Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 10, No. 3, 2023

Review of Research on Fault Diagnosis of Rolling Bearings Based on Deep Learning

Caidie Duan, Mingchuan Zhang

School of Henan University of Science and Technology, Luoyang 471023, China

Abstract: Deep learning has powerful capabilities in deep feature extraction and expression, and has been successfully applied in equipment fault diagnosis, overcoming the shortcomings of traditional diagnostic methods that rely on expert experience. It can save costs while improving diagnostic accuracy. This article briefly introduces three commonly used neural networks: Deep Belief Networks (DBN), Convolutional Neural Networks (CNN), and Long Short-Term Memory Networks (LSTM), and points out the problems in rolling bearing diagnosis and analyzes future development directions.

Keywords: Rolling bearing; Fault diagnosis; Deep belief networks; Convolutional Neural Networks; Long Short-Term Memory Networks.

1. Introduction

As an important mechanical component in large rotating machinery, bearings play an important role in supporting and transmitting rotational forces. According to statistics, the majority of equipment failures are caused by damage to rolling bearings. If a failure occurs, it can cause abnormal vibration of the equipment, resulting in unstable operating conditions. In severe cases, it can directly damage the equipment and cause huge losses. Therefore, bearing fault diagnosis has attracted widespread attention from researchers.

Traditional bearing fault diagnosis methods have obvious limitations, requiring operators to possess professional knowledge and signal processing techniques. The shallow structure of traditional machine learning algorithms has very limited ability in learning and extracting nonlinear relationships of features. The above limitations limit the further application of traditional machine learning in the field of fault diagnosis. Although the appeal method has achieved certain results, there are problems such as heavy workload in manual feature extraction, heavy reliance on expert experience, and low accuracy.

In recent years, with the development of deep learning technology, the above problems can be solved. Deep learning has been successfully applied in fields such as object detection, speech recognition, and computer vision. This article analyzes the application of three main neural networks of deep learning in fault diagnosis, and points out the challenges and suggestions of deep learning in bearing fault diagnosis.

2. The Development of Deep Learning

Deep learning [1] is a branch in the field of machine learning. It uses multi-layer neural networks to simulate the working mode of human brain, so as to achieve image recognition, natural language processing, target detection and other tasks. Since the 1980s, deep learning has been continuously developing, especially in recent years, with the support of big data and computing resources, deep learning has made significant progress in the field of vision and speech recognition.

The earliest deep learning model is the multilayer perceptron. This model uses a multi-layer network structure and can be trained through a back-propagation algorithm to achieve complex classification tasks. Deep learning algorithms have better ability to approximate complex functions and can effectively implement complex high-dimensional functions. With the rapid development of deep learning, neural networks applied to bearing fault diagnosis at present include: deep confidence network (DBM), convolutional neural network (CNN), long- and short-term memory network (LSTM).

3. Deep Learning

3.1. Bearing Fault Diagnosis Based on Deep Belief Network

Deep Belief Network (DBN) [2] is a deep neural network model based on unsupervised learning. Its structure is similar to that of multiple stacked restricted Boltzmann machines (RBMs). Each RBM is composed of a visible layer and a hidden layer, which are connected by weights. It is a kind of generating stochastic neural network, which can learn the probability distribution of a large amount of data. DBN has good feature learning ability and generalization ability, and can be used for feature extraction and classification tasks in image, voice, natural language processing and other fields. In the field of bearing fault diagnosis, DBN can be used to extract effective features from vibration signals, acoustic emission signals, and multimodal signals for fault diagnosis. The RBM structure is shown in Fig.1, where each RBM is composed of visible and hidden layers. The visible and hidden layers are linked by weights, where vi represents visible nodes, hj represents hidden nodes, and w represents weights. The upper layer is passed layer by layer to the lower layer.

Tao et al. [4] were the first to apply deep confidence networks to the field of rolling bearing fault diagnosis, by constructing encoders to extract fault features, and using the extracted low-dimensional fault features as model inputs, ultimately achieving fault category discrimination. Wang et al. [5] believe that a single vibration signal is not sufficient to characterize non-stationary signals under variable operating conditions, and therefore propose a deep confidence network and multi-sensor feature fusion method. This method uses multiple sensors to simultaneously measure bearing vibration signals at different positions, and then extracts features from the measured signals, which has high diagnostic accuracy for different types of bearing faults. Chen et al. [6] synthesized the advantages of multi-sensor fusion and multi feature fusion and proposed a DBN fault recognition method based on multi feature fusion and multi sensor feature fusion, which further improved the accuracy. Gan et al. [7] constructed a two-layer DBN for rolling bearing fault diagnosis based on the two-stage diagnosis of wavelet packet energy features. The first layer of deep confidence network was used to identify the location of the fault, and the output of the first layer was used as input to the second layer model to determine the degree of damage. By synthesizing the results of the two-layer network model, a very similar task was completed and compared with SVM and BPNN, which not only achieved higher accuracy, and it can also overcome the overlapping problem caused by noise and other disturbances. In order to further improve the accuracy of model classification, methods combining deep confidence networks with other optimization algorithms have been continuously proposed. Shao et al. [8] developed particle swarm optimization to optimize the structure of DBN and applied it to analyze simulation and experimental signals of rolling bearings, achieving more accurate and robust results compared to other intelligent methods. Yang et al. [9] proposed a random adaptive particle swarm optimization algorithm (RSAPSO) to optimize DBN.

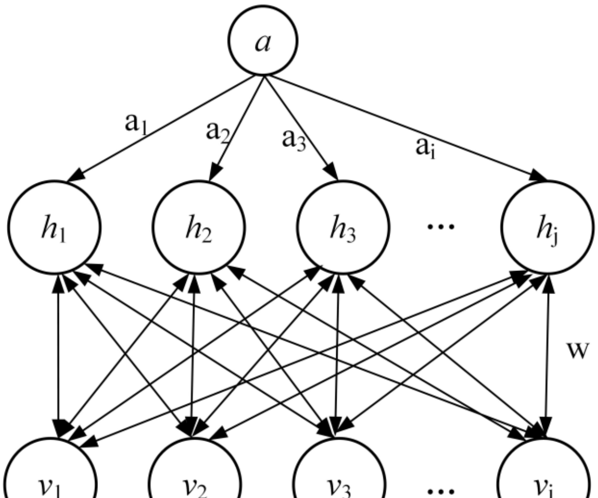

Fig.1 Restricted Boltzmann Machine Network Architecture[3]

3.2. Bearing Fault Diagnosis Based on Convolutional Neural Network

Convolutional Neural Networks (CNN) is a deep feedforward neural network, which is one of the representative algorithms of deep learning, and is widely used in image recognition, natural language processing and other fields. The convolutional module consists of convolutional layer, pooling layer, and fully connected layer. The classic architecture of CNN is shown in Fig.2. The convolutional layer performs convolution operations on data, and the pooling layer compresses the data. There are generally two pooling methods, maximum pooling and average pooling. The fully connected layer combines all local features into global features, which are used to calculate the final score for each class and output it. Convolutional neural networks are composed of multiple stacked convolutional modules, which have the ability to obtain deep level features of data. When classifying, it is necessary to connect the fully connected layers for fault diagnosis.

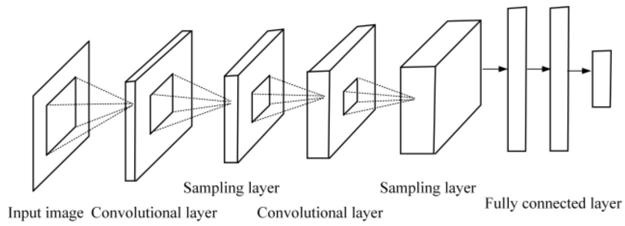

Fig.2 Classic Architecture of Convolutional Neural Networks[10]

The mathematical model of convolutional neural networks is:

$$x_j^l f(\sum_{i\in T_j} x_j^{l-1} \times k_{ij}^l + b_j^l)[10]$$

Where Tj is the input feature map, k is the convolutional kernel, b is the network bias， x_j^l is the output of layer l, x_j^{l-1} is the input of layer l。

Since CNN was first used to identify bearing faults in 2016, there have been many studies based on this type of network model, continuously improving CNN's fault detection and generalization capabilities from various aspects such as data preprocessing, model noise resistance, training speed, and number of parameters. Among them, the one-dimensional CNN model has been widely used due to its fixed layer structure, relatively simple structure, fewer parameters, and relatively high training speed and accuracy. Huang et al. [11] directly used the original vibration signal as the input of the convolutional neural network, and the trained model improved the recognition and classification accuracy of the four bearing states to a certain extent. Zhang et al. [12] inputted bearing vibration features adaptively extracted from one-dimensional CNN into support vector machines, improving the classification accuracy of bearing faults. Qu et al. [13] used a deep network structure of 1DCNN to achieve adaptive extraction of original vibration signal features, and achieved good experimental results in the bearing dataset of Western Reserve University. Chen et al. [14] proposed a bearing diagnosis method based on multi-scale convolutional neural networks and feature alignment to improve the translation invariance of convolutional neural networks and consider the periodic characteristics of vibration signals. This method uses a one-dimensional convolutional neural network to extract multi-scale features from vibration signals, and then uses feature alignment technology to align features of different scales to improve diagnostic performance. The experimental results show that this method has good performance and stability in bearing fault diagnosis. Zhang et al. [15] proposed a deep convolutional neural network (WDCNN) method based on wide convolutional kernels. This method takes the original vibration signal as input and uses a wider convolutional kernel in the first convolutional layer to extract features and suppress high-frequency noise. It also achieves high accuracy under different workloads and noise environment conditions。

The above method is to extract features from the original vibration signal for fault diagnosis. In order to fully utilize the advantages of convolutional neural network feature extraction, one-dimensional vibration signals are converted into two-dimensional time-frequency maps. Li et al. [16] focused on the low signal-to-noise ratio characteristics of rolling bearing vibration signals and used Short Time Fourier Transform (STFT) to process the vibration signals to obtain the STFT time-frequency map, which was used as input for the 2DCNN diagnostic model, verifying the high recognition accuracy and strong robustness of this method. Yuan et al. [17] combined CWT time-frequency maps with 2DCNN and optimized the model by changing structural parameters, improving the diagnostic accuracy and generalization performance of the model. Zhang et al. [18] combined convolutional neural networks with bidirectional gated loop units to extract spatial features of two-dimensional images, and used bidirectional gated loop units to extract temporal features for fault classification. In this study, Wen et al. [19]proposed a novel CNN based on LeNet-5 for fault diagnosis. Through a conversion method converting signals into two-dimensional (2-D) images, the proposed method can extract the features of the converted 2-D images and eliminate the effect of handcrafted features.

3.3. Fault diagnosis based on short-term memory network

Long Short-Term Memory (LSTM) is a variant of Recurrent Neural Network (RNN), which is used to learn historical information in time series and has great advantages in time series data processing. The main feature of LSTM is to control the input, output, and forgetting of information by introducing gating mechanisms, thereby avoiding the problem of gradient vanishing or exploding in traditional RNNs. A single LSTM neuron is shown in Fig.3, where σ represents the activation function Sigmoid. The tanh function is used to adjust the value. The output range is - 1 to 1. Jin et al. [20] proposed a two-way short-term memory neural network, and fused the features extracted from the two dimensions to achieve effective fault diagnosis. Ou et al. [21] constructed a deep learning network model based on the fusion of two-way short-term memory structure and multi-scale convolution structure. Dong et al. [22] proposed a fault diagnosis method based on improved one-dimensional convolution and bidirectional long- and short-term memory (1DCNN-BiLSTM) neural network fusion. This method constructed a 1DCNN-BiLSTM dual channel model, realized full feature extraction and fusion, and introduced the compression and excitation network module. Experiments proved that the model is better than the single channel model. Han et al. [23] proposed a new fault classification method based on bidirectional short-term memory and capsule network with convolutional neural network. This method combines the advantages of BiLSTM in modeling sequence data and capsule network in classifying features, and has achieved good results in experiments. Han et al. [24] proposed a fault diagnosis model based on optimizing the long-term and short-term memory network, and improved the accuracy of fault diagnosis by optimizing the number of hidden layer nodes of LSTM network. Great achievements have been made in fault diagnosis based on long-term and short-term memory networks.

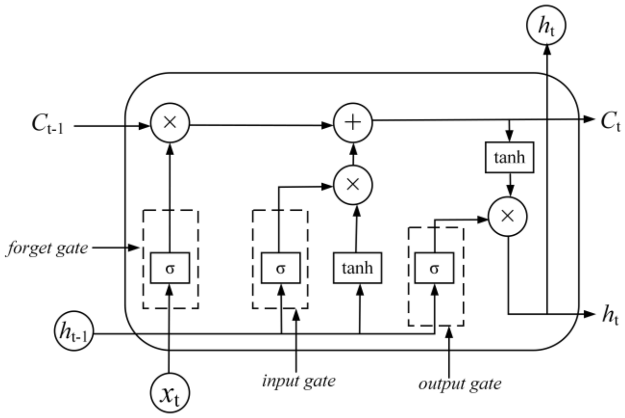

Fig.3 Structure diagram of LSTM neurons[25]

3.4. Summarize the deep learning

Table 1. Advantages and disadvantages of deep learning methods

Deep learning model	advantage	disadvantage
Deep Belief Network	1)No need for large amounts of labeled data 2)Deep structure 3)The algorithm is simple and easy to train 4)Widely used	1)High computational complexity 2)High training time cost 3)Difficulty in parameter adjustment 4)High data volume requirements
Convolutional Neural Network	1)Efficient feature extraction 2)Processing multi-channel data 3)Transferability	1)High demand for data 2)High occupancy of computing resources 3)Easily overfitting
Long- and Short-Term Memory Network	1)Effectively solving the problems of gradient vanishing and gradient explosion in traditional RNN	1)Training time consumption 2)Model performance degradation

The deep learning method focuses on mining the rules of the data layer, utilizing its powerful feature extraction ability to mine the same features between different working conditions to solve the problem of variable working conditions. This method trains model weights through a large amount of data, intelligently mining fault features, and avoids cumbersome steps such as manually extracting features. It is an effective method for achieving variable condition bearing fault diagnosis. This chapter mainly describes the research status of deep confidence network, convolutional neural network and short-term memory neural network in fault diagnosis. Table 1 summarizes the characteristics, advantages, and disadvantages of various deep learning methods.

4. Challenges and research ideas of deep learning in rolling bearing diagnosis

Compared to traditional fault diagnosis methods, which require manual feature extraction, deep learning relies on its powerful learning ability to automatically extract fault features based on a multi-layer neural network structure, achieving variable condition bearing fault diagnosis. Therefore, in the field of fault diagnosis, it has attracted the attention of many scholars and achieved great achievements, but there are still some challenging issues that need to be addressed.

(1) The bearing failure data sample is missing. In actual industrial production, it is difficult to obtain bearing fault data as data samples, so effective feature extraction is a challenge. A feasible approach is to introduce small sample learning [26] to achieve fault diagnosis.

(2) Lack of real scene data. Most research on bearing fault diagnosis based on deep learning is based on simulated data or laboratory manufactured data. These data cannot reflect the impact of complex and ever-changing operating conditions and environmental factors on bearing performance in real scenarios, leading to certain limitations in their reliability in practical applications. And digital twin technology [27] has made preliminary achievements in the energy internet, power industry, and other fields. Therefore, collecting bearing operation data in real scenarios through digital twin technology and conducting bearing fault diagnosis is the future development trend.

(3) Multiple fault classification. Due to the possibility of early faults, weak faults, system faults, and composite faults in rolling bearings, there is no reliable method for deep learning to diagnose rolling bearings in this regard. A feasible approach is to utilize the characteristics of CNN that are suitable for processing massive data, and combine the monitoring data of rolling bearings into a two-dimensional data graph to achieve multi-level and nonlinear complex feature extraction using CNN.

5. Summary

This article mainly introduces the basic principles of the three main models of deep learning and elaborates on the current research status based on these models. Deep learning methods provide new ideas for improving the accuracy and efficiency of fault diagnosis in rolling bearings. Finally, it is pointed out that there are still some problems in the application of deep learning in the diagnosis of rolling bearing faults. Some challenges exist in the areas of missing fault samples, lack of real scene data, and multi fault classification, and feasible suggestions and ideas are proposed.

References

[1] LeCun Yann, Bengio Y, Hinton G. Deep learning[J]. nature, Vol.521(2015),p. 436-444.

[2] Lopes Noel, Ribeiro Bernardete. Deep belief networks (DBNs)[J]. Machine Learning for Adaptive Many-Core Machines-A Practical Approach, Vol.7 (2015), p. 155-186.

[3] Liu Dongdong. Application of Deep Learning in Bearing Fault Diagnosis[J]. Science and technology wind. Vol.2022No9, p. 91-93.

[4] Tao Jie, Liu Yilun, Yang Dalian, et al. Fault diagnosis of rolling bearing using deep belief networks[C]. Energy and Environment Engineering, Proceedings of the 2015 International Symposium on Material, 2015.

[5] Wang Songjin, Peng Zanxin, Yin Han. Fault diagnosis of gearbox bearing based on multi-sensor signal processing[J]. Modular Machine Tool & Automatic Manufacturing Technique, Vol.2020No11, p. 5-10.

[6] Chen Zhuyun, Li Weihua. Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network[J]. IEEE Transactions on Instrumentation and Measurement, Vol.66(2017), p. 1693-1702.

[7] Gan Meng, Wang Cong, Zhu Chang'an. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings[J]. Mechanical Systems and Signal Processing, Vol.72-73(2016), p. 92-104,

[8] Shao Haidong, Jiang Hongkai, Zang Xun, et al. Rolling bearing fault diagnosis using an optimization deep belief network[J]. Measurement Science and Technology, 2015, Vol.26(2017) No.11, p. 115002.

[9] Yang Jianjian, Chang Bochang, Wang Xiaolin, et al. Design and application of deep belief network based on stochastic adaptive particle swarm optimization[J]. Mathematical Problems in Engineering, Vol.2020, p. 10.

[10] Eren L, Ince T, Kiranyaz S. A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier[J]. Journal of Signal Processing Systems, 2019, 91: 179-189.

[11] Hung Shuzhan, Tang Jian, Dai Juying, et al. Signal status recognition based on 1DCNN and its feature extraction mechanism analysis[J]. Sensors, 2019, Vol.19 (2019) No.9, p. 2018.

[12] Zhang Xiaolin, Han Peng, Li Xu, et al. Research on bearing fault diagnosis of wind turbine gearbox based on 1DCNN-PSO-SVM[J]. IEEE Access, Vol.8(2020), p. 192248-192258.

[13] Qu Jianling, Yu Lu, Yuan Tao, et al. Adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural networks[J]. Journal of Instruments and Meters, Vol.39(2018) No.07, p. 134-143.

[14] Chen Junbin, Huang Ruyi, Zhao Kun, et al. Multiscale convolutional neural network with feature alignment for bearing fault diagnosis[J]. IEEE Transactions on Instrumentation and Measurement, Vol.70(2021), p. 1-10.

[15] Zhang Wei, Peng Gaoliang, Li Chuanhao, et al. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals[J]. Sensors, Vol. 17(2017) No2, p. 425.

[16] Li Heng, Zhang Hydrogen, Qin Xianrong, et al, A bearing fault diagnosis method based on short-time Fourier transform and convolutional neural network [J].Vibration and shock, Vol.37(2018)No19 , p.124-131.

[17] Yuan Jianhu, Han Tao, Tang Jian, et al Intelligent fault diagnosis method for rolling bearings based on wavelet time-frequency maps and CNN [J]. Mechanical Design and Research, Vol.33(2017)No02,p.93-97.

[18] Zhang Xunjie, Zhang Min, Li Xianjun. Rolling bearing fault pattern recognition based on two-dimensional images and CNN-BiGRU network [J] Vibration and shock, Vol.40(2021)No23, p. 194-201.

[19] Wen Long, Li Xinyu, Gao Liang, et al. A new convolutional neural network-based data-driven fault diagnosis method[J]. IEEE Transactions on Industrial Electronics, Vol.65(2018)No.7, p. 5990-5998.

[20] Jin Jiangtao, Xu Zifei, Li Chun, et al. Rolling bearing fault diagnosis based on convolutional bidirectional long short memory network and chaos theory [J]. Vibration and Shock, Vol.41 (2022) No.17, p. 160-169.

[21] Ouyang Li, He Shui, Zhu Liangyu, et al. An intelligent fault diagnosis method for bearings based on the fusion of bidirectional short-term memory structure and multi-scale convolution structure [J]. Vibration and Shock, Vol.41 (2022) No.19, p. 179-187.

[22] Dong Yongfeng, Sun Yuehua, Gao Lichao, et al. Fault diagnosis method based on improved one-dimensional

convolution and bidirectional short-term memory neural network [J]. Computer Applications, Vol. 42 (2022) No.4, p. 1207-1215.

[23] Han Tian, Ma Ruiyi, Zheng Jigui. Combination bidirectional long short-term memory and capsule network for rotating machinery fault diagnosis[J]. Measurement, Vol.176(2021), p. 109208.

[24] Han Yongming, Ding Ning, Geng Zhiqiang, et al. An optimized long short-term memory network based fault diagnosis model for chemical processes[J]. Journal of Process Control, Vol.92(2020), p. 161-168.

[25] Hao S, Ge F X, Li Y, et al. Multisensor bearing fault diagnosis based on one-dimensional convolutional long short-term memory networks[J]. Measurement, 2020, 159: 107802.

[26] Che Changchang, Wang Huawei, Xiong Minglan, et al. Few-shot fault diagnosis of rolling bearing under variable working conditions based on ensemble meta-learning[J]. Digital Signal Processing. Vol.131(2022), p. 10377.

[27] Koch Gregory, Zemel Richard, Salakhutdinov Ruslan. Siamese neural networks for one-shot image recognition[C]. ICML deep learning workshop. 2015.