

Dual-Path Neural Network based on Deep Features and Transfer Learning for Bearing Fault Diagnosis

Juntao Tong *

College of Engineering, Zhejiang Normal University, Jinhua, Zhejiang, 321004, China

* Corresponding Author Email: technosense@163.com

ABSTRACT

The safe operation of mechanical equipment is a basic requirement in the industrial manufacturing process. As an important component of mechanical equipment, rolling bearings are prone to unforeseen failures due to long-term operation under complex conditions. The occurrence of failures, no matter how big or small, will cause economic losses. Therefore, reliable diagnosis of rolling bearings is crucial. Aiming at the problem of insufficient feature recognition of convolutional neural networks under strong noise background, a deep feature extraction network integrating multiscale convolutional neural network (MSCNN) and bidirectional gated recurrent unit (BiGRU) is proposed. MSCNN and BiGRU are used to extract multiscale features and temporal features from noisy vibration signals respectively, and different weights are assigned to the fused features through the attention mechanism module to achieve important feature selection. Furthermore, in order to solve the problem of different feature distributions between the source domain and the target domain under variable working conditions, transfer learning is introduced in the proposed deep feature extraction network. The difference in feature distribution between the source domain and the target domain is measured by a multi-level distance formula, and the measurement result is added to the loss function. The back propagation of the loss is used to achieve the alignment of feature distribution between the source domain and the target domain. Finally, the model uses the SoftMax function as a classifier for rolling bearing fault diagnosis. Experimental comparison and analysis show that the proposed model has good migration ability and achieves a higher fault identification accuracy.

KEYWORDS

Fault Diagnosis; Transfer Learning; Multi-scale Features; Time Series Features; Attention Mechanism.

1. INTRODUCTION

As an important component of mechanical equipment, rolling bearings are very susceptible to fatigue during long-term operation. Once a failure occurs, it will affect the stable operation of the entire mechanical equipment. Therefore, research on fault diagnosis of rolling bearings has important theoretical significance and practical engineering value[1].

In order to solve the problem of rolling bearing fault diagnosis, convolutional neural network (CNN) has been widely used in the field of fault diagnosis due to its excellent feature extraction ability. Tian et al. [2] proposed a rolling bearing fault diagnosis method based on two-dimensional CNN. The WT method was used to convert the one-dimensional bearing vibration signal into a two-dimensional time-frequency image to form a new image data set. The image was then input into the Resnet34 network for feature extraction and fault identification. Fan et al. [3] combined convolutional neural network with support vector machine and proposed a rolling bearing fault diagnosis method based on

information fusion, empirical mode decomposition and polar coordinate adaptive distribution. It can accurately identify the composite fault between bearing and rotor. However, although CNN has a strong feature extraction ability, it is difficult to fully obtain the intrinsic expression of the fault signal in time.

Compared with CNN, recurrent neural network (RNN) has a strong advantage in obtaining time correlation. Among the different networks based on RNN that have emerged later, long short term memory network (LSTM) is the most widely used. For example, Afridi et al. [4] proposed a rolling bearing fault prediction system based on LSTM network and using original time series data. The system performed well in terms of root mean square error and had excellent generalization ability. Wang et al. [5] proposed a rolling bearing vibration signal fault detection method that combines continuous wavelet transform, dual-channel convolutional neural network and LSTM network, with a detection accuracy of 99.98%. Although it has achieved certain application effects, the complex structure of LSTM makes network training more difficult. The gated recurrent unit (GRU) based on LSTM effectively solves the problem of model training difficulty. Considering the respective advantages of CNN and GRU, in order to improve the diagnostic performance of the network under strong noise, this study proposed a rolling bearing fault diagnosis model based on deep feature extraction and transfer learning. Firstly, the MSCNN-BiGRU deep feature extraction network is constructed to extract multi-scale features and temporal features under strong noise, and the attention mechanism module is used to select important features. Secondly, transfer learning is introduced to measure the difference in feature distribution between the source domain and the target domain through the multi-level distance formula, and the measurement results are added to the loss function. The back propagation of the loss is used to align the feature distribution of the source domain and the target domain to solve the problem of different feature distribution between the source domain and the target domain under variable working conditions. Finally, the softmax layer is used to perform fault mode recognition.

2. PRELIMINARIES

2.1. Convolutional Neural Networks

Convolutional neural networks are mainly composed of convolutional layers, pooling layers, and fully connected layers [6]. The convolutional layer uses convolution kernels to perform convolution operations on local areas of input data and uses activation functions to further generate output features. The calculation process is as follows:

$$y_n^l = f \sum_{i=1}^m (x_i^{l-1} \times w_n^l + b_n^l) \quad (1)$$

Where: y_n^l is the n th feature tensor output by layer l ; w_n^l is the weight matrix; x_i^{l-1} is the i -th feature quantity output by layer $l - 1$; b_n^l is the bias matrix; f is the activation function.

The pooling layer can further reduce the network parameters. The maximum pooling method is selected, and its calculation formula is:

$$p_{\max}^{l(i,j)} = \max_{(j-1)w \leq t \leq jw} \{u^{l(i,t)}\} \quad (2)$$

Where: $u^{l(i,t)}$ is the t th neuron of the i th feature tensor in the l th layer; w is the convolution kernel width.

For multi-classification tasks, the fully connected layer often uses the softmax function for classification, and its expression is as follows:

$$\alpha_i = \text{softmax}(\varphi_i) = \frac{e^{\varphi_i}}{\sum_{i=1}^C e^{\varphi_i}} \quad (3)$$

Where: C is the number of categories; φ_i is the output value of the i -th category; α_i is the probability of the i -th category.

2.2. Maximum Mean Difference (MMD)

Rolling bearings often operate under variable working conditions. To achieve effective diagnosis, it is necessary to comprehensively consider the domain data from different working conditions for analysis. The source domain and the target domain have the same feature space and category space, but their marginal distribution and conditional distribution are different [7].

In order to measure the distribution difference between the two domains, MMD is introduced for quantitative analysis. It maps the features in the source domain and the target domain simultaneously into the reproducing kernel Hilbert space, and calculates the distance between the means of the two domains to characterize the data distribution difference between the domains. The calculation formula is:

$$\text{MMD}(X, Y) = \left\| \sum_{i=1}^{n_s} \phi(x_i) - \sum_{j=1}^{n_t} \phi(x_j) \right\|_{H_k}^2 \quad (4)$$

Where: n_s and n_t are the number of samples in the source and target domains respectively.

3. APPROACH

3.1. Feature Extraction based on Fusion Deep Network

The working environment of rolling bearings is complex, and the effective components in the collected signals are often submerged by noise. Considering that the MSCNN network can adaptively extract multi-scale features from the signal and thus obtain richer fault feature information, and the BiGRU network can learn the temporal correlation of vibration signals from both the front and back directions, this study organically combines the advantages of the two networks and introduces the attention mechanism to construct a deep feature extraction network based on MSCNN-BiGRU-Attention.

3.2. Diagnostic Model based on Deep Transfer Network

MK-MMD and JMMD are respectively embedded in the loss function to measure the difference in feature distribution between the source domain and the target domain, and the measurement results are added to the loss function to construct a deep migration network with certain domain adaptation capabilities. Its structure consists of two modules: state recognition and domain adaptation. The state recognition module includes the MSCNN-BiGRU-Attention feature extraction network, the bottleneck layer and the label classifier, where the bottleneck layer and the label classifier both use a single-layer fully connected network with 256 and 5 neurons respectively. The domain adaptation module consists of the MK-MMD metric and the JMMD metric.

The main implementation steps of the diagnostic model based on the proposed deep migration network are as follows:

Step 1: Data preprocessing. Use sliding windows to perform non-overlapping cuts on the original vibration signal; at the same time, in order to keep the input value of each sample within a certain range, a normalization method is used.

Step 2: Network pre-training. In the first 30 trainings, only labeled source domain samples are used to pre-train the state recognition module so that the feature extraction network, bottleneck layer and label classifier obtain better model parameters.

Step 3: Domain transfer training. On the basis of the labeled source domain sample training model, unlabeled target domain samples are added for simultaneous training. Among them, MK-MMD generates LM loss, JMMD generates LJ loss, and the two together with the label classification loss

constitute the total loss, and then back propagation is used to reduce the difference in feature distribution between the source domain and the target domain.

Step 4: Fault diagnosis. Use the fully trained model to perform state recognition on the unlabeled target domain samples of the test set to complete the diagnosis.

3.3. Objective Optimization Function

The loss function L of the proposed model consists of three parts, namely, label classification loss L_P , MK-MMD metric loss L_M and JMMD metric distribution loss L_J . The loss function L formula is described as follows:

$$L(\theta_G, \theta_B, \theta_P) = L_P + L_M + L_J \quad (5)$$

Where: θ_G represents the feature extraction network parameters, θ_B is the bottleneck layer parameter, and θ_P is the label classifier parameter.

3.3.1. Label Classification Loss

The feature extraction network, bottleneck layer and label classifier are supervised by the labeled source domain samples. The label classification loss is measured by the cross entropy loss function, which is as follows:

$$L_P = -\frac{1}{n_s} \sum_{x_i \in D_B} \sum_{C=1}^C P_{x_i \rightarrow C} \lg G_P \left(G_B(G_G(x_i)) \right) \quad (6)$$

Where: G_G represents the feature extraction network, G_B is the bottleneck layer, and G_P represents the label classifier.

3.3.2. Multi-core Maximum Mean Metric Loss

Although MMD has a certain ability to measure differences, it is based on a single kernel transformation and is difficult to measure distribution differences in multiple spaces. For this reason, a multi-kernel mode is introduced, that is, a linear combination of multiple kernels, as follows:

$$K \stackrel{\Delta}{=} \{k = \sum_{u=1}^m \beta_u k_u : \sum_{u=1}^m \beta_u = 1, \beta \geq 0, \forall u\} \quad (7)$$

Where: β_u is the weighting parameter of different kernels.

Based on the above formula, we get the multi-core MMD metric (MK-MMD) with stronger representation ability, and then get the MK-MMD metric loss L_M , which is expressed as follows:

$$L_M = \|E_S[\phi(x^{n_t})] - E_t[\phi(x^{n_t})]\|_H^2 \quad (8)$$

3.3.3. Joint Maximum Mean Metric Loss

Based on the literature [8], JMMD is introduced to measure the difference in marginal distribution and conditional distribution between the source domain and the target domain. The formula of JMMD is as follows:

$$L_J(D_S, D_t) = \|E_P \left(\bigotimes_{l=1}^{|L|} \phi^l(z^{ll}) \right) - E_Q \left(\bigotimes_{l=1}^{|L|} \phi^l(z^{ll}) \right)\|_{\bigotimes_{l=1}^{|L|} H, H'}^2 \quad (9)$$

Where: $\bigotimes_{l=1}^{|L|} \phi^l(z^{ll}) \dots \bigotimes_{l=1}^{|L|} \phi^l(z^{ll})$ is the feature map of the tensor product in the Hilbert space, L is the set of higher network layers, l is the layer in the corresponding set, z^{sl} represents the activation function of the l th layer of the source domain, and z^{tl} represents the activation function of the l th layer of the target domain.

4. EXPERIMENT AND ANALYSIS

4.1. Data Collection

This study is based on the mechanical fault comprehensive simulation test bench designed by Spectra Quest for fault data collection. It is mainly composed of speed controller, rotating shaft and acceleration sensor. The acceleration sensor model is PCB352C33, the experimental bearing model is ER12KCL, and the diameter of the experimental rolling bearing is 19.05 mm.

The experimental data were collected at 1400 r/min, 1800 r/min and 2200 r/min. The sampling frequency was 12.8 kHz and the acquisition time was 10 s. The rolling bearing has four fault states: inner ring (IF), outer ring (OF), rolling element (BF) and comprehensive (CF) and a normal state (N).

4.2. Data Preprocessing

The original data is cut into non-overlapping parts. Based on the sampling frequency, three sampling cycles are selected, that is, the sample length is set to 1024, and finally 125 samples are obtained for each type of fault, as shown in Table 1. In order to eliminate the adverse effects caused by the sample singular data, normalization processing is performed.

Table 1. Fault Dataset Description

Bearing condition	Number of samples		
	1400r/min	1800r/min	2200r/min
N	125	125	125
IF	125	125	125
OF	125	125	125
BF	125	125	125
CF	125	125	125

4.3. Data Collection

A 2 dB strong noise background is selected, and A, B, and C are used to represent three speed data sets of 1400 r/min, 1800 r/min, and 2200 r/min, respectively. The six groups of migration tasks are represented by A to B, A to C, B to A, B to C, C to A, and C to B, respectively. For example, A to B represents: the source domain is a 1400 r/min data set and the target domain is a 1800 r/min data set. 80% of the labeled source domain sample set and the unlabeled target domain sample set are selected as the training set, and the remaining 20% of the target domain sample set is used as the test set. To verify the migration performance of the proposed model, it is compared with five domain adaptation models, including DCORAL, DAN, JAN, DANN, and CDAN. Among them, DAN, JAN, and DCORAL are deep domain adaptation models based on explicit distance metrics, and DANN and CDAN are deep domain adaptation models based on adversarial learning. The experiment used the Pytorch deep learning framework, with a batch size of 64, a learning rate of 0.001, and an epoch of 100. Five experiments were conducted simultaneously on the six groups of migration tasks, and the average of the experimental results was taken to evaluate the model performance. The average accuracy and standard deviation of each model in all tasks are shown in Table 2.

Table 2. Fault diagnosis results of different models

Model	Average accuracy/%	Accuracy standard deviation/%
DCORAL	75.21	5.28
DAN	88.49	3.32
JAN	89.70	2.07
DANN	90.02	3.36
CDAN	90.27	1.83
OURS	95.28	1.49

From the comparison of migration results, it can be seen that the DCORAL model has the worst diagnostic performance, with a recognition accuracy of only 75.21%. Among them, the DAN and DANN models only focus on marginal distribution adaptation, with recognition accuracy rates of 88.49% and 90.02% respectively, but their accuracy rates are significantly higher than DCORAL. The JAN model adapts marginal distribution and conditional distribution at the same time, with an accuracy rate of 89.70%, slightly higher than the DAN model. The CDAN model is improved on the basis of the DAN model, with a recognition accuracy rate of 90.27%. In comparison, the proposed migration method has the highest recognition accuracy of 95.28%, and the smallest standard deviation, which is significantly better than the five comparison models.

5. CONCLUSION

Rolling bearing fault diagnosis model based on deep feature extraction and transfer learning under strong noise. Its deep feature extraction network performs deep fault feature mining on noisy vibration signals by fusing MSCNN and BiGRU and introducing an attention mechanism module. In order to solve the problem of different data distribution between the source domain and the target domain caused by variable working conditions, transfer learning is introduced into the model. The difference in feature distribution between the source domain and the target domain is reduced through multi-level distance measurement, thereby realizing fault diagnosis of unlabeled target domain samples.

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- [1] K Wu, X. M., et al. "Application of machine learning in bearing fault diagnosis." *Equip. Manuf. Technol* 3 (2022): 03-118.
- [2] Tian, Zhenzhen, et al. "A Bearing Fault Diagnosis Method Based on 2D Images and Transformer." 2023 13th International Conference on Information Technology in Medicine and Education (ITME). IEEE, 2023.
- [3] Hongwei, Fan, et al. "A novel intelligent diagnosis method of rolling bearing and rotor composite faults based on vibration signal-to-image mapping and CNN-SVM." *Measurement Science and Technology* 34.4 (2023): 044008.
- [4] Afridi, Yasir Saleem, et al. "LSTM-Based Condition Monitoring and Fault Prognostics of Rolling Element Bearings Using Raw Vibrational Data." *Machines* 11.5 (2023): 531.
- [5] Wang, Yu, et al. "Research on fault detection of rolling bearing based on CWTDCNN-LSTM." *Engineering Letters* 31.3 (2023).
- [6] Wang, Zhuozheng, et al. "A novel fault diagnosis approach for chillers based on 1-D convolutional neural network and gated recurrent unit." *Sensors* 20.9 (2020): 2458.
- [7] Guo, Liang, et al. "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data." *IEEE Transactions on Industrial Electronics* 66.9 (2018): 7316-7325.
- [8] Long, Mingsheng, et al. "Deep transfer learning with joint adaptation networks." *International conference on machine learning*. PMLR, 2017.