Journal of Electronic Research and Application, 2025, Volume 9, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Research on Bearing Fault Diagnosis Method Based on Deep Learning

Ting Zheng*

Sichuan Institute of Industrial Technology, Deyang 618500, China

***Corresponding author:** Ting Zheng, zhengting0211@163.com

Abstract: Bearing is an indispensable key component in mechanical equipment, and its working state is directly related to the stability and safety of the whole equipment. In recent years, with the rapid development of artificial intelligence technology, especially the breakthrough of deep learning technology, it provides a new idea for bearing fault diagnosis. Deep learning can automatically learn features from a large amount of data, has a strong nonlinear modeling ability, and can effectively solve the problems existing in traditional methods. Aiming at the key problems in bearing fault diagnosis, this paper studies the fault diagnosis method based on deep learning, which not only provides a new solution for bearing fault diagnosis but also provides a reference for the application of deep learning in other mechanical fault diagnosis fields.

Keywords: Deep learning; Bearing failure; Diagnostic methods

Online publication: February 14, 2025

1. Overview of deep learning

1.1. The meaning of deep learning

Deep learning is a machine learning method, the core of which is to build a large number of hidden layers, through the training and learning of massive data, to build a deep nonlinear network structure to approximate complex functions. Compared with traditional machine learning methods, deep learning has significant advantages in processing high-dimensional, non-linear data, and can capture deep abstract features in the data. The essence of deep learning lies in its ability to construct multi-layer neural network structures. Each layer of neural network contains a large number of neurons, which are connected by weights to form a network. During training, the network constantly adjusts the weights to minimize the error between the predicted value and the true value, using forward propagation and backpropagation algorithms. In this way, deep learning models can progressively approximate complex functional relationships, enabling efficient learning and understanding of the data. In practical applications, deep learning models are trained on a large amount of data and can learn the inherent laws in the data, so that they can perform well on new and unseen data. This ability to automatically extract features from data has enabled deep learning to achieve remarkable results in areas such as image recognition, natural language processing, and speech recognition [1].

1.2. Characteristics of deep learning

Deep learning is characterized by the hierarchical depth of its model architecture. Compared with traditional machine learning models, deep learning sets multi-level and multi-dimensional hidden nodes in the hidden layer for training and learning. These hidden nodes are not only numerous, but also hierarchical, and the nodes of each layer will obtain information from the training results of the previous layer and use it as an input sample for the next layer. Through this way of training and learning layer by layer, the model can capture more complex and abstract features in the data, thus effectively improving the accuracy and generalization ability of the model [2].

In addition, another great feature of deep learning is its strong ability in feature learning. Deep learning neural network combines classifiers with feature learning so that the system can automatically extract and learn feature information [3]. This ability of automatic feature learning effectively reduces the workload of manual design and feature extraction and improves the efficiency and accuracy of the model. In bearing fault diagnosis, the deep learning model can automatically extract the time domain, frequency domain, and time-frequency domain features of the vibration signal through structures such as convolutional neural network (CNN), and combine and transform these features layer by layer to map the sample features in the original feature space to the new feature space. Through this multi-level and multi-dimensional feature extraction and learning, the deep learning model can accurately diagnose the bearing fault type in a complex industrial environment, and provide strong support for equipment maintenance and fault prevention.

2. Problems of deep learning in bearing fault diagnosis

2.1. The deep learning fault diagnosis method is a single model

Most of the existing research focuses on the use of a single model, such as convolutional neural network (CNN), recurrent neural network (RNN), or autoencoder (AE), which performs well on specific tasks, but the limitations of a single model gradually emerge when faced with complex and variable fault modes. In addition, the generalization ability of a single model is limited, and it is difficult to adapt to the fault diagnosis task under different working conditions and environments. There are various types of bearing faults, and each fault has different signal characteristics, and a single model is often difficult to fully cover these characteristics. In actual industrial applications, the challenges of bearing fault diagnosis are more complex. In the industrial environment, the noise and interference of data acquisition are inevitable, and the performance of a single model will be significantly degraded when processing these non-ideal data, making it difficult to achieve stable and reliable fault diagnosis [4].

2.2. Lack of experience in the hybrid collaborative operation of the model

Firstly, the strategy of model selection and combination is not systematic. How to choose the right model combination, how to determine the weight of each model, and how to optimize the interaction between models, these problems have not formed a set of mature methodologies in the existing research, resulting in the effect of model combination in practical application is not satisfactory, it is difficult to achieve the expected diagnostic accuracy. Secondly, the limitations of the data set further aggravate this problem. In real industrial settings, high-quality failure data is difficult to obtain. This results in insufficient diversity and representation of data in the process of model training, which affects the generalization ability and synergistic effect of the model. Moreover, the information transfer mechanism between the models is not perfect [5]. The existing information transfer mechanism is too simple, and lack flexibility and robustness, the effect of multi-model collaborative operation is limited, and it is difficult to give full play to the advantages of each model. Additionally, deep learning models, especially complex neural network models, are often seen as “black box” models whose internal decision-making processes are difficult to explain. This is an important flaw in fault diagnosis, as engineers and maintenance personnel need to understand the diagnostic basis of the model to carry out subsequent fault treatment and prevention. However, the existing methods still fall short in this respect, lacking effective means to improve the interpretability and transparency of the model [6].

2.3. The accuracy of feature extraction and recognition is low

Since the bearing fault signal has nonlinear and non-stationary characteristics and is affected by many factors in actual working conditions, it is difficult to extract the fault characteristics accurately [7]. Especially in complex working conditions, such as the coexistence of multiple faults and the early stage of faults, the weak features of the signal are more difficult to capture, which makes deep learning models face great challenges in feature extraction. In terms of recognition accuracy, the performance of deep learning models is also affected by a variety of factors. One of which is insufficient training data. The acquisition of bearing fault data usually needs to be carried out under specific working conditions, which are difficult to simulate completely in the actual industrial environment, resulting in insufficient diversity and representation of training data.

The second is the impact of noise and interference on the performance of the model. In the actual industrial environment, the signal collected by the sensor is often interfered with by various noises, which will not only reduce the quality of the signal but also introduce additional features, making the model prone to misjudgment in the recognition process. The third is the overfitting problem of the deep learning model. In the training process, if the model is too complex, it is easy to overfit the noise and outliers in the training data, which leads to the decline of the generalization ability on the test data and the reduction of the recognition accuracy [8].

3. The application of deep learning in bearing fault diagnosis

3.1. Automatic encoder

As an unsupervised learning method, the autoencoder effectively overcomes the difficulties of traditional methods in feature extraction by automatically learning features in advanced representations. In bearing health monitoring, the structure of the autoencoder is very simple, usually composed of an encoder and a decoder. The encoder is responsible for mapping the input data to a low-dimensional implied representation, while the decoder maps this implied representation back into the original data space. By minimizing the difference between the input data and the reconstructed data (usually using the mean square error as a loss function), the autoencoder can learn an efficient representation of the input data [9].

In specific applications, autoencoders can be used to extract key features from bearing vibration signals. The vibration signal of the bearing usually contains rich fault information, but this information is often covered up by noise and nonlinear factors. By learning the advanced representation of these signals, the autoencoder can effectively extract the fault characteristics, to improve the accuracy and robustness of fault diagnosis. For example, in a typical bearing fault diagnosis experiment, vibration signals under different working conditions are collected and the autoencoder is used for feature extraction. The encoder compresses the high-dimensional vibration signal into a low-dimensional implied representation, and the decoder attempts to reconstruct the original signal from this implied representation [10]. By training the autoencoder, it is possible to obtain a low-dimensional feature space that can efficiently represent the bearing state. Additionally, the autoencoder can not only be used for fault diagnosis of a single bearing but also can be extended to the joint diagnosis of multiple bearings. By building an autoencoder model capable of processing multiple inputs, vibration signals of multiple bearings can be analyzed simultaneously, enabling more comprehensive fault monitoring.

3.2. Restricted Boltzmann machine

Restricted Boltzmann machine (RBM), as an important component of deep learning, has been widely concerned about its application in bearing fault diagnosis [11]. RBM is an unsupervised learning algorithm that can learn the internal structure of data from a large amount of unlabeled data. RBM forms a bipartite graph model by establishing a connection between the input layer and the hidden layer, in which the input layer nodes are fully connected to the hidden layer nodes, but the nodes within the same layer are not connected. By learning the probability distribution of the input data, RBM can automatically extract the feature representation of the data, a feature that gives it a significant advantage when dealing with complex fault signals.

In bearing fault diagnosis, RBM can be used for feature extraction and feature representation. By training the RBM model, high-order features can be extracted from the vibration signals of bearings, which can reflect the fault state of bearings more accurately. For example, Xiong *et al.* used particle swarm optimization to optimize deep confidence network (DBN) for bearing fault diagnosis. DBN is a depth model composed of multiple RBMs superimposed. By stacking multiple RBMs, the accuracy of feature extraction can be further improved. Particle swarm optimization algorithm is used to optimize the parameters of DBN so that the model can converge to the optimal solution faster in the training process, to improve the accuracy and robustness of fault diagnosis. Tao *et al.* considered the fault diagnosis system of multi-signal fusion and used RBM to conduct joint modeling of multi-source signals. In practical applications, the fault information of bearings is often not only included in the vibration signal but also may be reflected in a variety of signals such as temperature and current. Through multi-signal fusion, the fault characteristics of bearings can be captured more comprehensively and the reliability of diagnosis can be improved. The application of RBM in multi-signal fusion can not only process the data of different modes but also extract richer feature representations through joint modeling, to improve the accuracy of fault diagnosis [12].

3.3. Convolutional neural network

Convolution neural network (CNN) as a kind of deep learning model, has been in image recognition, natural language processing, and other fields and has made remarkable achievements. In recent years, CNN has also shown strong potential in mechanical fault diagnosis, especially in bearing fault diagnosis. Guo *et al.* proposed a layered convolutional neural network with an adaptive learning rate for classifying bearing failures and further determining their severity. This method can effectively identify the type and degree of bearing failure through multi-level feature extraction and provides a new idea for the maintenance and fault prediction of industrial equipment [13].

In this study, Guo *et al.* first preprocessed bearing fault data, including steps such as data acquisition, noise filtering, and normalization. In the process of data acquisition, high-speed data acquisition cards, and acceleration sensors were used to obtain bearing vibration signals under different working conditions with high sampling rates. Noise filtering is used to remove the high-frequency noise in the signal through wavelet transform and other methods to improve the signal-to-noise ratio [14]. The normalization process is to unify the signals of different amplitudes to the same order of magnitude, which is convenient for subsequent feature extraction and model training. In the process of model training, Guo *et al.* adopted the optimization algorithm of adaptive learning rate. The traditional fixed learning rate may lead to slow convergence or local optimization in the training process. The adaptive learning rate can dynamically adjust the learning rate according to the gradient changes in the training process, which makes the model converge quickly in the early stage and fine-adjust in the late stage, thus improving the training efficiency and performance of the model. Specifically, the research team used the Adam optimization algorithm, which combines the advantages of the momentum method and RMSprop to perform well in complex optimization problems [15].

4. Conclusion

In short, this paper introduces the basic concepts, characteristics, and common models of deep learning in detail. By analyzing the problems existing in the current deep learning in bearing fault diagnosis, such as single method, lack of experience in the hybrid collaborative operation of models, and low accuracy of feature extraction and recognition. Given these problems, a comprehensive diagnosis method based on automatic encoder, restricted Boltzmann machine and convolutional neural network is proposed, which is helpful to improve the diagnosis efficiency and broaden the practical application in complex working conditions.

Disclosure statement

The author declares no conflict of interest.

References

[1] Ren H, Qu J, Chai Y, et al., 2017, Research Status and Challenges of Deep Learning in Fault Diagnosis. Control and Decision, 2017(8): 1345–1358.

[2] Hu N, Chen HP, Cheng Z, et al., 2019, Fault Diagnosis Method of Planetary Gearbox Based on Empirical Mode Decomposition and Deep Convolutional Neural Network. Chinese Journal of Mechanical Engineering, 2019(7): 9–18.

[3] Li M, Li J, Liu Y, 2023, Application of CEEMDAN-TFPF Noise Reduction Method Based on Autocorrelation in Gear Fault Diagnosis. Science of Digital Manufacturing, 2023(3): 199–204.

[4] Yan B, Li X, 2023, Research on Ultrasonic Location System and Algorithm for Cracks in Stone Relics. Chinese Journal of Scientific Instrument, 2023(8): 155–163.

[5] Kang S, Yang J, Wang Y, et al., 2023, Fault Diagnosis Method of Rolling Bearing Under Different Working Conditions Based on Federated Multiple Representation Domain Adaptation. Chinese Journal of Scientific Instrument, 2023(6): 165–176.

[6] Song Z, Xu L, Hu X, et al., 2021, Research on Axle Box Bearing Fault Diagnosis Method of EMU Based on Improved Shapelets Algorithm. Chinese Journal of Scientific Instrument, 2021(2): 66–74.

[7] Chen R, Zhang Y, Yang L, et al., 2020, Health State Evaluation of Harmonic Retarder Based on Whole Cycle Data

and Convolutional Neural Network. Chinese Journal of Scientific Instrument, 2020(2): 245–252.

[8] Gong W, Chen H, Zhang M, et al., 2020, Intelligent Diagnosis Method of Motor Bearing Minor Fault Based on Deep Learning. Chinese Journal of Scientific Instrument, 2020(1): 195–205.

[9] Dong X, Guo L, Gao H, et al., 2019, Cost Sensitive Convolutional Neural Networks: An Unbalanced Classification Method for Mechanical Fault Data. Chinese Journal of Scientific Instrument, 2019(12): 205–213.

[10] Wen Z, Chen J, Liu L, et al., 2022, Fault Diagnosis of Wind Power Gearbox Based on Wavelet Transform and Optimal CNN. Journal of Zhejiang University (Engineering and Technology), 2022(6): 1212–1219.

[11] Meng Z, Guan Y, Pan Z, et al., 2021, Research on Fault Diagnosis of Rolling Bearing Based on Secondary Data Enhancement and Deep Convolution. Chinese Journal of Mechanical Engineering, 2021(23): 106–115.

[12] Xin R, Miao F, Wang T, et al., 2022, Intelligent Fault Diagnosis of Mechanical Bearing Based on Deep Learning. Journal of Jilin University of Chemical Technology, 2022(11): 25–29.

[13] Xu Y, 2020, Deep Learning Algorithm for Working State Detection of Electromechanical Equipment. Electronic Measurement Technology, 2020(11): 34–38.

[14] Li C, Wu Y, Yang S, 2023, Research Progress of Data-Driven Fault Diagnosis with Unbalanced Distribution. Chinese Journal of Scientific Instrument, 2023(8): 181–197.

[15] Hu Y, Liu S, Li Q, et al., 2019, Fault Diagnosis of Rolling Bearing Based on EMD Amplitude Entropy and Support Vector Machine. Mechanical Research and Application, 2019(2): 166–169.