

Fault Classification for On-board Equipment of High-speed Railway Based on Attention Capsule Network

Lu-Jie Zhou¹ Jian-Wu Dang^{1,2} Zhen-Hai Zhang¹

¹School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

²Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphic & Image Processing, Lanzhou 730070, China

Abstract: The conventional troubleshooting methods for high-speed railway on-board equipment, with over-reliance on personnel experience, is characterized by one-sidedness and low efficiency. In the process of high-speed train operation, numerous text-based on-board logs are recorded by on-board computers. Machine learning methods can help technicians make a correct judgment of fault types using the on-board log reasonably. Therefore, a fault classification model of on-board equipment based on attention capsule networks is proposed. This paper presents an empirical exploration of the application of a capsule network with dynamic routing in fault classification. A capsule network can encode the internal spatial part-whole relationship between various entities to identify the fault types. As the importance of each word in the on-board log and the dependencies between them have a significant impact on fault classification, an attention mechanism is incorporated into the capsule network to distill important information. Considering the imbalanced distribution of normal data and fault data in the on-board log, the focal loss function is introduced into the model to adjust the imbalanced data. The experiments are conducted on the on-board log of a railway bureau and compared with other baseline models. The experimental results demonstrate that our model outperforms the compared baseline methods, proving the superiority and competitiveness of our model.

Keywords: On-board equipment, fault classification, capsule network, attention mechanism, focal loss.

Citation: L. J. Zhou, J. W. Dang, Z. H. Zhang. Fault classification for on-board equipment of high-speed railway based on attention capsule network. *International Journal of Automation and Computing*, vol.18, no.5, pp.814-825, 2021. <http://doi.org/10.1007/s11633-021-1291-2>

1 Introduction

Chinese train control system level 3 (CTCS-3) is widely used in the 300km/h high-speed railway. It is the key technical equipment for the Chinese railway to control electric multiple unit (EMU) trains, ensure traffic safety and improve transportation efficiency. On-board equipment is an important train operation control equipment in CTCS-3. On-board equipment has high reliability, but failures often occur due to uninterrupted operation in a complex and changeable environment for a long time. The large-scale equipment system is composed of various working modules, and each module is closely related. The failure of some modules will often produce a chain reaction, and in serious cases, it will lead to the failure of the whole production process^[1]. Timely and accurate fault location of on-board equipment is an important link to ensure train operation safety and equipment health maintenance. When the on-board equipment is

working, operation status information of each unit module is stored in the on-board safety computer in the form of a text log. After the end of the train operation, the status of each unit module is analyzed by downloading the on-board log. At present, the on-board equipment diagnosis is mainly through the technical staff to check the on-board log to identify the fault type. This way increases the labor cost and operation difficulty and has the possibility of misjudgment and omission.

For years, many scholars have researched intelligent fault classification and diagnosis, including Bayesian network^[2,3], support vector machine (SVM)^[4], backpropagation neural network^[5], etc., which have been applied in the fault classification of on-board equipment effectively. The quantities of data in the on-board log are great, and the relationship between the operation status of the equipment is complex. The probability of normal operation of on-board equipment is much greater than that of failure probability, so there is an imbalance between normal and fault samples. There are two problems in the existing research methods of the on-board equipment fault classification. First, the traditional feature extraction methods of the on-board log, such as the topic model^[2, 4] and vector space model (VSM)^[3, 5], ignore the relationship between contexts, and it is not easy to extract the deep structure and semantic features of the on-board log.

Research Article

Manuscript received July 11, 2020; accepted March 2, 2021; published online March 24, 2021

Recommended by Associate Editor Dong-Hua Zhou

Colored figures are available in the online version at <https://link.springer.com/journal/11633>

© Institute of Automation, Chinese Academy of Sciences and Springer-Verlag GmbH Germany, part of Springer Nature 2021

Second, most classifiers are based on the class balance hypothesis and aim to maximize the classification accuracy, which cannot scientifically evaluate imbalanced samples' classification effect.

The artificial neural network has been widely used in fault classification because of its good nonlinear fitting ability^[6]. With the development of deep learning technology, convolutional neural networks (CNN) based on deep learning have gradually become the research trend in classification tasks because of their ability to extract local deep features of samples^[7, 8]. CNN uses the convolution operation to extract the low-level features and pooling operation to retain significant features. However, the pooling operation can filter out the local position information and overall sequence structure of the text when modeling a text sequence such as the on-board log^[9]. Capsule network (CapsNet) was proposed by Sabour et al.^[10] to address the limitations of deep neural networks. The CapsNet uses vector-output capsules to replace the scalar-output feature extractors used in CNN and uses a dynamic routing mechanism to solve information loss caused by pooling operations. Yang et al.^[11] proposed a text classification model based on CapsNet. This work proves that the classification effect of CapsNet is better than that of CNN and long short-term memory (LSTM) models. However, CapsNet cannot selectively pay attention to the key contents of the text. Different words in the on-board log have different effects on the fault classification results. Effective feature extraction of key content helps make the network pay close attention to the key information in the training stage of the classification model. The attention mechanism^[12] can better solve this problem. In natural language processing, an attention mechanism can effectively improve the effectiveness of tasks such as generative dialog^[13] and target-based sentiment analysis^[14]. Kim et al.^[15] proposed a text classification model based on attention and CNN, but the low-efficiency of CNN coding limits this model.

This paper proposes a fault classification model for high-speed railway on-board equipment based on attention capsule networks to better distill the information from the on-board log and deal with class imbalance. The primary contributions of this study can be summarized as follows:

- 1) An attention mechanism of word embedding is incorporated into the network to capture the most important information in the on-board operation status statement.

- 2) The capsule network based on dynamic routing is used to learn the part and whole association information of the on-board log to improve the feature extraction ability and classification effect of the model.

- 3) In the presence of class imbalance, well-classified samples comprise the majority of the loss and dominate the gradient. Therefore, based on the cross-entropy loss function, a weighting factor and a dynamically modulat-

ing factor are introduced to construct a multi-class focal loss function to down-weight the loss assigned to well-classified samples.

To verify the correctness and effectiveness of the model, this work uses the on-board data provided by a railway bureau to compare this model with several other baseline models. The experimental results show that the model has a good effect on the fault classification of high-speed railway on-board equipment.

2 Background material

CTCS-3 is composed of on-board equipment and lineside equipment. The on-board equipment is connected with external equipment such as EMU and monitoring equipment through the external interface. The overall structure of CTCS-3 is shown in Fig. 1^[4, 16, 17]. The on-board equipment of CTCS-3 is designed with a distributed structure, and the functions of each module are relatively independent. Each module is connected by bus. The main control unit of on-board equipment mainly includes automatic train protection control unit (ATPCU) and CTCS-2 control unit (C2CU), which are the core computing control unit of CTCS-3 and CTCS-2, respectively. The driver machine interface (DMI) is used to realize the information exchange between driver and on-board equipment. The train interface unit (TIU) is used for the interface between on-board equipment and EMU. The radio transmission module (RTM) is used to connect the on-board radio and global system for mobile-railway (GSM-R) to realize the two-way transmission of information between on-board equipment and lineside equipment. The vital digital input/output (VDX) is the interface between the on-board equipment and the TIU, used for the input and output of relevant safety signals. The balise transmission module (BTM) is used to receive the balise information and feed it back to the main control unit. The track circuit receiver (TCR) is used to receive track circuit information. The speed and distance unit (SDU) is used to receive the pulse signal collected by speed sensors and radars and generate speed, distance, and direction information. The juridical recorder unit (JRU) is used to record the original information collected by on-board equipment and the control information output by on-board equipment during train operation^[4, 16, 17].

In the research of fault classification, the classification criterion is essential^[18]. To classify the types of on-board faults, this paper refers to the training materials for high-speed railway technicians^[16] and the relevant literature^[4, 17] on the fault research of on-board equipment and combines the work experience of on-site technicians. After the summary, it can be found that the modules of CTCS-3 on-board equipment with frequent faults are mainly concentrated in the seven parts: ATPCU, DMI, TIU, RTM, VDX, BTM, and SDU. When each unit module fails, it will produce specific fault types. Therefore, aiming at the modules with centralized fault occurrence, 20 typical fault types with high frequency are defined,

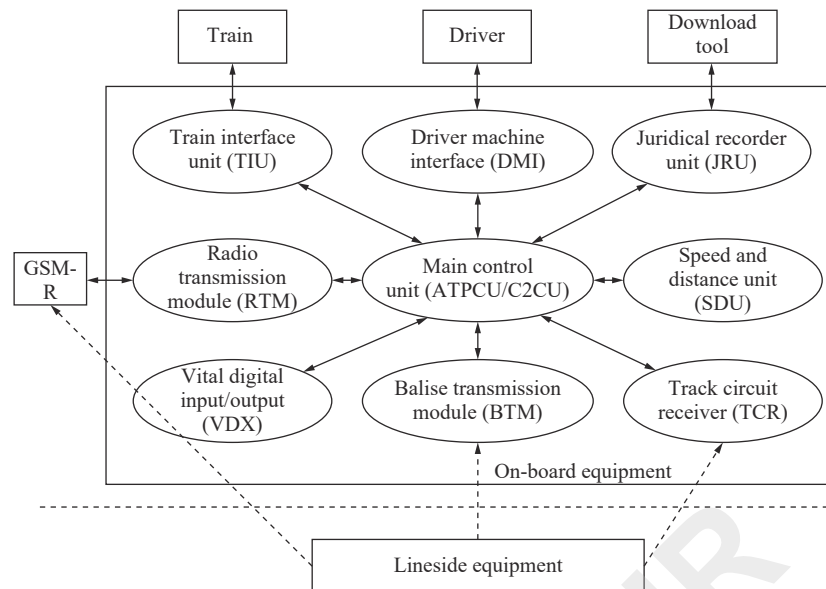


Fig. 1 Structure of CTCS-3

covering most faults. The fault modules, fault types, and some operation state statements of on-board equipment are shown in Table 1. It can be seen that the operation status statements are mainly described in a short text. The fault descriptions of the same fault type are diverse, and the same description will appear in different faults. The probability of normal operation of on-board equipment is much greater than failure, so the samples collected in the normal state (majority class) are many more than the fault samples (minority class). Therefore, it is necessary to establish a model suitable for imbalanced text classification to achieve the fault classification of high-speed railway on-board equipment.

3 Fault classification model of on-board equipment based on attention capsule network

To solve this problem effectively, an attention capsule network (ATT-Capsule) model for fault classification of high-speed railway on-board equipment is proposed, which is illustrated in Fig. 2. It consists of five parts: an embedding layer, an attention layer, a convolutional layer, a primary capsule layer, and a fully connected capsule layer. The embedding layer uses the word2vec^[19] method to convert the operation status statements of the on-board log into low-dimensional word embedding. The attention layer focuses on the important information by calculating the correlation score between words and creates a context vector for each word. The convolutional layer uses convolution filters to extract N-gram features from different positions of the text vectors to construct feature maps. The primary capsule layer combines the N-gram features extracted from the same location. Finally, the fully connected capsule layer is used to synthesize the characteristic information of the

primary capsule layer to generate the final fault type.

3.1 Embedding layer

The word2vec method is used to convert each word in the operation status statements into a low-dimensional real-value vector, capturing the syntactic and semantic information in the on-board log. After preprocessing, the on-board log is represented as the serialized data. The words in the sample are spliced sequentially to compose an input embedding matrix $X \in \mathbf{R}^{n \times d}$, where n is the length of the longest operation state statement in the sample set, and d is the dimension of the word embedding.

3.2 Attention layer

An attention mechanism is incorporated into the model to make the fault classification model focus on the important and distinguishable information to the classification results. The attention mechanism^[20, 21] of word embedding is mainly aimed at the text content. The idea is to calculate the correlation score between each word and other words in the text and create a context vector for each word. The context vector is concatenated with the word embedding as a new word representation fed to the convolutional layer. This method enables the network to focus on specific significant words in the text with a higher correlation score with other words, which contain more important distinguishing information.

Suppose $x_i \in \mathbf{R}^d$ is the d -dimensional word embedding of the i -th word in a sample and h_i is the context vector corresponding to x_i . Take turns to take each word as the target word and find its corresponding h_i . The h_i is combined in a weighted sum:

Table 1 Fault type of on-board equipment

Fault module	Number	Fault type	Operation state statements	Fault module	Number	Fault type	Operation state statements
BTM	F1	BTM port invalid	[BTMS] BTM1 status telegram invalid. StatusPort invalid in BTM1.	VDX	F11	VDX telegram invalid	BI-H A VDX1 telegram state = 4 (invalid). BI-H A telegram from VDX1 is not valid.
	F2	BSA startup error	Report failure inactive BTM1: Startup test strategy mismatch. BSA Permanent Error, inactive BTM1.		F12	VDX port invalid	BI-H VDX1:IN3 I/O failed.
	F3	BSA temporary error	[BTMS] BSA temporary error. BSA Temporary Error, active BTM1.	TIU	F13	Emergency brake relay (EBR) state wrong	BI-H EBR1 feedback timeout. VDX EBR1 port switched to invalid.
	F4	BSA permanent error	[BTMS] BSA permanent error. BSA Permanent Error, inactive BTM1.		F14	Brake feedback relay (BEB) state wrong	Wrong feedback. Timeout expires 66523. Time 64 623 BI-H EBFR state wrong.
	F5	BTM test timeout	[BTMS] startup test timeout. BSA TestInProgress, active BTM1.		F15	Bypass relay (BP) state wrong	Bypass failed. VDX bypass port switched to invalid.
	F6	All zero balise message	[BGH] Expected balise not found. IL A Detect balise reported.		F16	Cab activation (CabAct) relay state wrong	Direction control failure. Invalid direction signal combination received.
ATPCU	F7	Kernel mode transition invalid	(MS) A-kernel mode transition invalid.	SDU	F17	Radar error	Speed sensor failure 1.
	F8	MA A/B code inconsistent	VC: end of MA! a=1 145 772 832, b=1145582832. VC: start of MA! a=1 143 838 732, b=1143676832.		F18	Tacho error	Tacho Error 1.
	F9	Level transition A/B code inconsistent	VC: etcs level! a=3, b=5.	RTM	F19	Radio timeout	Level changed to LSTM, NID=45, orderby=2. [RS] NVCONTACT time_out reaction SB.
	F10	RBC handover A/B code inconsistent	VC: RBCHandover! a=1, b=0.	DMI	F20	DMI hardware failure	IO reported stopping failure. B-code: MMI down in active cabin.

*Abbreviations: BTMS balise transmission module supervisor; BTM1: Balise transmission module 1; BSA: Balise service available; BGH: Balise group handover; MS: Maintenance service; VC: Vital compare; MA: Movement authority; etcs: European train control system; RBC: Radio block center; BI-H: Brake interface handler; VDX1: Vital digital input/output 1; IN3: Input port 3; I/O: Input/output; EBR1: Emergency brake relay 1; EBFR: Emergency brake feedback relay; LSTM: Level specific transmission module; NID: Identification number; RS: Radio signal; SB: Stand-by mode; MMI: Man machine interface.

$$h_i = \sum_{j=1, j \neq i}^n \alpha_{i,j} \times x_j \tag{1}$$

where $\alpha_{i,j}$ are called attention weights, and $\alpha_{i,j} \geq 0$, $\sum_{j=1}^n \alpha_{i,j} = 1$. Using softmax normalization to realize the allocation of attention weights.

$$\alpha_{i,j} = \frac{\exp(\text{score}(x_i, x_j))}{\sum_{j'=1}^n \exp(\text{score}(x_i, x_{j'}))} \tag{2}$$

where the score function in (2) is used to calculate the correlation score between two words, which can be calculated by training a feedforward neural network.

$$\text{score}(x_i, x_j) = v_a^T \tanh(W_a [x_i \oplus x_j]) \tag{3}$$

where v_a and W_a are the weights to be learned in network training. The higher the correlation scores, the greater the attention weights.

The context vector h_i is concatenated with the word embedding x_i as the extended vector x_i' :

$$x_i' = h_i \oplus x_i \tag{4}$$

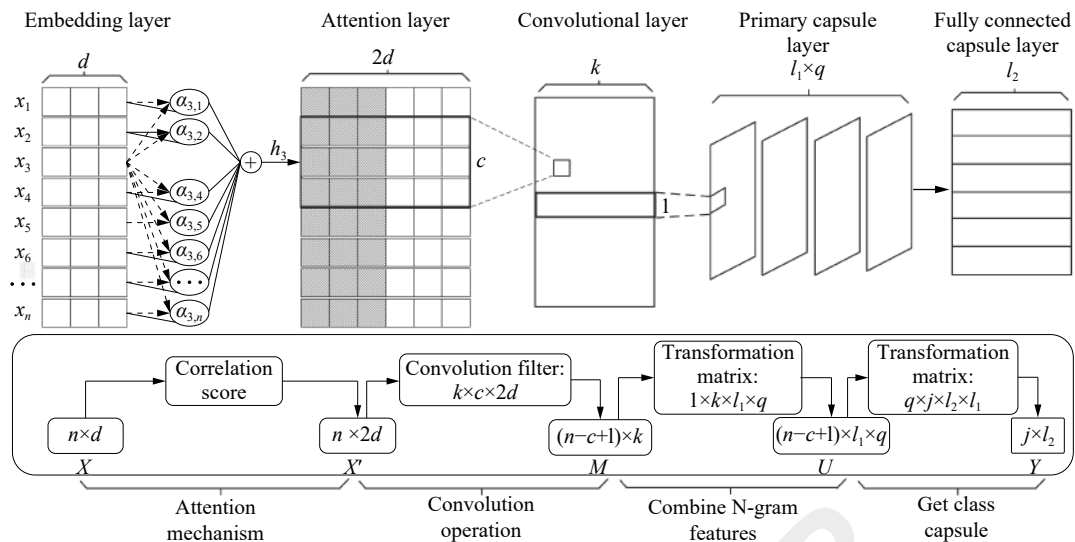


Fig. 2 Structure of fault classification model for on-board equipment

where the extended vector $x_i' \in \mathbf{R}^{2d}$. A new text matrix $X' \in \mathbf{R}^{n \times 2d}$ is constructed by stitching together x_i' , which will be fed to the convolutional layer.

3.3 Convolutional layer

This layer is a standard convolutional layer that extracts the N-gram features of the input text matrix at different positions through the convolution operation. The convolutional layer is connected to a local area of the upper layer by a convolution filter. The locally weighted sum is passed to the non-linear activation function, and the final output value of the convolutional layer is produced.

Suppose that there are k convolution filters with the stride of 1 in the convolutional layer. $w_i \in \mathbf{R}^{c \times 2d}$ represents the i -th filter for the convolution operation, where c is the window size of the filter used to identify the N-gram local feature and $2d$ is the dimension of the input text matrix. Each filter performs a convolution operation on the sliding over the text matrix from top to bottom. The feature map m_i generated by the i -th filter is

$$m_i = f(w_i \cdot l_{i:i+c-1} + b_i) \in \mathbf{R}^{n-c+1} \quad (5)$$

where $l_{i:i+c-1}$ represents a continuous c word embedding, b_i is bias, f is a non-linear activate function rectified linear unit (ReLU). When there are k filters, k feature maps can be obtained, which are defined as

$$M = [m_1, m_2, \dots, m_k] \in \mathbf{R}^{(n-c+1) \times k}. \quad (6)$$

3.4 Primary capsule layer

The primary capsule layer is the first capsule layer in the network, which uses vector-valued capsules instead of the scalar-valued feature extractors of a convolutional

neural network to combine the N-gram features extracted from the same location. The primary capsule layer can extract different attributes of a certain feature in the text, such as the location information of the word, the syntactic and semantic information of the text.

The primary capsule layer is a combination of different attributes of the vector-matrix $M_i' (i = 1, 2, \dots, n - c + 1)$ with N-gram length 1 in the convolutional layer, and M_i' is the i -th row vector of M . Suppose that the dimension of the primary capsule is l_1 , the i -th primary capsule filter is $z_i \in \mathbf{R}^{1 \times k}$. Each filter is convoluted with M_i' in step of 1, then the feature map p_i of each filter can be generated:

$$p_i = g(z_i \cdot M_i' + e_i) \in \mathbf{R}^{n-c+1} \quad (7)$$

where e_i is the bias term, g is a non-linear activate function. Since each capsule includes l_1 filters, the output vector of each capsule is $u_i \in \mathbf{R}^{(n-c+1) \times l_1}$. For $i \in \{1, 2, \dots, q\}$, the output of the primary capsule layer can be obtained, which are defined as

$$U = [u_1, u_2, \dots, u_q] \in \mathbf{R}^{(n-c+1) \times l_1 \times q}. \quad (8)$$

3.5 Fully connected capsule layer

The last layer of the network is the fully connected capsule layer used to get the class capsule:

$$Y = [y_1, y_2, \dots, y_j] \in \mathbf{R}^{j \times l_2} \quad (9)$$

where $y_j \in \mathbf{R}^{l_2}$ represents the j -th class capsule. The capsule matrix U obtained from the primary capsule layer is linearly transformed to obtain the prediction vector $u_{j|q}$, and the final class capsule Y is produced by the dynamic routing algorithm. The structure of the fully connected capsule layer is shown in Fig. 3. The output of

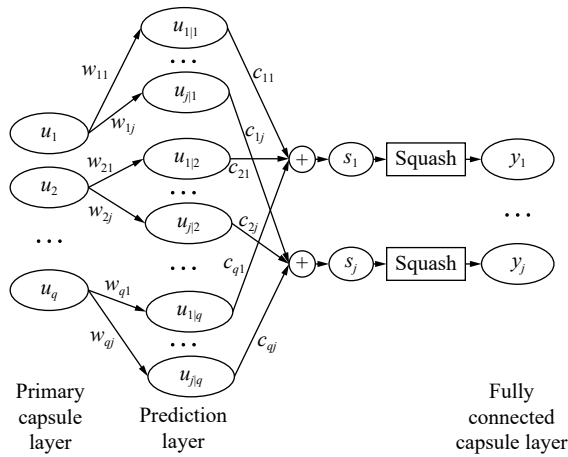


Fig. 3 Structure of the fully connected capsule layer

the class capsule is a vector, and the norm of the capsule vector represents the probability for each type.

3.6 Dynamic routing

The structural relationship between the primary capsule layer and the fully connected capsule layer is shown in Fig. 3. The calculation process includes two stages: transformation matrix and dynamic routing. First, the prediction vector is obtained by transforming the matrix of each capsule in the primary capsule layer:

$$u_{j|q} = u_q \times w_{qj} \tag{10}$$

where u_q is the output of the primary capsule and w_{qj} is a transformation matrix. Then, the prediction vector can be calculated:

$$S_j = \sum_q c_{qj} \times u_{j|q} \tag{11}$$

where c_{qj} is the coupling coefficient, which can be determined by the iterative dynamic routing process. The coupling coefficient represents the connection weight between each lower capsule layer and the corresponding upper capsule layer. For each capsule q , the sum of all weights c_{qj} is 1. According to the method of Sabour et al.^[10], S_j is compressed and redistributed by the squash function, and the norm of S_j is transformed to 0–1.

$$y_j = \frac{\|S_j\|^2}{1 + \|S_j\|^2} \times \frac{S_j}{\|S_j\|} \tag{12}$$

where y_j is the output vector of the j -th capsule in the fully connected capsule layer. The first half of (12) is a nonlinear squashing function, and its main function is to constrain the length of y_j . The second half of (12) is to unify the S_j and make its direction consistent with y_j . So the squashing function only changes the length of S_j and does not change the direction of S_j .

The dynamic routing algorithm learns the nonlinear mapping relationship between the prediction layer and the full connection layer in an iterative way. It depends on the softmax function to update the coupling coefficient c_{qj} constantly.

$$c_{qj} = \frac{\exp(b_{qj})}{\sum_k \exp(b_{qk})} \tag{13}$$

$$b_{qj} \leftarrow b_{qj} + u_{j|q} \times v_j \tag{14}$$

where b_{qj} represents the prior probability that capsule q couples to capsule j , and its initial value is 0. The similarity between the vectors is judged by the inner product of the prediction vector $u_{j|q}$ of the primary capsule and the output vector v_j of the full connection layer capsule. Then update b_{qj} iteratively and update the coupling coefficient c_{qj} accordingly.

The process of dynamic routing is summarized in Algorithm 1.

Algorithm 1. Dynamic routing

Input: Prediction vectors $u_{j|q}$, routing iteration times T .

Output: Class capsule vectors y_j

- 1) **for** all capsule q in lower-level and capsule j in higher-level: $b_{qj} \leftarrow 0$
- 2) **for** T iterations **do**
- 3) **for** all capsule q in lower-level and capsule j in higher-level:
- 4) $c_{qj} \leftarrow \text{softmax}(b_{qj})$
- 5) **for** all capsule j in higher-level capsule:
- 6) $S_j \leftarrow \sum_q c_{qj} \times u_{j|q}$
- 7) $y_j \leftarrow \text{squash}(S_j)$
- 8) $b_{qj} \leftarrow b_{qj} + u_{j|q} \times v_j$
- 9) **return** y_j

3.7 Loss function

As for the loss function, the focal loss is applied to the ATT-Capsule model. Focal loss is proposed by Lin et al.^[22] as a binary classification problem for dense object detection initially, which addresses the few-shot problem by reshaping the standard cross-entropy loss, it down-weights the loss assigned to well-classified examples. In this paper, a loss function is constructed by referring to the focal loss function to solve imbalanced text multi-class classification.

The standard cross-entropy loss function is shown in (15):

$$f_{CE} = -\frac{1}{D} \sum_{i=1}^D \sum_{j=1}^C \hat{p}_{ij} \log p_{ij} \tag{15}$$

where D is the number of training samples, C is the

number of target classes. \hat{p} is a variable. If the prediction class is the same as the actual class, it is equal to 1. Otherwise, it is equal to 0. p is the probability of the prediction class. Cross entropy loss function treats all samples equally. To control the contribution of each sample to the loss, a weight factor α is introduced to weaken the influence of majority class samples on the loss. The α -balanced CE loss can be written as

$$f_{\text{BCE}} = -\frac{1}{D} \sum_{i=1}^D \sum_{j=1}^C \alpha_j \hat{p}_{ij} \log p_{ij}. \quad (16)$$

Equation (16) is to balance the difference between the number of samples. To further differentiate between easy/hard samples, a dynamically modulating factor $(1-\hat{p}_{ij})^\gamma$ is introduced based on (16), where γ is a tunable focal parameter. By reshaping the loss function to down-weight easy examples and thus focus training on hard negatives. The multi-class focal loss function can be written as

$$f_{\text{FL}} = -\frac{1}{D} \sum_{i=1}^D \sum_{j=1}^C \alpha_j (1-\hat{p}_{ij})^\gamma \hat{p}_{ij} \log p_{ij}. \quad (17)$$

In the case of multi-class classification, a α_j ($j = 1, 2, \dots, C$) is set for each class and α_j is used to control the weights of different classes. In model training, to solve the problem of vanishing gradient and improve the convergence rate, batch normalization^[23] is added after the convolution operation of the model, and then the activation function is used for the operation. The ATT-Capsule model uses the adaptive moment estimation (Adam) optimization method to minimize the multi-class focal loss. The hyperparameters of multi-class focal loss are determined by experiments.

4 Experiments

4.1 Dataset and evaluation metrics

To verify the effectiveness of the proposed ATT-Capsule model for fault classification of high-speed railway on-board equipment, this paper takes the on-board equipment of CTCS-3 as the research object, and carries out experiments on 20 kinds of on-board equipment faults listed in Table 1 through the proposed ATT-Capsule model. The experimental data are taken from the on-board log provided by the electricity department of a railway bureau. The dataset consists of 3 152 samples with 21 classes, among which the fault numbers F1 to F20 are classified as classes 1 to 20, and the normal operation class N is classified as class 21. The ratio in which the data are divided among training, verification, and testing is 6:2:2 so that model can be trained and verified on sufficient data

and at the same time has enough data to test the effect of the model on the classification of on-board equipment fault.

The fault classification of on-board equipment is an imbalanced multi-class classification problem, and the accuracy is not enough to fully evaluate the fault classification performance of the model. Because even if the minority class fault samples are misclassified, the overall fault classification accuracy of the classifier is still very high. In order to scientifically evaluate the fault classification effect of the proposed model, precision (*Macro-P*), recall (*Macro-R*), and F1-Measure (*Macro-F₁*) are used as the evaluation metrics, which can be computed by

$$\text{Macro-P} = \frac{1}{K} \sum_{i=1}^K P_i \quad (18)$$

$$\text{Macro-R} = \frac{1}{K} \sum_{i=1}^K R_i \quad (19)$$

where P_i and R_i represent the precision and recall of i . F1-Measure is a combination of recall and precision and helps understand the results much better than the other metrics shown in (20). The following formula gives

$$\text{Macro-F}_1 = \frac{1}{K} \sum_{i=1}^K F_i = \frac{1}{K} \sum_{i=1}^K \frac{2 \times P_i \times R_i}{P_i + R_i}. \quad (20)$$

4.2 Experiment settings

First, the word2vec method is used to convert each word in the on-board equipment operation status statements into word embedding, and in the experiments, the dimension of the word embedding is set to 300. In the ATT-Capsule on-board equipment fault classification model, three kinds of filter windows are set for the convolutional layer to extract the low-level local features of different operation status statement lengths. The number of capsules in the primary capsule layer is set to 10, and the dimension is 12. An Adam optimizer with a learning rate of 1×10^{-3} is used.

To evaluate the performance of the ATT-Capsule model in fault classification of high-speed railway on-board equipment, this paper will evaluate the fault classification effect of the model from three aspects: 1) Discuss the influence of model parameters on on-board equipment fault classification. 2) Compare our proposed model with several strong baselines to evaluate the effectiveness of our model in fault classification. 3) Verify the effect of introducing an attention mechanism into the capsule network on the fault classification of on-board equipment.

In order to verify the fault classification effect of the proposed model, several representative classification models are selected as baseline models for fault classification

on the same on-board equipment fault data set, including statistical machine learning method, LSTM, and its bidirectional variant, CNN and its variation methods, and capsule-based models.

Support vector machine (SVM)^[4]: SVM uses a kernel function to map data points in low-dimensional space to high-dimensional space to realize the classification of non-linear separable sample data.

Random forest (RF)^[24]: RF is an ensemble classifier with several decision trees. The predicted class for a sample is computed by aggregating the predictions of decision trees through majority voting.

LSTM^[25]: LSTM has memory ability and is suitable for dealing with sequence data. It can obtain sentence features with long-distance dependency between words.

Bi-directional LSTM (BiLSTM)^[26]: BiLSTM uses forward and backward LSTM to capture the hidden information, which constitutes the final output.

TextCNN^[27]: TextCNN is a feedforward neural network with convolution operation.

Dynamic CNN (DCNN)^[28]: DCNN extracts sentence features by wide convolution and dynamic K-max pooling.

CapsNet^[10]: This is a basic capsule network, which consists of a convolutional layer, a primary capsule layer, and a fully connected capsule layer.

Gated recurrent unit (GRU)-CapsNet^[29]: This network uses the GRU layer to learn latent representations of input word embedding. The subsequent capsule network layer learns high-level features from that hidden representation and outputs the prediction class.

4.3 Parameters sensitive analysis

In order to explore the influence of model parameters on fault classification effects of on-board equipment in high-speed railways, three essential parameters are investigated, which are the filter window size c and the number of filter windows k in the N-gram convolutional layer define in (5) and (6), the routing iteration times T , and the weight factor α_j and γ define in (17).

First, taking the fault data set of on-board equipment as the input, the routing iteration times in the fault classification model is set to 4, and the network is trained by the standard cross-entropy loss function. The influence of convolution filter parameters on the fault classification of on-board equipment is tested by changing the size and number of filter windows. As can be seen from Table 2, compared with the single-size filter window, using a multi-size filter window to extract the features of on-board operation state statements can improve the adaptability of the model to the change of state statement length, so as to improve the precision and recall of fault classification. When the filter window size of the fault classification model is the same, compared with the 200 filter windows, the F1-Measure of on-board equipment

Table 2 Fault classification effect under convolution filter parameters

Filter window size	Filter window number	Marco-P	Marco-R	Marco-F ₁
3	200	0.856 7	0.760 1	0.784 6
3	300	0.873 8	0.775 1	0.798 1
3,4	300	0.866 9	0.790 1	0.810 4
3,4,5	200	0.881 4	0.792 2	0.822 3
3,4,5	300	0.887 2	0.796 9	0.825 3

fault classification is improved under the number of filter windows is 300. The results show that the multi-size filter window can extract the low-level local features of different on-board operation state statement lengths more comprehensively. And appropriately increasing the number of filter windows can also improve the fault classification effect.

For the multi-class focal loss function, its main idea is to use an appropriate function to measure the contribution between easy/hard samples and minority /majority class samples, so the values of α_j and γ will affect the model, resulting in the difference of the judgment results of on-board fault type. Buda et al.^[30] proposed that most of the real-world cases can be divided into two types: step imbalance and linear imbalance. In step imbalance, the number of samples is equal within minority classes and equal within majority classes but differs between the majority and minority classes. In the task of fault classification of on-board equipment, the number of normal classes is much higher than that of other fault classes. Therefore, this task belongs to a step imbalance. There are 21 classes in the task of on-board fault classification, among which the fault numbers F1 to F20 are classified as classes 1 to 20, and the normal operation class N is classified as class 21. In the multi-class focal loss function of (17), the weight of different classes needs to be controlled by $\alpha_j (j = 1, 2, \dots, C)$, where C is the number of target classes. Since this task is step imbalance, α_j can be expressed as

$$\alpha_j = \begin{cases} \sigma_1, & \text{if } j = 1, 2, \dots, 20 \\ \sigma_2, & \text{if } j = 21 \end{cases} \quad (21)$$

The first step is to test the effect of α_j on fault classification. The value of γ is set to 0 and fix σ_1 is set to 1. It is only necessary to adjust σ_2 to weakening the influence of on-board equipment normal operation samples on the loss. The value of σ_2 is searched on the range {0.2, 0.4, 0.6, 0.8, 1}. As shown in Table 3, the model proposed has good fault classification performance when $\sigma_2 = 0.8$. Then, the σ_2 is set to 0.8 and σ_1 remains unchanged to investigate the effect of γ on fault classification. Referring to the research of Lin et al.^[22], the value of γ is searched on the range of {0.5, 1, 2, 3, 4, 5}. Through experiments, the highest Marco-P, Marco-R, and Marco-F₁

Table 3 Fault classification effect under focal loss parameters

σ_2	γ	Marco-P	Marco-R	Marco-F ₁
1	0	0.887 2	0.796 9	0.825 3
0.8	0	0.893 6	0.809 0	0.834 3
0.6	0	0.880 5	0.803 7	0.828 6
0.4	0	0.889 9	0.803 5	0.829 4
0.2	0	0.884 5	0.811 2	0.830 1
0.8	0.5	0.894 1	0.805 2	0.826 6
0.8	1	0.898 5	0.805 9	0.834 2
0.8	2	0.875 5	0.803 5	0.823 4
0.8	3	0.907 7	0.811 2	0.844 2
0.8	4	0.884 5	0.807 2	0.830 1
0.8	5	0.870 5	0.805 9	0.820 4

are obtained when $\gamma = 3$. It shows that the balance between easy/hard samples and minority/majority class samples can be found when $\sigma_2 = 0.8$ and $\gamma = 3$. It can improve the imbalance classification effect of on-board equipment fault to a certain extent.

The relational operation between capsules is often determined by dynamic routing, and different routing iteration times will affect the effect of fault classification. In dynamic routing between the primary capsule layer and the fully connected capsule layer, the routing iteration time is searched on the range of {1, 2, 3, 4, 5, 6, 7, 8}. The experimental results are shown in Fig. 4. According to the results, it can be observed that the effect of fault classification gets better with the increase of dynamic routing iteration times initially and reaches the peak when the routing iteration time is set to 4. Its Marco-P, Marco-R, and Marco-F₁ reach 0.907 7, 0.811 2, 0.844 2, respectively, which achieve the best fault classification effect. This may be because the dynamic routing algorithm is easy to converge at this time. After that, the effect of fault classification decreases as the routing iteration times increases. Considering the reason, when the times of iterations are less than 3, the dynamic connection between

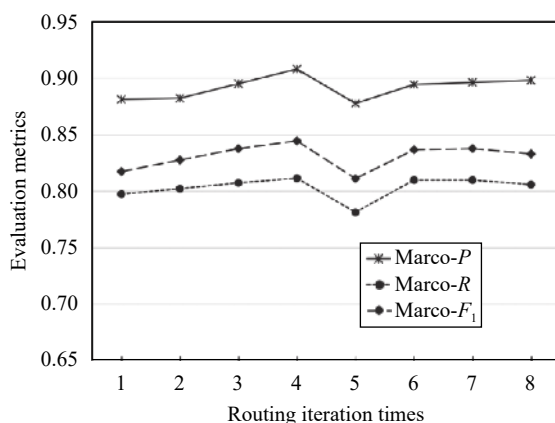


Fig. 4 Comparison of different routing iteration times

the primary capsule layer and the fully connected capsule layer cannot be fully connected, and the optimal routing relationship between the capsules cannot be found, resulting in poor performance. When the routing iteration times increased to 5, the performance decreases slightly. With the increase of iteration times, it takes a longer time and easily leads to over-fitting, which leads to the degradation of fault classification performance. When the number of routing iterations is 4, the model can achieve high precision, recall and F1-Measure in the fault classification of on-board equipment.

4.4 Model evaluation and comparison

To verify the effectiveness of the ATT-Capsule model in the fault classification of high-speed railway on-board equipment, the model is compared with other baseline models. Each model is tested with the optimal parameters to ensure the effectiveness of the comparative experimental results. Considering the influence of the imbalance of on-board fault samples on the classification model, precision, recall, and F1-Measure are used as the evaluation metrics. The results are shown in Table 4.

As shown in Table 4, compared with the baseline models, the ATT-capsule model proposed in this paper has the best fault classification effect for on-board equipment. SVM and RF are traditional machine learning algorithms. When judging the fault types of on-board equipment, the two models will treat the on-board equipment in the normal state (majority class) and the fault samples (minority class) equally based on the class balance hypothesis. However, in the fault classification of on-board equipment, it is important to accurately identify the fault type (minority class), so the two models are not effective in fault classification. Simultaneously, the operation state statements of on-board equipment are complex and vary in length, so it is essential to extract features from samples. High-quality features will improve the effectiveness of the fault classification model. Traditional machine learning depends on artificial feature design,

Table 4 Experimental results of on-board equipment fault classification

Models	Marco-P	Marco-R	Marco-F ₁
SVM	0.833 6	0.709 9	0.745 1
RF	0.871 4	0.724 5	0.753 2
LSTM	0.835 6	0.741 1	0.766 3
BiLSTM	0.887 2	0.731 2	0.768 6
TextCNN	0.867 8	0.731 7	0.770 8
DCNN	0.845 8	0.732 3	0.768 9
CapsNet	0.895 8	0.760 0	0.792 8
GRU-CapsNet	0.860 8	0.738 7	0.768 5
ATT-Capsule	0.907 7	0.811 2	0.844 2

which has some limitations in feature extraction of operation state statements. Compared with SVM, RF improves the precision and recall of fault classification by 3.78% and 1.46%. The RF model adopts the ensemble learning strategy based on the decision tree to comprehensively judge the fault types of on-board equipment, enhancing the generalization ability of the model and improving the fault recognition effect.

The performance of most deep learning methods in fault classification of on-board equipment is better than that of traditional machine learning methods. The deep learning method can automatically extract the embedding features of operation state statements, reduce the need for feature engineering, and improve the quality of feature extraction of on-board fault samples. When embedding is used as model input, TextCNN is better than LSTM and BiLSTM in fault classification. The F1-Measure of LSTM and BiLSTM are 0.766 3 and 0.768 6, respectively, while that of TextCNN is 0.770 8, which is higher than that of the former two models. LSTM can reflect the relationship between two distant words, which is suitable for long text modeling. The on-board operation state statements are short text structure, which is more suitable to use a CNN model to extract N-gram features from different positions of the state statement in parallel to serve the final fault type output. However, the pooling layer of CNN can only extract the most significant or average semantic features in state statements, ignoring the semantic information, which is helpful to fault classification but has a low probability of occurrence. In sequence modeling, the pooling operation will cause information loss of the local position and overall sequence structure and destroy the word order feature of the operation state statements.

Compared with other baseline models, the ATT-Capsule model proposed in this paper has the highest precision, recall, and F1-Measure in the on-board fault classification of high-speed railway. Compared with CapsNet, the F1-Measure of the CNN model in fault classification is increased by 2.2%, and the recall is increased by 2.83%. In the case of the imbalanced number of fault samples, the recall is significantly improved. The results show that the CapsNet uses vector-output capsules to replace CNN scalar-output to enrich the attribute feature information in the on-board operation state statements. The dynamic routing between capsule layers can dynamically assign the attribute feature in operation state statements to all kinds of categories, which can retain all the semantic features and word order features in the sentence. ATT-Capsule introduces the attention mechanism into the capsule network and makes the model pay more attention to the features that play a key role in the fault classification results. Simultaneously, the model dynamically adjusts the impact of imbalanced samples on the loss function in the training process, which helps identify the fault type (minority class) accurately. Compared with CapsNet, the

ATT-Capsule fault classification model increased from 0.76 to 0.811 2 in the recall. Although GRU-CapsNet also uses the capsule layer, GRU pays more attention to the remote information capture between words and does not fully extract the short-distance hierarchical features of the operation state statements. Hence, the effect of feature extraction is not good, which affects the final fault classification effect of on-board equipment.

4.5 Attention mechanism effect verification

To verify the influence of the attention mechanism proposed in this paper on the fault classification effect of high-speed railway on-board equipment, a fault classification model is established by combining the attention mechanism with the capsule network and CNN: ATT-Capsule, Capsule, ATT-CNN, and CNN. ATT-Capsule is the model proposed in this paper. The Capsule model only removes the attention layer based on the ATT-Capsule, and the input, parameters, and training process of the Capsule model are consistent with the model proposed in this paper. CNN model includes a convolutional layer, a max-pooling layer, and a full connection layer. The ATT-CNN model introduces the attention mechanism based on the CNN model. These two fault classification models also use the Adam optimization method to minimize the multi-class focal loss over the training data. The input, filter window parameters, and focal loss function parameters of the two models are consistent with the model in this paper. The *Marco-P*, *Marco-R*, and *Marco-F₁* of the fault classification result of on-board equipment are obtained by using each model, and these three metrics are used as the evaluation criteria of fault classification performance. The experimental results are shown in Fig. 5.

The experimental results show that the fault classification performance of the model with the attention mechanism is better than that without the attention mechanism. Compared with Capsule, the *Marco-P*, *Marco-R*, and *Marco-F₁* of ATT-Capsule in fault classification increased by 2.02%, 1.25%, and 1.89%, respectively. When

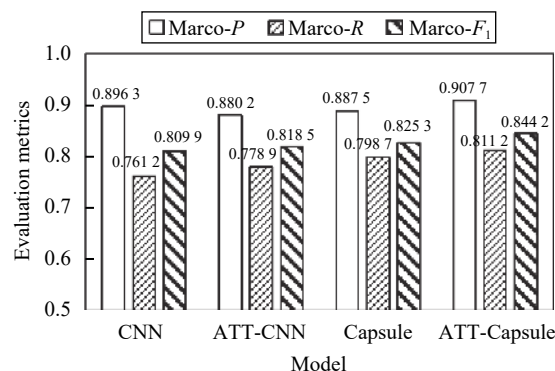


Fig. 5 Comparison of different models

the CNN model is combined with the attention layer, the *Marco-F₁* of fault classification increases to 0.8185, while the *Marco-F₁* of CNN is only 0.8099, it shows that these attention-based methods can obtain more important and differentiated information for fault classification results from on-board operation status statements under the supervision of on-board equipment type tags F1-F20 and N. Compared with the ATT-CNN model, the ATT-Capsule model increases 3.23% and 2.57% respectively in the *Marco-R* and *Marco-F₁* of fault classification, and the recall is improved obviously. This metric is the main basis to measure the correct classification of on-board fault samples. It shows that the Capsule network can learn the part and whole association information of the on-board log to get abundant information of features from the input operation state statements, reduce the loss of semantic information, and improve the effect of on-board equipment fault classification. It also shows the value and feasibility of introducing the attention mechanism into the capsule network.

5 Conclusions

The fault classification for on-board equipment of high-speed railways is investigated in this paper. Taking the on-board log as data source, a fault classification model based on an attention capsule network is proposed. The attention mechanism is introduced to calculate the dependencies between words in the on-board log and capture important information, which solves the problem that the capsule network cannot selectively pay attention to the information that is important and distinguishable to the classification results. To effectively capture the part-whole relationship information and reduce information loss, the capsule network is used to activate high-level features by dynamic routing agreement between low-level features. A multi-class focal loss function is used to train the model to deal with the imbalance of samples. Through experiments on the on-board log provided by a railway bureau, our results conclusively show that the ATT-Capsule model is superior to other models in terms of *Marco-P*, *Marco-R*, and *Marco-F₁*. It also provides theoretical basis and application value for the fault classification of high-speed railway on-board equipment.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No.61763025), Gansu Science and Technology Program Project (No.18JR3RA104), Industrial support program for colleges and universities in Gansu Province (No.2020C-19), Lanzhou Science and Technology Project (No.2019-4-49).

References

- [1] X. Ma, Y. B. Si, Z. Y. Yuan, Y. H. Qin, Y. Q. Wang. Mul-

- tistep dynamic slow feature analysis for industrial process monitoring. *IEEE Transactions on Instrumentation and Measurement*, vol.69, no.12, pp.9535–9548, 2020. DOI: [10.1109/TIM.2020.3004681](https://doi.org/10.1109/TIM.2020.3004681).
- [2] Y. Zhao, T. H. Xu. Text mining based fault diagnosis for vehicle on-board equipment of high speed railway signal system. *Journal of the China Railway Society*, vol.37, no.8, pp.53–59, 2015. DOI: [10.3969/j.issn.1001-8360.2015.08.009](https://doi.org/10.3969/j.issn.1001-8360.2015.08.009). (in Chinese)
- [3] X. Liang, H. F. Wang, J. Guo, T. H. Xu. Bayesian network based fault diagnosis method for on-board equipment of train control system. *Journal of the China Railway Society*, vol.39, no.8, pp.93–100, 2017. DOI: [10.3969/j.issn.1001-8360.2017.08.013](https://doi.org/10.3969/j.issn.1001-8360.2017.08.013). (in Chinese)
- [4] W. Shangguan, Y. H. Yuan, J. Wang, F. W. Hu. Research of fault feature extraction and diagnosis method for CTCS on-board equipment (OBE) based on labeled-LDA. *Journal of the China Railway Society*, vol.41, no.8, pp.56–66, 2019. DOI: [10.3969/j.issn.1001-8360.2019.08.008](https://doi.org/10.3969/j.issn.1001-8360.2019.08.008). (in Chinese)
- [5] L. J. Zhou, Y. Dong. Research on fault diagnosis method for on-board equipment of train control system based on GA-BP neural network. *Journal of Railway Science and Engineering*, vol.15, no.12, pp.3257–3265, 2018. DOI: [10.19713/j.cnki.43-1423/u.2018.12.031](https://doi.org/10.19713/j.cnki.43-1423/u.2018.12.031). (in Chinese)
- [6] Z. J. Lou, Y. Q. Wang. New nonlinear approach for process monitoring: Neural component analysis. *Industrial & Engineering Chemistry Research*, vol.60, no.1, pp.387–398, 2021. DOI: [10.1021/acs.iecr.0c02256](https://doi.org/10.1021/acs.iecr.0c02256).
- [7] K. Aukkapinyo, S. Sawangwong, P. Pooyoi, W. Kusakuniran. Localization and classification of rice-grain images using region proposals-based convolutional neural network. *International Journal of Automation and Computing*, vol.17, no.2, pp.233–246, 2020. DOI: [10.1007/s11633-019-1207-6](https://doi.org/10.1007/s11633-019-1207-6).
- [8] A. X. Li, K. X. Zhang, L. W. Wang. Zero-shot fine-grained classification by deep feature learning with semantics. *International Journal of Automation and Computing*, vol.16, no.5, pp.563–574, 2019. DOI: [10.1007/s11633-019-1177-8](https://doi.org/10.1007/s11633-019-1177-8).
- [9] L. C. Li, Z. Y. Wu, M. X. Xu, H. L. Meng, L. H. Cai. Combining CNN and BLSTM to extract textual and acoustic features for recognizing stances in mandarin ideological debate competition. In *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, San Francisco, USA, pp.1392–1396, 2016. DOI: [10.21437/Interspeech.2016-324](https://doi.org/10.21437/Interspeech.2016-324).
- [10] S. Sabour, N. Frosst, G. E. Hinton. Dynamic routing between capsules. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ACM, Long Beach, USA, pp.3859–3869, 2017.
- [11] M. Yang, W. Zhao, L. Chen, Q. Qu, Z. Zhao, Y. Shen. Investigating the transferring capability of capsule networks for text classification. *Neural Networks*, vol.118, pp.247–261, 2019. DOI: [10.1016/j.neunet.2019.06.014](https://doi.org/10.1016/j.neunet.2019.06.014).
- [12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ACM, Long Beach, USA, pp.6000–6010, 2017.
- [13] X. Zhang, Q. Yang. Transfer hierarchical attention network for generative dialog system. *International Journal of Automation and Computing*, vol.16, no.6, pp.720–736, 2019. DOI: [10.1007/s11633-019-1200-0](https://doi.org/10.1007/s11633-019-1200-0).

- [14] B. Liang, Q. Liu, J. Xu, Q. Zhou, P. Zhang. Aspect-based sentiment analysis based on multi-attention CNN. *Journal of Computer Research and Development*, vol. 54, no. 8, pp. 1724–1735, 2017. DOI: [10.7544/j.issn1000-1239.2017.20170178](https://doi.org/10.7544/j.issn1000-1239.2017.20170178). (in Chinese)
- [15] Y. Kim, H. Lee, K. Jung. AttnConvnet at semeval-2018 task 1: Attention-based convolutional neural networks for multi-label emotion classification. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, pp. 141–145, 2018. DOI: [10.18653/v1/S18-1019](https://doi.org/10.18653/v1/S18-1019).
- [16] China Railway Corporation. *Typical Faults of Train Control On-board Equipment*, Beijing, China: China Railway Publishing House, pp. 29–87, 2013. (in Chinese)
- [17] J. Z. Zhang. Research on Fault Combination Prediction Method for on-board Equipment of CTCS Based on Cross Entropy Theory, Master dissertation, Beijing Jiaotong University, China, 2019. (in Chinese)
- [18] Y. B. Si, Y. Q. Wang, D. H. Zhou. Key-performance-indicator-related process monitoring based on improved kernel partial least squares. *IEEE Transactions on Industrial Electronics*, vol. 68, no. 3, pp. 2626–2636, 2021. DOI: [10.1109/TIE.2020.2972472](https://doi.org/10.1109/TIE.2020.2972472).
- [19] T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient Estimation of Word Representations in Vector Space, [Online], Available: <https://arxiv.org/abs/1301.3781>, 2013.
- [20] Z. W. Zhao, Y. Z. Wu. Attention-based convolutional neural networks for sentence classification. In *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, San Francisco, USA, pp. 705–709, 2016. DOI: [10.21437/Interspeech.2016-354](https://doi.org/10.21437/Interspeech.2016-354).
- [21] B. Z. Guo, W. L. Zuo, Y. Wang. Double CNN sentence classification model with attention mechanism of word embeddings. *Journal of Zhejiang University (Engineering Science)*, vol. 52, no. 9, pp. 1729–1737, 2018. DOI: [10.3785/j.issn.1008-973X.2018.09.013](https://doi.org/10.3785/j.issn.1008-973X.2018.09.013). (in Chinese)
- [22] T. Y. Lin, P. Goyal, R. Girshick, K. M. He, P. Dollar. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020. DOI: [10.1109/TPAMI.2018.2858826](https://doi.org/10.1109/TPAMI.2018.2858826).
- [23] S. Ioffe, C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, ACM, New York, USA, pp. 448–456, 2015.
- [24] S. del Rio, V. Lopez, J. M. Benitez, F. Herrera. On the use of MapReduce for imbalanced big data using Random Forest. *Information Sciences*, vol. 285, pp. 112–137, 2014. DOI: [10.1016/j.ins.2014.03.043](https://doi.org/10.1016/j.ins.2014.03.043).
- [25] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation, [Online], Available: <https://arxiv.org/abs/1406.1078>, Sep 3, 2014.
- [26] W. Shangguan, Y. Y. Meng, J. M. Yang, B. G. Cai. LSTM-BP neural network based fault diagnosis for on-board equipment of Chinese train control system. *Journal of Beijing Jiaotong University*, vol. 43, no. 1, pp. 54–62, 2019. DOI: [10.11860/j.issn.1673-0291.2019.01.006](https://doi.org/10.11860/j.issn.1673-0291.2019.01.006).
- [27] Y. Kim. Convolutional neural networks for sentence classification. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar, pp. 1746–1751, 2014. DOI: [10.3115/v1/D14-1181](https://doi.org/10.3115/v1/D14-1181).
- [28] N. Kalchbrenner, E. Grefenstette, P. Blunsom. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, USA, pp. 655–665, 2014. DOI: [10.3115/v1/P14-1062](https://doi.org/10.3115/v1/P14-1062).
- [29] P. Rathnayaka, S. Abeysinghe, C. Samarajeewa, I. Manchanayake, M. Walpola. Sentylic at IEST 2018: Gated recurrent neural network and capsule network based approach for implicit emotion detection. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, ACL, Brussels, Belgium, pp. 254–259, 2018. DOI: [10.18653/v1/W18-6237](https://doi.org/10.18653/v1/W18-6237).
- [30] M. Buda, A. Maki, M. A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, vol. 106, pp. 249–259, 2018. DOI: [10.1016/j.neunet.2018.07.011](https://doi.org/10.1016/j.neunet.2018.07.011).



Lu-Jie Zhou received the B.Sc. degree in traffic information engineering & control from Lanzhou Jiaotong University, China in 2015. She is currently a Ph.D. degree candidate in traffic information engineering & control from Lanzhou Jiaotong University, China.

Her research interests include intelligent fault diagnosis and natural language

processing.

E-mail: 792321186@qq.com (Corresponding author)

ORCID: 0000-0003-4808-6942



Jian-Wu Dang received the Ph.D. degree in electrification & automation of railway traction from Southwest Jiaotong University, China in 1996. He is a professor, doctoral supervisor, vice president of Lanzhou Jiaotong University, China. He is a national candidate for the New Century Ten Million Talent Project and one of the first batch of special science and technology experts in Gansu Province.

He is an expert with outstanding contributions from the Ministry of Railways and won the 6th Zhan Tianyou Railway Science and Technology Award. He has published five monographs and published more than 170 academic papers.

His research interests include intelligent information processing, intelligent transportation, and image processing.

E-mail: dangjiw@mail.lzjtu.cn



Zhen-Hai Zhang received the Ph.D. degree in traffic information engineering & control from Lanzhou Jiaotong University, China in 2014. He is an associate professor, master supervisor of Lanzhou Jiaotong University, China. Now he is in charge of the National Natural Science Foundation, the postdoctoral fund of China, the Natural Science Foundation of Gansu Province.

He has published 14 relevant academic papers and participated in the compilation of 2 teaching materials.

His research interest is intelligent transportation.

E-mail: zhangzhenhai@lzjtu.cn