

Research Progress of Rotating Machinery Fault Diagnosis Based on Deep Learning

Du Zhehua^{1,*}, Lin Xin²

¹Wuhan Second Ship Design and Research Institute, 430205 Wuhan, China

²Hubei Province Engineering Consulting Co., LTD., 430071 Wuhan, China

Abstract. In modern production, the precision and the importance of rotating machinery is higher and higher in the direction of large-scale, high speed and automation development, so that the traditional fault diagnosis methods are insufficient to deal with massive, multi-source and high-dimensional data, cannot meet the requirements of security and reliability. Therefore, several typical deep learning models are briefly introduced at first and the application of deep learning in fault diagnosis of rotor system, gear box and rolling bearing in recent years is studied and analyzed based on its strong feature extraction ability and advantages of clustering analysis. Finally, the advantages and disadvantages of deep learning model are summarized and the fault diagnosis methods of rotating machinery are summarized and prospected based on engineering practice.

1 Introduction

With the continuous increase of high-speed, heavy-duty and automation requirements of contemporary rotating machinery and equipment, when equipments fail, it is very easy to cause great economic losses and even cause major accidents. Therefore, speed, efficiency, and accuracy of fault diagnosis methods have increasingly higher requirements.

The current fault diagnosis technology uses the detected information characteristics to judge the working status of the equipment. Generally, fault diagnosis methods can be divided into the following categories: model-based method, signal-based method, knowledge-based method and combination method. These methods can no longer adapt to the trend of intelligent manufacturing.

In view of powerful data representation learning and analysis capabilities of deep learning technology, fault diagnosis methods based on this technology have attracted widespread attention from all walks of life. Through multi-layer nonlinear network training, the potential characteristics of samples can be learned and the ability of classification or prediction can be improved. Therefore, deep learning is considered as a powerful tool for big data processing and fault diagnosis of mechanical equipment.

2 Typical depth models

2.1 Deep Belief Networks

Deep belief network (DBN) is a probabilistic generative model proposed by Hinton [1]. DBN can be used in unsupervised learning to extract features by reducing the dimension of processed signals. It can also be used as a classifier in supervised learning. DBN consists of several restricted Boltzmann machines (RBM) and Softmax regression layers, as shown in the figure. In the figure, V represents visible layer, h represents hidden layer, and W represents weight.

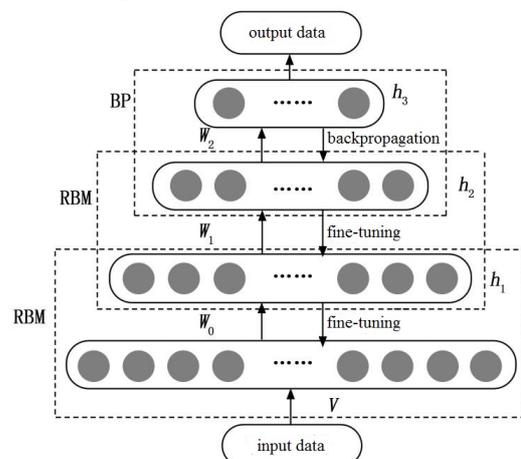


Fig. 1. Schematic diagram of DBN basic structure

The advantage of deep belief network in fault diagnosis of rotating machinery lies in that it can be applied to one dimensional vibration signal data and can

* Corresponding author: shunli878@163.com

achieve good performance even in the case of fewer samples.

2.2 Convolutional neural network

Convolutional neural network (CNN) was proposed by LeCun et al. [2]. It is a feed-forward neural network inspired by the sensory mechanism of animal visual cortex cells. It was first used in large-scale image classification and recognition. The basic network structure combined with rotating machinery fault diagnosis is shown in Fig. 2. It is composed of an input layer, two groups of alternating convolution layer and pooling layer, and full connection layer.

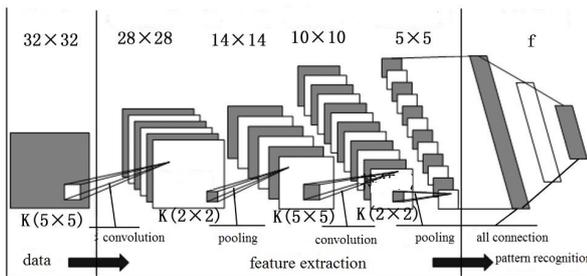


Fig. 2. Schematic diagram of CNN basic structure

Each feature map in the convolutional layer corresponds to a convolution kernel. These convolution kernels convolve the input of previous layer through a set of weights and form a set of characteristic outputs, which become the input of next layer [3]. After two sets of convolution layers and pooling layers, a fully connected layer is added. Fully connected layer is followed by a hidden layer, and finally Softmax regression layer completes classification.

2.3 Recurrent neural network

Recurrent Neural Network (RNN) is a type of deep learning network that processes sequence data, which is widely used in natural language processing. Unlike other neural networks, RNN network also establishes connections between neurons in the layers. Simply put, the output of its neuron can directly affect itself at the next time stamp.

RNN can be regarded as multiple assignments to the same neural network. The input of layer i neuron at time t , in addition to the output of layer $(i-1)$ neuron at that time, also includes its own output at time $(t-1)$. If RNN is expanded according to time point, structure diagram is shown in Figure 3.

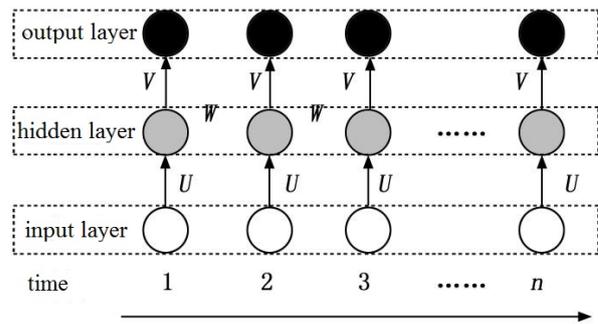


Fig. 3. Schematic diagram of RNN basic structure

2.4 Autoencoder and its variants

2.4.1 Autoencoder

Autoencoder (AE) is divided into two parts: encoder and decoder. It is a typical three-layer neural network structure proposed by Rumelhart [5], including input layer, hidden layer and output layer. Literature [6] proposed that autoencoder network can be regarded as a variation of traditional multi-layer perceptron. The basic idea is to reconstruct original input after passing input signal through a multilayer neural network. The potential structure of input signal is analyzed by unsupervised learning, and the response of middle layer is represented as potential feature. Its basic structure is shown in Figure 4.

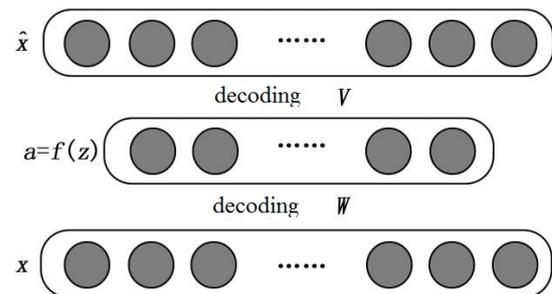


Fig. 4. Schematic diagram of AE basic structure

2.4.2 Denoising autoencoder

Denoising autoencoder (DAE) is proposed by Vincent et al. from the perspective of robustness. In daily research, it is found that the partially occluded or damaged image can still be accurately identified. DAE is inspired by this. By artificially adding some pure Gaussian noise to the original input signal, or randomly discarding a certain feature of the input layer, clean signal is locally damaged, and then the damaged signal is output through traditional self-encoding. Its basic structure is shown in Figure 5.

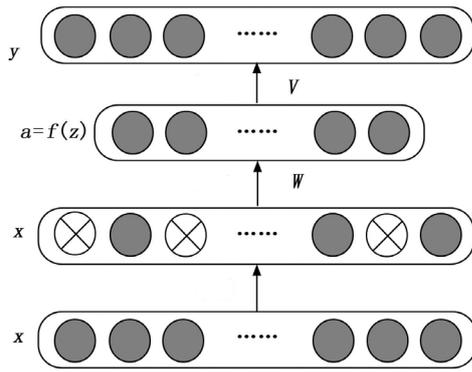


Fig. 5. Schematic diagram of DAE basic structure

2.4.3 Sparse autoencoder

Some researchers introduced the idea of sparse representation into autoencoder and proposed sparse autoencoder (SAE). Autoencoder is originally proposed based on the idea of dimensionality reduction. However, when there are more hidden layer nodes than input nodes, the autoencoder will lose the ability to automatically learn the sample characteristics. At this time, certain constraints on the hidden layer nodes are required. Therefore, each layer of neurons is regularized on the basis of self-encoding. The sparsity constraint function is added to the hidden layer of autoencoder to enhance the sparsity of the output of autoencoder, so that better hidden layer features can be obtained when the number of model neurons is large.

2.5 Generative adversarial network and its variants

Godfellow et al. were inspired by the "zero-sum game" in 2014 and proposed a generative adversarial network (GAN). Compared with the previous deep learning models, they take the distribution difference between data and models as the objective function. GAN model adopts an adversarial form. It first identifies the differences and then gradually narrows them. Network structure of GAN is shown in Figure 6.

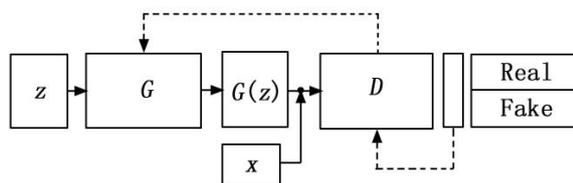


Fig. 6. Schematic diagram of GAN basic structure

The generator and discriminator in GAN can be regarded as two parties in an adversarial game. The goal of generator is to generate the required data, but it should be noted that the data here is only simulated data. The discriminator comes to discriminate the authenticity of generated data.

2.5.1 Stack generative adversarial network

Stacked generative adversarial network (StackGAN) is implemented by cascading multiple GAN models in

structure, which is similar to stacked self-encoding method. The specific method of model construction is to connect two GAN models in series, and the output of upper-level generator is input to lower-level generator as a hidden variable. Its basic structure is shown in Figure 7.

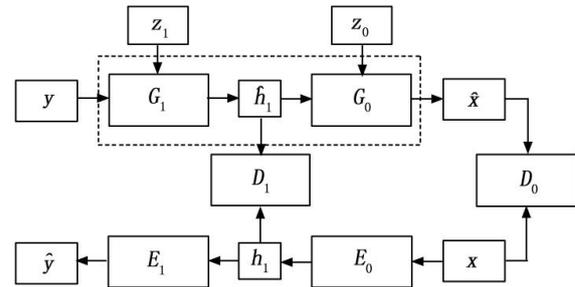


Fig. 7. Schematic diagram of StackGAN basic structure

2.5.2 Deep convolutional adversarial adversarial network

It is a combination of convolutional neural network (CNN) and confrontation network, which integrates the idea of CNN into confrontation model. It replaces fully connected layer in the generator with a deconvolution layer, as shown in Figure 8.

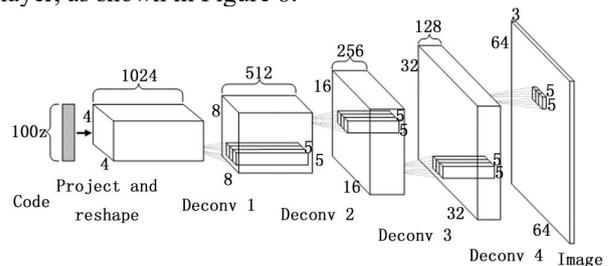


Fig. 8. Schematic diagram of DCGAN basic structure

2.5.3 Conditional generative adversarial network

Conditional generative adversarial network (CGAN) is a kind of GAN with conditional constraints. The specific structure is shown in Figure 9.

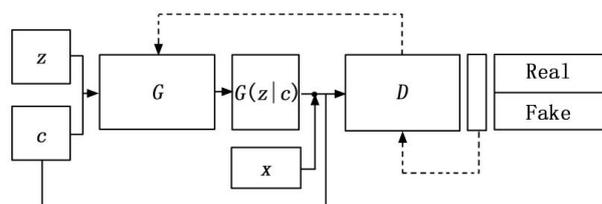


Fig. 9. Schematic diagram of CGGAN basic structure

3 Application in fault diagnosis

3.1 Rolling bearing fault diagnosis

Rolling bearing is the supporting element of various rotating machinery. It not only has a wide range of use, but also has a wide range of varieties and strict requirements. It is an indispensable part of rotating machinery.

3.1.1 The problem of accuracy and efficiency of algorithm training

Related research on this issue has proposed deep belief network (DBN) and particle swarm optimization support vector machine (PSO-SVM) rolling bearing fault diagnosis method. This method not only improves the diagnostic accuracy, but also greatly shortens the training time. There is also a fault diagnosis method based on deep adversarial network under unbalanced data sets. This method not only establishes a deep convolutional neural network model suitable for bearing fault diagnosis, but also introduces Focal loss function to enhance the accuracy of diagnosis.

3.1.2 The problem with fewer fault samples

In order to maintain the accuracy of fault diagnosis and reduce the number of samples, studies have shown that SAE with dynamic adjustment of learning rate can solve this problem to some extent. However, in the above-mentioned traditional supervision methods, model training still requires a large number of manual annotation samples. A new rolling bearing fault diagnosis method based on clustering and laminated autoencoder is proposed. This method not only eliminates the initial steps of manually calibrating the sample, but also makes it more intelligent.

3.2 Gearbox fault diagnosis

Gear vibration signal is nonlinear, non-stationary and has large data volume. Research on gear fault diagnosis method based on ELMD energy entropy and PSO-SAE [7] shows that it can solve such problems. Gearbox has a complicated structure and it is difficult to obtain a signal when it fails. A single sensor cannot effectively obtain the fault vibration characteristics. Research has shown that a gear fault diagnosis method based on information fusion of multiple sensors and stack noise reduction self-encoding [8] has less accuracy attenuation in a noisy environment, and has a strong anti-interference ability.

3.3 Rotor system fault diagnosis

The common forms of rotor system faults are rotor misalignment, rotor unbalance, dynamic and static friction, etc. For permanent magnet synchronous motor turn-to-turn short circuits and permanent magnet loss-of-magnetization faults, there are diagnostic deviation problems caused by factors such as complex data processing, independent and single feature, and insufficient samples. The literature [9] proposed a fault diagnosis method based on deep learning variational

autoencoding network (VAE) and sparse autoencoding network (SAE).

In view of rolling bearing defects in the motor, the relevant research put forward solutions. The innovative discrete wavelet transform (DWT) is used for feature extraction. Then orthogonal fuzzy neighborhood discriminant analysis (OFNDA) is used to reduce features. Finally, the state of components is predicted by dynamic recurrent neural network (RNN).

4 Conclusion

Fault diagnosis method based on deep learning benefits from its advantage of feature learning and strong clustering ability. Intelligent fault diagnosis model established by it is more automatic and efficient. Different deep learning models have different advantages and disadvantages. Researchers have all sorts of ways to improve on the typical deep learning model. Its essence is to give full play to the advantages of each neural network, and to make up for or avoid the shortcomings through various methods.

Deep learning itself has a very rapid development. Various algorithm ideas are fused with each other to complement each other's strengths. In the future, the application of deep learning in the fault diagnosis of rotating machinery will become more efficient and accurate.

References

1. G. Hinton, S. Osindero, Y. Teh, *Neural Computation*, **18**, 1527 (2006)
2. Y. Lecun, L. Bottou, Y. Bengio, *Proceedings of the IEEE*, **86**, 2278 (1998)
3. Y. Lecun, Y. Bengio, G. Hinton, *Nature*, **521**, 436 (2015)
4. P. Tamilselvan, P. Wang, *Reliability Engineering & System Safety*, **115**, 124 (2013)
5. D.E. Rumelhart, G.E. Hinton, R.J. Williams, *Nature*, **323**, 533 (1986)
6. Y. Kamp, H. Bourlard, *Biological Cybernetics*, **59**, 4 (1988)
7. H.M. Zhao, H.L. Liu, J.J. Xu, *Journal of Mechanical Science and Technology*, **33**, 4165 (2019)
8. C.Q. Shen, J.Q. Xie, D. Wang, *Applied Sciences-Basel*, **9**, 3374 (2019)
9. K. Xu, S.M. Li, J.R. Wang, *Proceedings of the Institution of Mechanical Engineers*, **234**, 1167 (2020)