	0.2133	1.3051

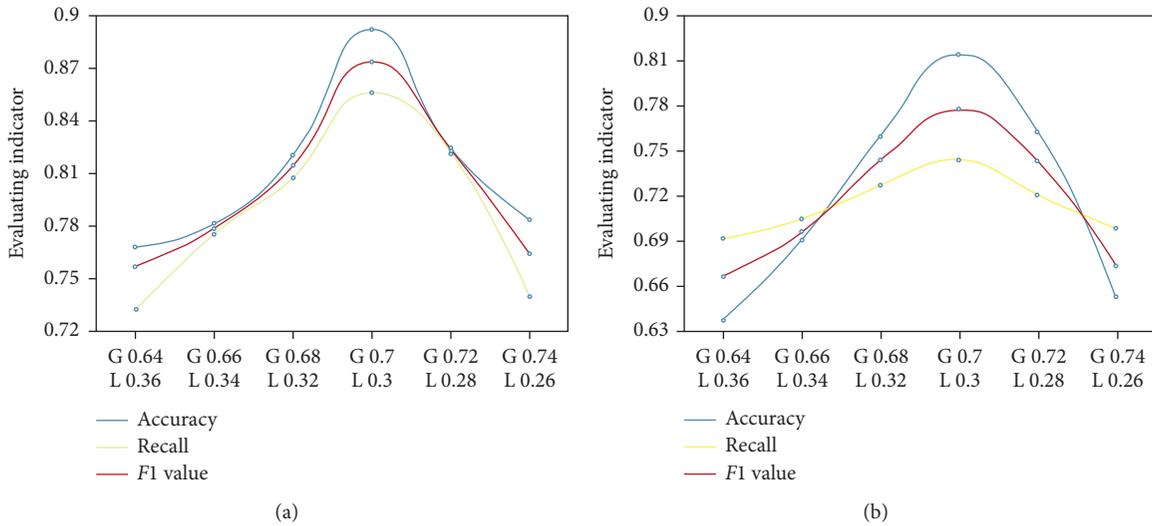


FIGURE 7: Evaluation indexes of the deep learning integration model under different weights. (a) First-level classification. (b) Second-level classification.

TABLE 4: Single-level text classification results of the stacking model.

Method	Level	Accuracy	Recall	F1 value
Stacking single-layer classification model	First-level classification	0.8814	0.8642	0.8727
	Second-level classification	0.7691	0.6747	0.7188

relationship, and its votes are all valid; the second-level effective votes account for the majority, and the invalid votes and the total votes are also large. This is also the main reason for the low accuracy of the secondary classification.

After re-calculating the classification indicators of voting results, the results are shown in Table 6. It can be seen from Table 6 that the classification indexes after multitask collaborative voting are improved compared with the classification indexes of the stacking model, especially the improvement range of secondary classification is more obvious.

In order to get a more comprehensive neural network in the classification of various categories, the fault sample data of signal switch equipment represented by the TF-IDF eigenvector is input into RF (Random Forest) and GBDT (Gradient Boost Decision Tree) for the experiment. RF represents the bagging algorithm. GBDT is the representative algorithm of boosting. 30% of real samples are used for evaluation, and the number of base classifiers is set to 50. The final experimental results are shown in Table 7. It can be seen from the table that the evaluation index of the deep learning integration model designed in this paper is obviously higher

TABLE 5: Voting results of multitask collaboration.

Level	Valid ticket	Invalid ticket	Ticket abandonment
First-level classification	345	0	0
Second-level classification	197	125	23

TABLE 6: Multilevel classification results of multitask cooperative voting.

Method	Level	Accuracy	Recall	F1 value
Stacking + multitask cooperative voting decision tree	First-level classification	0.8924	0.8561	0.8738
	Second-level classification	0.8152	0.7446	0.7783

TABLE 7: Experimental results of ensemble learning model classification.

Method	Level	Accuracy	Recall	F1 value
RF	First-level classification	0.8516	0.8419	0.8467
	Second-level classification	0.7526	0.7012	0.7568
GBDT	First-level classification	0.8598	0.8316	0.8455
	Second-level classification	0.7614	0.7278	0.7545
Stacking + multitask cooperative voting decision tree	First-level classification	0.8924	0.8561	0.8738
	Second-level classification	0.8152	0.7446	0.7783

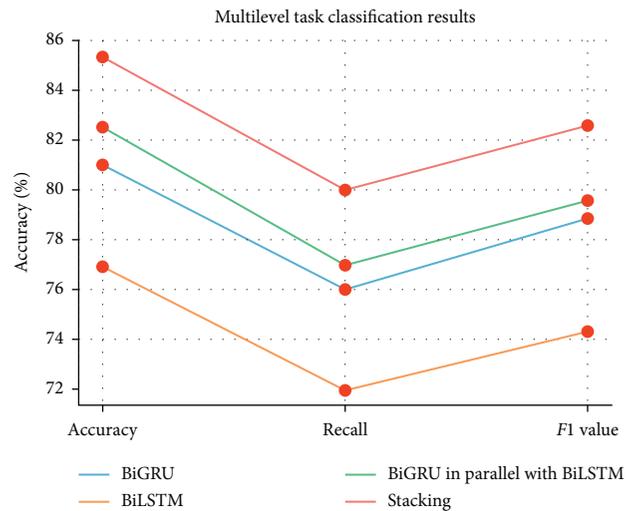


FIGURE 8: Comparison of classification performance of each model.

than that of the mature integration learning algorithms, RF and GBDT.

5.3. Implementation Summary. According to the above experimental analysis, the classification indexes of each model are calculated according to the average value of corresponding evaluation indexes at all levels. The performance is analyzed experimentally, regarding BiGRU model, BiLSTM model, stacking model, and BiGRU in parallel with BiLSTM model. The classification performance of each model is shown in Figure 8.

From Figure 8, it can be seen that the stacking model designed in this paper effectively improves the classification index of each level, solves the subordination of classification results based on the mechanism of multitask cooperative voting, and improves the overall classification performance

of the stacking model. Experiments show that the stacking model proposed in this paper is based on the stacking model of text mining technology has advantages in solving the multilevel classification problem of high-speed signal equipment.

6. Conclusion

Text data of high-speed railway equipment fault is important data for mining high-speed railway operation safety status and safety laws. It is necessary to realize multilevel classification of high-speed railway equipment fault data based on text mining technology. In this paper, a multilevel classification model is designed for the fault text data of high-speed railway signal equipment to solve the membership relationship between all levels of classification and effectively

improve the classification evaluation index. In this paper, the k -fold cross validation single-layer classification model based on the stacking idea ensures the algorithm difference and diversity of primary learners, effectively reduces the risk of classification overfitting, and improves the classification index compared with single neural network classifier, and the multitask cooperative voting mechanism ensures the membership relationship of classification results. The stacking single-level classification model and multilevel classification model in this paper have reference value in railway text classification.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (no. 2020YFF0304100), Key Project of China Railway Research Institute Group Limited (2019YJ115), Scientific Research Project of China Academy of Railway Sciences Corporation Limited (2052DZ1201), and National Science Foundation for Young Scientists of China (51707128).

References

- [1] Y. Lin, *High-Speed Railway Signal Technology*, China Railway Publishing House, Beijing, China, 2016.
- [2] L. He, Y. Jia, W.-H. Han, S. Tan, and Z.-K. Chen, "Research and development of large scale hierarchical classification problem," *Chinese Journal of Computers*, vol. 35, no. 10, pp. 2101–2115, 2012.
- [3] B. Li, "Expanding training dataset with class hierarchy in hierarchical text categorization," *Acta Scientiarum Naturalium Universitatis Pekinensis*, vol. 51, no. 02, pp. 357–366, 2015.
- [4] L. Wu, S. C. H. Hoi, and N. Yu, "Semantics-Preserving bag-of-words models and applications," *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1908–1920, 2010.
- [5] C. Zhai and J. Lafferty, "A study of smoothing methods for language models applied to information retrieval," *Acm Transactions on Information Systems*, vol. 22, no. 2, pp. 179–214, 2004.
- [6] P. D. Turney and P. Pantel, "From frequency to meaning: vector space models of semantics," *Journal of Artificial Intelligence Research*, vol. 37, no. 1, pp. 141–188, 2010.
- [7] X. Zhang, Z. Wang, and L. Liang, "A stacking algorithm for convolution neural network," *Computer Engineering*, vol. 44, no. 4, pp. 243–247, 2018.
- [8] C. Wang, F. Jiang, and H. Yang, "A hybrid framework for text modeling with convolutional RNN," in *Proceeding of the 23th ACM SIGKDD International Conference*, ACM, Halifax, Canada, August 2017.
- [9] L. E. I. Tao, R. Barzilay, and T. Jaakkola, "Molding CNNs for text: non-linear, non-consecutive convolutions," *Indiana University Mathematics Journal*, vol. 58, no. 3, pp. 1151–1186, 2015.
- [10] W. Wang, Y. Sun, Q. Qi, and W. Yiu, "Text sentiment classification model based on BiGRU-Attention neural network [J/OL]," *Application Research of Computers*, vol. 10, pp. 09–27, 2019.
- [11] M. Ren and G. Gan, "Sentiment analysis of text based on Bidirectional long short-term memory model," *Computer Engineering and Design*, vol. 39, no. 07, pp. 272–276, 2018.
- [12] X. Li, L. Shitianyun, P. Li et al., "Intelligent classification of railway safety hazards based on evolutionary ensemble classifier," *Journal of Transport Information and Safety*, vol. 37, no. 02, pp. 33–39, 2019.
- [13] L. Yang, P. Li, R. Xue et al., "Intelligent fault classification of railway signal equipment based on imbalanced text data mining," *Journal of the China Railway Society*, vol. 40, no. 2, pp. 59–66, 2018.