Deep transfer attention network for intelligent fault diagnosis of rolling bearings

Shihao Li, Feng Jia, Jianjun Shen and Junxing Ma

Key Laboratory of Road Construction Technology and Equipment, Ministry of Education, Chang’an University, Xi'an 710064, China

Abstract. Deep learning based on fault diagnosis methods of rolling bearings has attracted much attention with the rapid development of artificial intelligence. However, these methods trained by bearing datasets from one equipment cannot be well applied to correctly identify the fault information of the bearing datasets from another equipment. The main reason for this problem is that the datasets from the different equipment have different probability distributions. In order to solve the above problem, a deep transfer attention network is proposed for intelligent fault diagnosis of the rolling bearings. The steps of this method are listed as follows. Firstly, the feature extraction network is used to extract the features from different bearing datasets. Secondly, the different transfer attention is given to different areas of the sample data by the attention mechanism and distribution differences among the features is reduced by domain adaptation. Finally, the fault recognition network is trained to identify fault status for different bearings. The proposed method is verified by using fault datasets from different bearings equipment. The results show that the proposed method can achieve the fault identification of different bearings equipment.

1. Introduction
Deep fault diagnosis methods of rolling bearings have attracted many attentions with the rapid development of artificial intelligence, and it has been an important tool to ensure the safe and normal operation of bearings in the background of big data. [1] For example, Wu et al. proposed the multiple incipient sensor faults diagnosis method and a descriptor system approach for estimation of incipient faults and apply above two methods to high-speed railways. [2-3] In the deep fault diagnosis method, however, the training datasets and the testing datasets are both from the same rolling bearings equipment. When the training datasets and the testing datasets are from different rolling bearings equipment, the deep fault diagnosis method shows poor diagnostic performance in the testing datasets. The main reason is that the training datasets and the testing datasets have different probability distributions. In order to solve this problem, the transfer fault diagnosis method of rolling bearings is introduced. In transfer diagnosis method of rolling bearings, the domain adaptation is one of the commonly used methods to reduce the probability distribution difference of the training datasets and the testing datasets. Thereby, the transfer diagnosis method trained with the training datasets can be applied to accurately identify the fault status of the testing datasets. Guo et al. [4] proposed transfer diagnosis method based on a one-dimensional convolutional neural network. Shao et al. [5] proposed transfer adversarial fault diagnosis method based on adversarial idea. Zhu et al. [6] proposed deep transfer fault diagnosis method based on the multi-core maximum mean difference (M-MMD) method.
However, the existing transfer diagnosis method is to directly perform domain adaptation for the entire sample data. It is very obvious that different areas of sample data may have different transferability. For example, some corresponding areas of the training sample data and the testing sample data have obvious differences due to the influence of factors such as noise. Therefore, the probability distribution of these areas is significantly different. If these areas are forced to domain adaptation, the diagnostic accuracy of the transfer diagnostic method on the testing datasets could be reduced [7].

In order to solve the above problems and improve the accuracy of the transfer diagnosis method, the deep transfer attention network for intelligent fault diagnosis of rolling bearings is proposed. The proposed method is designed mainly based on the idea of attention mechanism. The idea of attention mechanism is mainly used for assigning different attention to different areas of sample data, so that the areas suitable for transfer are paid more attention and the areas not suitable for transfer are paid less attention. [8] Therefore, the proposed method can judge the areas suitable for transfer and the areas not suitable for transfer according to the different attention, thereby solving the problem of the different areas of sample data may have different transferability.

The main steps are listed as follows. Firstly, the datasets from rolling bearing equipment are mapped to high-dimensional feature space by feature extraction network, thereby the features of datasets from rolling bearing equipment are extracted. Secondly, the attention distribution model is designed to assign different attention to different areas of sample data, thereby capturing the transferable information of different areas of a single sample. Thirdly, the domain discrimination network is designed to reduce probability distribution difference of the training datasets and the testing datasets. Finally, a fault feature recognition network is designed to identify fault of datasets. The model was verified by collecting monitoring signals of different rolling bearings. Experimental results show that the proposed diagnosis method has better diagnosis results than traditional deep transfer fault diagnosis.

The main contributions of this paper are as follows. 1) The proposed method solves the problem that the different areas of sample data have different transferability. And the accuracy of fault recognition is improved based on the existing transfer diagnosis method. 2) The model was verified by collecting monitoring signals of different bearings, which proved the effectiveness of the proposed method.

2. The structure of deep transfer attention network

In this paper, the sample spaces of the training datasets and the testing datasets are called the source domain and the target domain respectively. Let \( D_s = \{x^i, p(x^i)\} \) denote the source domain and \( D_t = \{x^j, p(x^j)\} \) denote the target domain, where \( x^i \) represent the \( i \) source domain sample data and \( x^j \) represent the \( j \) target domain sample data, and \( p(x^i) \) and \( p(x^j) \) represent the probability distribution of the source domain and the target domain, respectively. [9] The source domain datasets and the target domain datasets come from different rolling bearings, so they have the following properties: 1) The source domain datasets and the target domain datasets have different probability distributions; 2) The source domain datasets and the target domain datasets have the same feature space and category space.

The proposed method mainly includes four parts: feature extractor network, attention distribution model, domain discrimination network and fault feature identification network. Its specific structure is shown in Figure 1.

2.1. Feature extractor network

The feature extractor network is mainly composed of a convolutional neural network. Let the feature extractor network be denoted as \( G_f \). Therefore, after inputting the source domain data and target domain data into the feature extraction network, the features obtained can be expressed as:
where $x'_t$ represents the $t$ sample in the union of the source domain and the target domain; $\theta_l$ represents the parameter to be optimized in the $l$-th layer network; $F(\cdot)$ represents the mapping function in the $l$-th layer network.

![Flow chart of deep transfer attention network.](image)

**Figure 1.** Flow chart of deep transfer attention network.

### 2.2. Attention distribution model

Since the transferability of different areas will affect the diagnosis accuracy of the transfer result. Therefore, an attention distribution model is constructed based on the idea of attention mechanism [10]. The specific method of the network is listed as follows. Firstly, the feature $x'_t$ of Equation (1) is equally divided into $m$ pixel areas, where the detailed description of the size of $m$ can be found in Section 3.2. Secondly, the maximum mean difference method is used to measure the transferability of the $m$ area. Finally, the weights of transferability are assigned to the corresponding area. The transferability measure of the $m$ area is expressed as follows:

$$
\text{dis}_s(x'^{s,\alpha}_m, x'^{t,\alpha}_m) = \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(x'^{s,\alpha}_i) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(x'^{t,\alpha}_j)
$$

(2)

where $\text{dis}_s(x'^{s,\alpha}_m, x'^{t,\alpha}_m)$ represents the transferability value of the $m$ area; $x'^{s,\alpha}_m$ represents the $m$ area of the source domain feature and $x'^{t,\alpha}_m$ represents the $m$ area of the target domain feature; $\phi(\cdot)$ represents the mapping formula that can project the source domain datasets of bearings and target domain datasets of bearings to the same feature space; and $n_s$ and $n_t$ represent the number of samples of source domain datasets of rolling bearings and target domain datasets of rolling bearings respectively.

The larger the value of the $\text{dis}_s(x'^{s,\alpha}_m, x'^{t,\alpha}_m)$, the lower the transferability. However, we hope that the higher the transferability, the larger the transferability value obtained. A feasible approach is to use the min-max normalization method to normalize the result obtained in Formula (2), and then calculate the difference between 1. The formula is expressed as follows:

$$
T'_s = \frac{\text{dis}_s(x'^{s,\alpha}_m, x'^{t,\alpha}_m) - \min \{ \text{dis}_s(x'^{s,\alpha}_m, x'^{t,\alpha}_m) \}}{\max \{ \text{dis}_s(x'^{s,\alpha}_m, x'^{t,\alpha}_m) \} - \min \{ \text{dis}_s(x'^{s,\alpha}_m, x'^{t,\alpha}_m) \}}
$$

(3)

where $T' = [T'_s, T'_t, \cdots, T'_m]$ represents a vector composed of $m$ transferability values. According to Equation (4), the maximum high-level abstract feature can be expressed as:
\( x'_i = T \cdot x'_i = [T_{ij} \cdot x_{ij}^s, T_{ij} \cdot x_{ij}^t, \ldots, T_{ij} \cdot x_{ij}^{m} ] \)  \hspace{1cm} (4)

### 2.3. Domain discriminator network

The domain discriminator network is mainly composed of the fully connected network and the binary classification function. Let the domain discriminator network be expressed as \( G_d \). Therefore, the output probability of the domain discriminator network can be expressed as

\[
\hat{d}_i = \left[ \begin{array}{c} d_o \\ d_t \end{array} \right] = \frac{1}{\sum_{e=0}^1 e^{\theta^T \phi(x'_i)}} \left[ e^{\theta^T \phi(x'_i)} \right]
\]

where \( \hat{d}_i \) represents the output probability vector of the \( i \)-th sample of bearings. \( d_o \) represents the probability of the sample belonging to the source domain. \( d_t \) represents the probability of belonging to the target domain. \( \theta \) represents the parameters to be optimized in the domain adaptation network.

According to the output probability of the domain discriminator network, the loss of domain adaptation can be obtained as:

\[
\ell_d = \frac{1}{n} \sum_{i=1}^{n} L_d (\hat{d}_i, d_i)
\]

where \( \ell_d \) represents the binary cross entropy loss function; \( \hat{d}_i \) represents the real domain label. When it is 0, it means the source domain; when it is 1, it means the target domain. At this time, when the network performs the forward propagation calculation, the error loss obtained by Equation (6) is minimized; when the backward gradient calculation is performed, the error is maximized. Finally, through maximum and minimum adversarial training, the feature extractor network can extract domain invariant features datasets of bearings.

### 2.4. Fault feature identification network

The fault feature identification network is mainly composed of fully connected layer. Let the fault feature extraction network be expressed as \( G_f \). Therefore, the output probability of the fault feature recognition network is expressed as follows:

\[
\hat{\gamma}_i = [\gamma_{i,1}, \gamma_{i,2}, \ldots, \gamma_{i,K}]
\]

where \( \hat{\gamma}_i \) represents the predicted probability of bearing failure status; \( \theta \) represents the parameter to be optimized in the fault identification network; \( K \) represents the total type of bearings fault status. According to the output probability of Equation (7), the loss error of the fault feature identification network is expressed as follows:

\[
\ell_f = \frac{1}{n} \sum_{i=1}^{n} L_f (\hat{\gamma}_i, \gamma_i)
\]

where \( \ell_f \) represents the cross entropy loss function.

### 2.5. Objective function

According to the loss function of domain adaptation and loss function of the fault feature identification network, the total loss function can be expressed as follows:

\[
\ell(\theta, \theta_d, \theta_f) = \ell_d + \lambda \ell_f
\]

where \( \lambda \) is the weight parameter.
where $\theta_f$, $\theta_r$ and $\theta_i$ respectively represent the parameters to be optimized of the feature extraction network, the domain discriminator network and the fault feature identification network; $\lambda$ represents the loss trade-off parameter.

According to the optimization function of transfer adversarial learning, the Adam optimization algorithm is adopted. The optimization function can be expressed as follows:

$$\left(\theta_f', \theta_r', \theta_i'\right) = \arg \min_{\theta_f, \theta_r, \theta_i} \max_{\lambda} \ell\left(\theta_f, \theta_r, \theta_i\right)$$

(10)

3. Experimental verification

In order to verify the validity of the proposed method, the monitoring signals are collected from two different bearings. Then use this data to train different diagnostic methods to form a control experiment.

Figure 2. SpectraQuest's mechanical failure simulator.

3.1. Data description

The fault datasets of bearings were collected from SpectraQuest's mechanical failure simulator [11], as shown in Figure 2. The simulator includes two bearings at different locations: underhang bearing and overhang bearing, where the underhang bearing is located between the rotor and the motor and the overhang bearing is located behind the rotor. Its specific location is marked with a red line in the figure.

Let the fault datasets for bearings from underhang bearing be dataset A, and the fault datasets for bearings from overhang bearing as dataset B. The fault status of dataset A and dataset B include: normal (N), outer race fault (OF), inner race fault (IF) and rolling fault (RF). And the number of samples of dataset A and datasets B are 1600 and 1500 respectively, and the length of a single sample is 1044. In the failure experiment, the motor speed is 1800rpm, and the data frequency is 5kHz. The specific information is shown in Table 1.

| Dataset | Bearing location | Health state | Rotating speed | No. of samples | Load | Sampling frequency | Label |
|---------|------------------|--------------|----------------|----------------|------|--------------------|-------|
| A       | underhang        | N/OF/OR/RF   | 1800 rpm       | 1600           | 6g   | 5 kHz              | 1/2/3/4 |
| B       | overhang         | N/OF/OR/RF   | 1800 rpm       | 1500           | 20g  | 5 kHz              | 1/2/3/4 |

3.2. Experimental parameters

Before conducting the experiment, Network structure and the hyper-parameter of the model are set.

The network structure with setting mainly includes: feature extractor network, domain identification network and fault feature recognition network. The feature extractor network consists of a total of 8 layers of networks: including an input layer, two convolutional layers, two pooling layers, a Relu activation layer and a dropout layer. The domain discriminator network has the same structure as the fault feature identification network, which consists of a four-layer network. The above two
networks have a four-layer network structure: including 2 fully connected layers, a Relu layer and a Softmax layer.

The hyper-parameter that needs to be set includes the trade-off parameter \( \lambda \) in Equation (9) and the total number of pixel areas \( m \) in Section 2.2. And other hyper-parameters use default parameters. The hyper-parameter \( \lambda \) is mainly used to weigh the error loss of the domain discriminator network and the fault feature identification network; the hyper-parameter \( m \) is mainly used for the calculation cost and training accuracy of the method. The experimental results are shown in Figure 3 and Figure 4 respectively. The experimental results show that when \( \lambda = 0.6 \) and \( m = 25 \), the model has the best training results and the most suitable computational cost.

### 3.3. Experimental comparative analysis

In order to illustrate the rationality of the proposed method, the proposed method is compared with the other four typical methods. The four typical methods are: Intelligent fault diagnosis model based on convolutional neural network (CNN) [12], a deep domain confusion diagnosis method based on the multi-core maximum mean difference method (MDDC) [13], a domain adversarial model based on the adversarial neural network (ADDA) [14] and Multi-domain adaptation model (MADA) [15]. The diagnosis result of the above method is shown in Table 2.

| Experimental task | CNN   | MDDC  | ADDA  | MADA  | Proposed |
|-------------------|-------|-------|-------|-------|----------|
| A \( \rightarrow \) B | 78.4% | 85.8% | 89.5% | 91.4% | 96.7% |
| B \( \rightarrow \) A | 80.2% | 84.8% | 88.3% | 90.5% | 95.8% |
| Average           | 79.3% | 85.3% | 88.9% | 90.9% | 96.3% |

The results in Table 2 show that the CNN model does not use the transfer diagnosis method, thus the accuracy rate is the lowest, and the final average accuracy rate is 79.3%. The average accuracy rate of the MDDC model for the two sets of experiments can reach 85.3%. The average accuracy of the ADDA model and the MADA model can reach 88.9% and 90.9% respectively. The final accuracy of the method proposed in this paper can reach 96.3%, which is greatly improved compared to other methods in fault recognition accuracy.

### 4. Conclusions

In order to solve the application problem of intelligent fault diagnosis technology in different bearings equipment, a deep transfer attention network for intelligent fault diagnosis of rolling bearings is proposed. Firstly, this method solves the problem of local transferability for bearing datasets, and improves the diagnostic accuracy of the transfer attention network for intelligent fault diagnosis on
different bearings datasets. Secondly, the effectiveness of the proposed method is verified by two sets of test tasks and control experiments.

The method proposed in this paper mainly solves the problem of fault diagnosis between different rolling bearings equipment. However, when the number of fault status categories in the target domain is less than the number of fault categories in the source domain, new methods need to be further explored.

Acknowledgements
This work was supported by the Natural Science Basic Research Plan in Shaanxi Province of China (No. 2020JQ-365), the Fundamental Research Funds for the Central Universities of CHD (No. 300102250302), and the China Postdoctoral Science Foundation (No. 2020M683393).

References
[1] Pan S J and Yang Q 2010 A Survey on Transfer Learning IEEE Transactions on Knowledge and Data Engineering 22(10) p. 1345-1359
[2] Wu Y, Jiang B and Lu N 2019 A Descriptor System Approach for Estimation of Incipient Faults With Application to High-Speed Railway Traction Devices IEEE Transactions on Systems Man Cybernetics-Systems 49(10) p. 2108-2118
[3] Wu Y K, et al. 2017 Multiple incipient sensor faults diagnosis with application to high-speed railway traction devices Isa Transactions 67 p. 183-192
[4] Guo L, et al. 2019 Deep Convolutional Transfer Learning Network: A new method for intelligent fault diagnosis of machines with unlabeled data IEEE Transactions on Industrial Electronics 66(9) p. 7316-7325
[5] Shao J, Huang Z and Zhu J 2020 Transfer Learning Method Based on Adversarial Domain Adaptation for Bearing Fault Diagnosis IEEE Access 8 p. 119421-119430
[6] Zhu J, Chen N and Shen C 2020 A New Deep Transfer Learning Method for Bearing Fault Diagnosis Under Different Working Conditions IEEE Sensors Journal 20(15) p. 8394-8402
[7] Lei Y, et al. 2020 Applications of machine learning to machine fault diagnosis: A review and roadmap Mechanical Systems and Signal Processing 138
[8] Wang X, et al. 2019 Transferable Attention for Domain Adaptation Proceedings of the AAAI Conference on Artificial Intelligence 33 p. 5345-5352
[9] Tan C, et al. 2018 A Survey on Deep Transfer Learning, in Artificial Neural Networks and Machine Learning - Icann 2018, Pt iii, V. Kurkova, et al., Editors. p. 270-279
[10] Vaswani A, et al. 2017 Attention Is All You Need. arXiv.
[11] Marins M A, et al. 2018 Improved similarity-based modeling for the classification of rotating-machine failures Journal of the Franklin Institute-Engineering and Applied Mathematics 355(4) p. 1913-1930
[12] Tang S, Yuan S and Zhu Y 2020 Convolutional Neural Network in Intelligent Fault Diagnosis Toward Rotatory Machinery IEEE Access 8 p. 86510-86519
[13] Long M and Wang J 2015 Learning Transferable Features with Deep Adaptation Networks
[14] Ghifary M, Kleijn W B and Zhang M 2014 Domain Adaptive Neural Networks for Object Recognition in Pricai 2014: Trends in Artificial Intelligence, D.N. Pham and S.B. Park, Editors. p. 898-904
[15] Zhongyi P, et al. 2018 Multi-Adversarial Domain Adaptation arXiv. arXiv, p. 9 pp.-9 pp.