Intelligent Fault Diagnosis of Satellite Communication Antenna via a Novel Meta-learning Network Combining with Attention Mechanism

Y H Chang¹, J L Chen¹ and S L He²

¹State Key Lab. For Manufacturing Systems Engineering, Xi’an Jiaotong University, Xi’an, 710049, China
²School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin, 541004, China

E-mail: jlstri@163.com (J L Chen)

Abstract. Shipborne satellite communication antenna which is used for remote control plays an irreplaceable role in ships, it is necessary to monitor its operation state. However, obtaining sufficient fault information in mass monitoring data is particularly difficult, which greatly degrades performance of existing intelligent algorithms. In this paper, a novel meta-learning network is proposed to realize state recognition of shipborne antenna under small samples prerequisite. The network is constructed to improve generalization even though inputs collected under different operating conditions. Meta-learning network consists of sampler, feature extractor, auxiliary classifier and discriminator. It trains an adaptive pseudo-distance to evaluate the degree of correlation between different data, then realize classification task. Feasibility and effectiveness of the network are verified by three bearing datasets. Results show that the proposed method uses few samples to successfully classify mechanical data of shipborne antenna even with different rotating speed and random noise.

1. Introduction

Shipborne antenna is a typical and complex mechanical system which determines the effect of communication, target tracking, measurement and control [1]. As a key component in shipborne antenna, the failure probability of rolling bearing will be aggravated by long period of high load operation, which may cause the whole mechanical system stop operating. Considering the harsh working environment and mass of data, traditional diagnosis method is difficult to effectively detect faults in bearings. Currently, intelligent fault diagnosis methods have gradually become the mainstream in the field of fault diagnosis [2-3]. However, the premise of the success of intelligent diagnosis methods is that they rely on a large amount of labeled training data, which is contrary to the reality. Because most of the data recorded by device are useless health data, few of them contain wealthy fault information. Consequently, small samples problem has seriously hindered the development of intelligent diagnostic methods [4].

As a further development of deep learning, meta-learning [5-6] has gradually received much attention. The purpose behind meta-learning is to enable the network have the ability to learn independently to cope with small samples problem. In this paper, a novel meta-learning network (MN) is proposed to overcome small sample difficulties and apply it to fault classification. MN is composed
of four network substructures. Firstly, the training dataset is divided into multiple training subsets by random sampler. Next, the feature extractor is used to extract sensitive features from mechanical data. Afterwards, the application of the attention mechanism [7-8] helps network filter the extracted features and select the discriminative feature information. Finally, the discriminator implements the classification task. MN trains an adaptive pseudo-distance to evaluate the degree of correlation between different operating state data, thereby realizing the recognition of operating state of mechanical equipment.

2. Training data setup
In this section, the Case Western Reserve University (CWRU) bearing datasets are introduced as training data. Then, the operation of random sampler is described to construct the training subsets.

2.1. Introduction of the CWRU dataset
The goal of the MN is to use enough labeled data from the laboratory to train network and apply it to other scenarios. The Case Western Reserve University (CWRU) bearing datasets are widely used as common datasets for fault diagnosis of bearings [9]. The investigation of the MN would be carried out with the CWRU. Further details can be found at the CWRU Bearing Data Center website [9].

2.2. Training subsets setup
The CWRU datasets which collected under 30Hz rotating speed were used to construct training dataset. Four health conditions were labeled as normal condition (NC), inner raceway fault (IF), roller fault (RF), and outer raceway fault (OF). A collected signal was segmented from a random start position to create a sample, which contains 2048 data points. So each category of fault contains 400 samples, and a total of 1600 training samples are obtained. As shown in [10], spectrogram features of vibration signal are shown to be superior for neural network. Therefore, the raw wave signals are transformed into the time-frequency domain by using Short-Time Fourier Transform. Mathematically, it can be described as follows.

$$STFT(k, m) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}m}$$

Where $x(n)$ is a waveform of a sample. $\omega(n)$ represents the window function. Furthermore, 5 samples were regarded as a small-samples dataset, every 5 samples were sampled as a training subset, while every 10 samples were sampled as a validation subset. The process is done by random sampler.

3. Proposed method
The proposed method will be presented in this section. The overall architecture of the MN is depicted in Fig. 1.

3.1. Feature extractor based on convolutional framework
The feature extractor is composed of six layers, including four convolutional layers and two maximum pooling layers, wherein two maximum pooling layers are connected to the first two convolutional layers respectively. Kernel number of these convolutional layers is 64. The size of kernel is set as $3 \times 3$. The convolution operation can be defined as follows.

$$Y_j^i = f(X_j^i) = f \left( \sum_{l} X^{i-1}_{j} \otimes w_j^l + b_j^l \right)$$

Where $Y_j^i$ is an intermediate result of convolutional layer. $\otimes$ represents a convolutional operation. $w_j^l$ is referred to a convolutional filter. $b_j^l$ is the additive bias. $f$ is a nonlinear active function. After that, the characteristics of heterogeneous training subsets extracted by feature extractor are concatenated with those of validation subsets. The concatenate result can be represented as follows.
Where $Z_i$ is the synthetic feature vector. $f(X_i)$ is the feature of training subset. $f(X_j)$ is the feature of validation subset. $\text{Concatenate}(\cdot, \cdot, \cdot)$ is the concatenate operation.

**Figure 1. Overview of the Meta-learning Network.**

### 3.2. Auxiliary classifier with attention mechanism

The auxiliary classifier consists of four convolutional layers, two average pooling layers, and two maximum pooling layers. Kernel number of these convolutional layers is 64. The size of kernel is set as $3 \times 3$. The function of auxiliary classifier with attention mechanism mainly includes two parts. The first part is channel attention, the other part is spatial attention. The calculation formula is as follows.

$$Z_c = M_c(Z) \odot Z$$ \hfill (4)

$$Z_s = M_s(Z) \odot Z$$ \hfill (5)

Where $Z$ represents the input from feature extractor. $\odot$ is the convolutional operation. $M_c$ and $M_s$ can be expressed as follows.

$$M_c(Z) = \sigma(\text{MLP}(\text{AvgPool}(Z)) + \text{MLP}(\text{MaxPool}(Z))) = \sigma(W_i(Z_{c_{avg}}) + W_i(Z_{c_{max}}))$$ \hfill (6)

$$M_s(Z) = \sigma(f^{3 \times 3}([\text{AvgPool}(Z) \odot \text{MaxPool}(Z)])) = \sigma(f^{3 \times 3}([Z_{s_{avg}} \odot Z_{s_{max}}]))$$ \hfill (7)

Where $M_c$ represents the channel attention operation. $M_s$ represents the spatial attention operation. $\sigma$ is a attention parameter. $W$ is the weight matrix. $\odot$ is the convolutional operation. AvgPool is an average pooling operation. MaxPool is a maximum pooling operation.

### 3.3. Discriminator based on convolutional framework

The discriminator is composed of two convolutional layers, two maximum pooling layers and two full-connection layers. Kernel number is set as 128 and 64, and the size of kernel is $3 \times 3$. The discriminator attempts to train an adaptive pseudo-distance to evaluate the degree of correlation between different samples, and then make a prediction. The discriminator is calculated as follows.
\[ r_{ij} = g_{\phi}(C(f_{\phi}(Z_i), f_{\phi}(Z_j))), \quad i, j = 1, 2, ..., K \]  

(8)

Where \( r_{ij} \) is the correlation score. \( f_{\phi}(Z) \) represents the auxiliary classifier. \( g_{\phi} \) represents the discriminator. \( K \) is the total number of categories in training and validation subsets. The loss function \( L \) of this network is defined by using minimum Mean Square Error (MSE) function.

\[
\phi \leftarrow L = \arg \min_{\phi} \sum_{i=1}^{m} \sum_{j=1}^{n} (r_{ij} - 1(\hat{Z}_i = \hat{Z}_j))^2.
\]  

(9)

Where \( \hat{Z}_i \) represents the real label, and \( \hat{Z}_j \) represents the predicted label. Therefore, the error of each layer can be calculated by the back-propagation algorithm, the objective is to minimize the error by reducing the contributions of the network parameters.

4. Case study
Two heterogeneous cases are analyzed in this section. Three currently excellent intelligent methods were used to compare with the proposed method during case study, including convolutional neural network (CNN), transfer component analysis (TCA) and transfer learning network (DaNN) [11].

4.1. Case 1: SQ dataset
A simulated fault test was carried out on the integrated mechanical fault simulation test rig developed by SpectraQuest Company. The test rig is shown in Fig. 2. Two fault bearings were manufactured with IF and OF, as shown in Fig. 2. Each bearing data was collected under three rotating frequencies (approximately 10, 20 and 30 Hz) by a tachometer.

![Figure 2. Graph of the SQ integrated mechanical fault simulation test rig and photographs of the manufactured bearings. Top: Inner raceway defect. Bottom: Outer raceway defect.](image)

Since the CWRU datasets were acquired under 30Hz rotating speed, the SQ datasets under same rotating speed were firstly used for experiments. 155 samples of each bearing were formed and transformed into spectrograms. So total number of spectrogram datasets is 465 with three labels (NC, IF, and OF).

After 2000 loops trained by CWRU datasets, the MN gained 100% testing accuracy for SQ data. And the comparison results are given in Table 1. By comparison, it can be found that the proposed method is more superior than current effective method in the face of small samples.

| Method  | Mean accuracy (%) |
|---------|------------------|
| CNN     | 91.1             |
| TCA     | 65.9             |
| DaNN    | 93.0             |
| MN      | 100.0            |

Table 1. Classification results of SQ Datasets by 10-fold cross validation

In addition, experiments on mixed rotating speed were carried out. The same method as above was adopted, the comparison results were also given as shown in Table 2. Though observing comparison...
results, TCA obtains the lowest accuracy of 46.7%. It implies that neural network adaptive extraction of deep features in data has great advantages over artificial feature indexes. Further, comparing results of DaNN and MN, MN gets the highest accuracy of 95.1%, which is slightly ahead. However, MN does not need to be fine-tuned for cross-domain testing like transfer learning, which is the superiority of this approach.

4.2. Case 2: Shipborne Antenna Dataset
In above case, the test rig is much different from the transmission system of real ship, so a full-size high-precision testing platform was built to simulate common operation of shipborne antenna as shown in Fig. 3. Bearings with a RF and a OF were manufactured and installed into the test rig, which is shown in Fig. 3. The testing data is acquired under 600 rpm, 1200 rpm and 1500 rpm.

Similarly, 205 samples from collected signals of each bearing under same speed were formed. So there are totally 1845 samples. Firstly, a cross-domain small samples experiment was carried out under a single rotating speed. Considering the complexity of Shipborne Antenna Datasets, the MN was trained 5000 loops to obtain the optimal weight. The results of comparison with other methods by cross validation are presented in Table 2. It can be found that the testing accuracy of case 2 decreased by 11.8% compared with case 1, which is due to the difference between rotating speeds of training data and testing data. (CWRU-30Hz, Shipborne Antenna-25Hz).

![Image of experiment platform](image-url)

**Figure 3.** Right: Photo of experiment platform. Middle: Experiment setup of the test rig. Left: photographs of the manufactured bearings. Top: Roller defect. Bottom: Outer raceway defect.

| Method | Single rotating speed | Mixed rotating speed |
|--------|-----------------------|----------------------|
| CNN    | 69.5                  | 49.2                 |
| TCA    | 49.2                  | 34.9                 |
| DaNN   | 76.0                  | 53.0                 |
| MN     | 88.2                  | 66.0                 |

The MN was used to test mixed data with rotating speed of 10Hz, 20Hz and 25Hz on the basis of previous experiments. The implementation process is the same as the single-speed experiment. The final testing results and comparison results are also given in Table 2. Results show that the testing accuracy of four methods has dropped significantly, which is due to the fact that complex transmission paths and mass of background noise drown out effective fault information. Training data of small samples leads to the network not learning enough diversity of knowledge, which is also one of the reasons for the low accuracy. However, the proposed method is still far ahead of other comparative methods, and it also proves its effectiveness and superiority in sense.
5. Conclusion
In this paper, a novel meta-learning network was proposed for fault classification of noisy vibration signals under small samples prerequisite. This method starts from laboratory common datasets for training, extracts the characteristics with good robustness and sensitivity adaptively, and finally learns the knowledge with strong generalization for classification. The MN trains a pseudo-distance to evaluate the correlation degree between different labeled data, then the trained network can be directly applied to other scenarios. Three different types of bearing datasets are used to verify its outstanding performance and generalization ability. The results show the proposed method is capable for classification under small samples prerequisite.

6. Acknowledgements
This research is supported financially by the National Key Research and Development Program of China (No.2019YFF0302204), National Natural Science Foundation of China (No. 51875436), National Natural Science Foundation of China – Civil Aviation Joint Research Fund of China Civil Aviation Administration (No. U1933101), China Postdoctoral Science Foundation (No. 2018M631145) and Shaanxi Natural Science Foundation (No. 2019JM-041). The authors would like to sincerely thank all the anonymous reviewers for the valuable comments that greatly helped to improve the manuscript.

7. References
[1] J. Chen, C. Zhang, X. Zhang, et al. "Planetary gearbox condition monitoring of ship-based satellite communication antennas using ensemble multi-wavelet analysis method," Mech. Syst. Signal Process., vol. 54-55, pp. 277-292, 2015.
[2] Asghar Dehghani, J. Zheng. “A new method to generate virtual samples for solving small sample set problems”, in: 2011 10th Information Conference on Machine Learning and Applications and Workshops, Honolulu, HI, USA, 2011.
[3] H.O.A. Ahmed, M.L.D. Wong, A.K. Nandi, “Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features”, Mech. Syst. Signal Process. 99 (2018) 459–477.
[4] Asghar Dehghani, J. Zheng. “A new method to generate virtual samples for solving small sample set problems”, in: 2011 10th Information Conference on Machine Learning and Applications and Workshops, Honolulu, HI, USA, 2011.
[5] Z. Xu, L. Cao, X. Chen. “Learning to Learn: Hierarchical Meta-Critic Networks”, IEEE Access. vol. 7, pp. 57069-57077, 2019.
[6] L. Yang, et al. “The effect of correlation on the accuracy of meta-learning approach”, in: Fifth IEEE International Conference on Advanced Learning Technologies, Taiwan, 2005.
[7] Q. Zhao, X. Cai, C. Chen et al. “Commented content classification with deep neural network based on attention mechanism”, in: 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 2017.
[8] H. Li, J. Li. “Recognition of robot based on attention mechanism and convolutional neural network”, in: 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019.
[9] CWRU Bearing Data Center Website <http://csegroups.case.edu/bearing data center/home >
[10] L. Deng, et al. “Binary coding of speech spectrograms using a deep auto-encoder,” in Eleventh Annual Conference of the International Speech Communication Association, 2010.
[11] Yaroslav Ganin, Victor Lempitsky et al. “Domain-adversarial training of neural networks”, Journal of Machine Learning Research, vol. 17, pp. 2096-2030, 2016.