2022 21st International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Intelligent judgment of rotating machinery based on multi-scale parallel network and attention mechanism

Zhixiang Fan
Jiangnan University
School of Artificial Intelligence and Computer Science
Wuxi,China
1486276099@qq.com

Pengjiang Qian
Jiangnan University
School of Artificial Intelligence and Computer Science
Wuxi,China
qianpjiang@jiangnan.edu.cn

***Abstract*—The primary problem solved in rotating machinery fault diagnosis is how to effectively extract fault features from the vibration signals with noise. To extract fault features accurately, this study proposes a multi-scale parallel convolutional neural network fault recognition algorithm, which can carry out feature fusion. The above method combines empirical feature extraction (e.g., fast Fourier transform) to enrich feature information, which can effectively implement deep learning. The effectiveness and reliability of the method are verified through example studies on JNU, SEU and PU rolling bearing experimental data sets. The algorithm has the higher classification capability and diagnostic accuracy compared with four common deep learning algorithms.**

Keywords-component; Mechanical intelligent diagnosis; Deep neural network; Multi-scale module; Attention mechanism.

I. INTRODUCTION

As the structure of modern industrial equipment becomes more and more complex and sophisticated, the reliability of machinery and equipment decreases, maintenance failure cycles become longer, and the losses caused by failures are greater. Thus, it is important to anticipate equipment failure as early as possible and carry out equipment maintenance in time.

Rotating machinery fault diagnosis technology is a science and technology to monitor, diagnose and predict the status and fault of continuously operating rotating machinery and equipment to ensure that the rotating machinery and equipment operate safely.

Its specific research content includes the study of fault mechanism, how to obtain information, vibration signal analysis and processing methods, pattern recognition methods, fault monitoring and diagnosis system development research.

This study focuses on the research of pattern recognition method. Extracting the fault characteristic parameters, displaying the effective information, accurately judging the working status of the equipment at this time, and giving the correct maintenance plan for the cause of the fault.

As computer computing power is advancing, and deep learning techniques are leaping forward in numerous fields, including support vector machines (SVM), convolutional neural networks (CNN), stacked self-coding and deep belief nets (DBN) are increasingly implemented in rotating machinery fault diagnosis. With the support of the above algorithms, deep learning has become an attractive approach and at the same time avoids subjective feature selection and human interpretation [1]. Zhang et al. used support vector machines (SVM) to identify the state of rolling bearings in real time [2], and Weimer et al. used convolutional neural networks (CNN) to overcome the need to redefine manually each novel situation in the production process, improving automation of equipment and accuracy of detection [3]. Verstraete et al. optimized the structure of the CNN model through the combination of CNN with generative adversarial network (GAN) to generate more labeled samples to solve part of the data set. Due to the small amount of data, the neural network could not extract feature information completely and accurately [4]. The AE algorithm model contains two main parts, including Encoder (encoder) and Decoder (decoder). The improved models are Stacked Self-Encoder (SAE) and Variational Self-Encoder (VAE), etc.

For validity and fairness of the experiments, all experiments were conducted in the benchmark codebase proposed by Zhao et al. which contains nine publicly available data sets and nine commonly used deep learning models [5], including AE, DAE, SAE, MLP, CNN, AlexNet [6], ResNet18[7] and LSTM [8] models. The results show the significant advantages of DL in the field of mechanical diagnosis.

The main goal of the algorithm in this study is to enrich the bearing fault feature information and increase the accuracy of equipment fault diagnosis through the extraction of multiple feature information (frequency domain information and time-frequency domain information).

In this study, we use JNU bearing data set, Pu bearing data set and SEU transmission data set.

II. RELATED WORK

A. Attentional Mechanism

The attention mechanism has been studied as early as the 1990s, and was applied to the field of vision in Volodymyr's paper “Recurrent Models of Visual Attention” in 2014, and later accompanied by Ashish Vaswani's “Attention is all you need” in 2017. The attention mechanism has been widely used in the design of networks for NLP and CV related problems with the proposal of Transformer model.

For the respective input item of the model, probably a different part of a picture or a word of a statement, a weight is set, and the size of this weight denotes the level of attention the model should pay to that part. On that basis, the weight size is adopted for the simulation of the focus of human attention during the processing of the information, thus significantly reducing the computational effort and enhancing the performance of the model.

2473-3636/22/$31.00 ©2022 IEEE
DOI 10.1109/DCABES57229.2022.00040

B. CoNet Attentional Mechanism

The Self-Attention structure in the original Transformer only calculates the attention matrix based on the interaction between query and key, thus ignoring the connection between adjacent key.

Therefore, Qilong Wang et al. proposed the CoT block structure[9]. First, a 3*3 convolution is applied to the Key to model static context information. Subsequently, the Key after modeling Query and context information is concatenated, and then two consecutive 1*1 convolutions are adopted to generate dynamic context with self-attention. The static and dynamic context information is finally fused into the output.

First, three variables Value, Query and Key are defined (Value is the mapping of X features, Query and Key use the original X values). First, a k*k group convolution is performed on Key to obtain the local context information K^1, which can be seen as a static modeling on the local information, and then merged with Query, and then two successive convolution operations are performed on the merged results, (1) is as follows.

$$A = [K^1, Q]W_\theta W_\delta \tag{1}$$

The performance of the self-attentive mechanism is enhanced by the bootstrap of local context modeling.

III. Our Proposed Method

A. Multi-scale module

To improve the ability of the network to extract feature information and the accuracy of fault judgment, we carefully designed the neural network modules shown in Figures.1,2. By different perceptual fields (convolutional kernel size) and different order of convolutional kernel operations, the network obtains more different feature information.

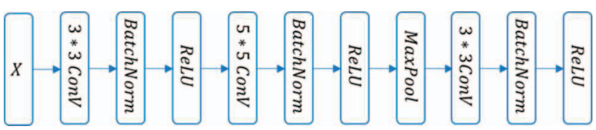

Figure 1. Multiscale Module 1

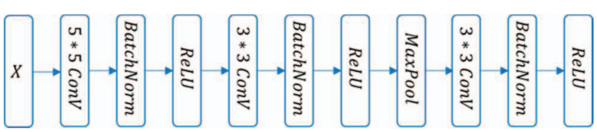

Figure 2. Multiscale Module 2

B. Network Model

Recent studies have shown that parallel convolutional neural networks are effective in certain applications [10]. Therefore, we propose the network model shown in Figure.3.

The network is composed of two parallel branches, each of which performs convolutional extraction. In one branch, a fast Fourier transform operation (2) is first performed on the training data slices, and the frequency domain feature information (FFT) is extracted, and in the other branch, a short-time Fourier transform operation (3 and 4) is first performed on the training data slices, and the time-frequency domain feature information (STFT) is extracted. Subsequently, both branches perform convolution operations by parallel multiscale modules for both frequency and time-frequency domain. The frequency domain information is processed twice in parallel analysis. Next, the features obtained from the convolution analysis are fused together to explore the rich context through the CoTNet attention mechanism module, and finally the obtained results are used as the input to the classifier.

In the last step, the final classification results are obtained by the powerful Resnet18 classification network.

$$x_i^{FFT} = FFT(x_i) \quad i = 1,2,\cdots \tag{2}$$

$$STFT(t,f) = \int_{-\infty}^{\infty} x(\tau)h(\tau - t)e^{-j2\pi f\tau}d\tau \tag{3}$$

$$x_i^{STFT} = STFT(x_i) \quad i = 1,2,\cdots \tag{4}$$

where FFT(·) and STFT(·) operators represent the transformation x_i into the frequency and time-frequency domains, respectively. The h(τ-t) in (3) is the analysis window function, using a Hanning window with a window length of 64.

Multiple convolution operations are performed in the two branch networks. As the network deepens, the number of convolution kernels doubles that of the previous layer to prevent overfitting after successive convolutions, and we use the ReLU function as the activation function after the respective convolution operation and the parameter for feature mapping reduction by the maximum pooling method.

$$ReLU(x) = \max(0, \mathrm{x}) \tag{5}$$

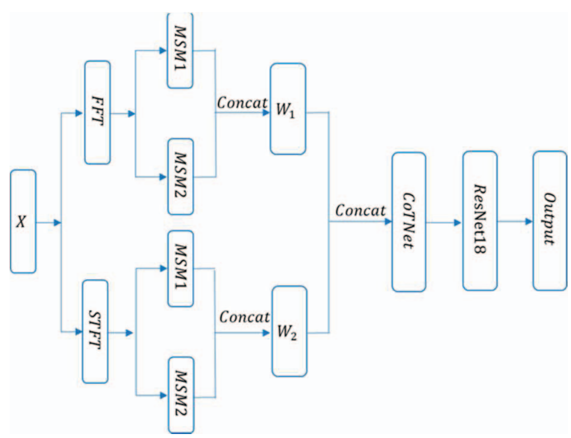

Figure 3. Algorithnm model diagram

IV. Experiment and Discussion

A. Data sets

The The Jiangnan University (JNU) bearing data set was provided by the Bearing Laboratory of Jiangnan University. the JNU data set is composed of three bearing vibration data sets at different speeds, 600 rpm, 800 rpm and 1000 rpm, with a data acquisition frequency of 50 kHz. each speed data set also includes one normal mode and three failure modes (inner ring failure, outer ring failure and rolling element failure). Thus, the total number of classifications is 12, depending on the operating conditions.

The Paderborn University (PU) data set was provided by the Paderborn University Bearing Data Center, and the PU data set consists of 32 sub-data sets. The bearings were (1) 6 undamaged bearings; (2) 12 artificially damaged bearings; and (3) 14 bearings that were actually damaged due to accelerated life tests. Considering the server performance, only 13 sub-data sets were used in the experiment.

The Southeast University (SEU) gearbox data set was provided by SEU. the SEU data set contains both the bearing data set and the gear data set. Both sub-data sets were obtained under two operating conditions, 20Hz-0V and 30Hz-2V (RS-LC), respectively. As depicted in Table III, the SEU data set has 20 fault categories, and each data set file is limited by the server performance to obtain only the 2nd channel data out of 8 channels to reduce the data volume.

B. Data Enhancement

We enhance the input signal by two data enhancement strategies. The first one is to randomly multiply the input signal by a random factor with the following equation:

$$x^* = \beta * x \tag{6}$$

where x is the input signal and β is a scalar that follows the $N(1,0.01)$distribution.

The second one is a random coverage of part of the signal. The formula is as follows:

$$x^* = mask * x \tag{7}$$

where mask is the binary sequence whose random position subsequence is zero.

C. Training Details

Adam optimizer was employed in the network training, in which several training parameters were given, including batch size=16 as well as learning rate=0.01. The respective model was trained and examined in the training alternately in terms of 100 epochs. The present algorithm and all comparison algorithms were trained. The comparison of these algorithms was drawn on the basis of the open-source framework developed by Zhao et al. The method of data augmentation presented in the open-source framework was employed in the data augmentation strategy of the comparison algorithm.

D. Evaluation Indicators

$$Acc = \frac{Sum_y}{Sum_t} \tag{8}$$

where Sum_y enotes the number of correctly classified samples; Sum_t epresents the total number of samples.

To avoid the influence of network model fluctuations during the training process, the respective experiment was repeated five times. Lastly, the average Acc is used as the evaluation index.

E. Experimental results

Under the same conditions, each algorithm five times were tested based on different data sets. Figure.4 presents the average predicted Acc.

From the overall view of Figure.4, our proposed algorithm achieves a first-class level in the diagnosis of rotating machinery faults.

The Resnet18 algorithm performs much better than other comparative algorithms (CNN, LSTM, AlexNet) on the three experimental data sets, which proves that Resnet18 has strong parsing ability on rotated data.

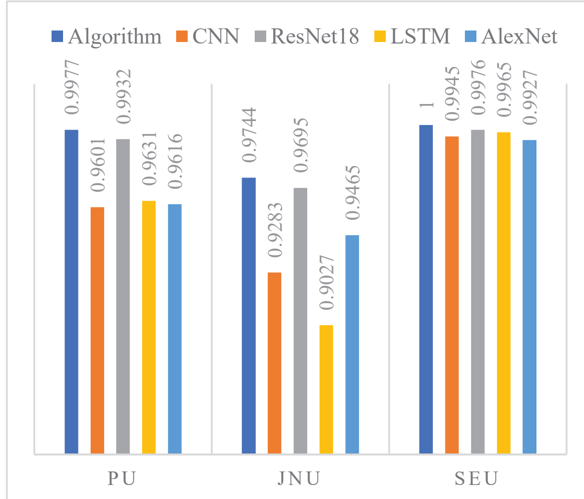

Figure 4. Average prediction accuracy of 5 algorithms on 3 data sets

The algorithm in this study uses Resnet18 as a classifier, and from Figure.4, the accuracy of this study's algorithm is higher than Resnet18 on mechanical fault diagnosis. On that basis, it is proven that the multi-scale parallel module can better extract feature information from the original data and suppress noise.

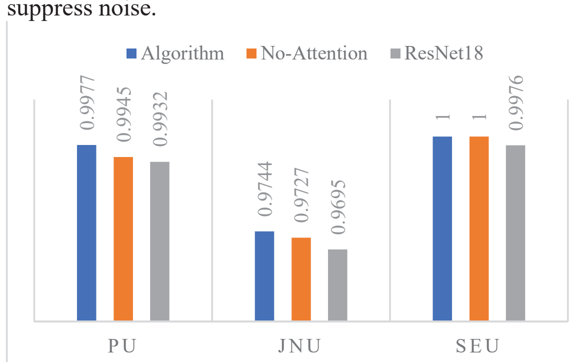

Figure 5. Graph of ablation experiment results

To better demonstrate the superiority of multi-scale parallel networks, a set of ablation experiments were conducted, and the algorithm of this study and the algorithm of this study with the CoTNet attention mechanism removed were compared. The result graph in Figure. 5 indicates that the algorithm with the CoTNet attention mechanism removed outperforms the Resnet18 algorithm and inferior to the algorithm in this study. It proves the superiority of multi-scale parallel networks and the ability of the CoTNet attention mechanism to enhance the performance of deep convolutional networks.

V. CONCLUSION

In this study, a novel network model is built based on typical deep convolutional neural network structures in combination with multiscale analysis techniques. As revealed by the experimental results in a number of data sets, the multiscale parallel network of this study shows a classification level that is more statistically significant than other networks in terms of intelligent diagnosis of rotating machinery.

REFERENCES

[1] Gan M, Wang C, Zhu C (2018) Fault feature enhancement for rotating machinery based on quality factor analysis and manifold learning. J Intell Manuf 29(2):463–480.

[2] X.L. Zhang, B.J. Wang, X.F. Chen, Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine, Knowl.-Based Syst. 89 (2015) 56–85 .

[3] Weimer D, Scholz-Reiter S, B. (2016) Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. CIRP Ann Manuf Technol 65(1):417–420.

[4] Verstraete DB, Lope Droguett E, Meruane V, Modarres M, Ferrada A. Deep semi-supervised generative adversarial fault diagnostics of rolling element bearings. Struct Health Monitor 2020;19:390–411.

[5] Zhao Z, Li T, Wu J, et al. Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study[J]. ISA Transactions, 2020.

[6] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, pp. 1097–1105.

[7] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

[8] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.

[9] Wang Q , Wu B , Zhu P , et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[10] Cui Z, Chen W, Chen Y (2016) Multi-scale convolutional neural networks for time series classification, arXiv preprint arXiv: 1603.06995.