Method of extracting gear fault feature based on stacked autoencoder

Shuo Liu\(^1\), Yulong Liu\(^1\), Yuhai Gu\(^1\)\(\textsuperscript{✉}\), Xiaoli Xu\(^1\)

\(^{1}\)Key Laboratory of Electromechanical Measurement and Control, Beijing Information Science & Technology University, Beijing, People’s Republic of China

\(\textsuperscript{✉}\)E-mail: guyuhai@bistu.edu.cn

Abstract: Gear and its transmission are widely used in different transmission systems, and its complicated and changeable condition brings a series of problems to the fault feature extraction and diagnosis. In recent years, deep learning techniques have been gradually applied to feature extraction and pattern recognition, and the features of feature extraction and fault diagnosis in complex working environments have shown certain advantages. This study is based on stacked autoencoder under deep learning model, and improve training network performance by modified activation function. Through the network training before and after the experiment done, and to extract the fault feature data comparison in testing, improving network after activation function to extract fault features showed a greater advantage, can be a very good application in practical fault feature extraction.

1 Introduction

Gear as one of the important transmission parts, the stability and accuracy of the drive is applied in more scope. The wide application also brings about its complicated and changeable working environment, which makes it difficult to diagnose gear fault. In the analysis of the vibration signal of the fault gear, the signal processing technique used is the short-time Fourier transform, wavelet transform, wavelet packet transform, empirical mode decomposition and so on [1].

In recent years, the feature extraction of deep learning shows unique advantages and has been gradually applied to solve the problems of complex industrial systems [2]. Stacked autoencoder (SAE) is one of the deep learning model architecture, data as the main drive, using single hidden layer in unsupervised learning mode to a certain processing of data, so as to make the network to extract features. Yu et al. [3] proposed a sparse coding vibration signal feature extraction based on improved dictionary learning and shift invariant component filtering. Miao et al. [4] use the vibration signal effective feature extraction method to effectively solve the extraction of device status information features in redundant data. Lee et al. [5] used stacked noise AE for feature extraction and classification of chip faults, which effectively improved the accuracy of fault diagnosis. Sun et al. [6] used the sparse AE to learn features and then identified the induction motor faults, which proposed a new regularisation method to effectively avoid overfitting.

The above improvements and modifications in different aspects have made the application of AE better. However, the most basic network influencing factors are less improved or optimised, so the application and fault diagnosis of stacked AE will bring some disadvantages. In view of this problem, this paper proposes to improve the activation function in the deep network, and to select the appropriate network through training, and apply it to the actual fault diagnosis.

2 Stacked AE

AE is mainly driven by data, and a single hidden layer is used to deal with data characteristics in unsupervised learning mode so that the neural network can be extracted [7]. AE encodes the input and give the only encoded representation (i.e. for the coding phase), and then according to the code refactoring input (i.e. for decoding phase), to achieve its target output as the original input, as shown in Fig. 1.

2.1 Stacked AE

SAE is a neural network, the type of automatic coding machine so deep learning stack network model structure is composed of multiple AE in series with the stack, as an intermediary between the AE also will be the next input layer, would be enough layers of AE stack, make up for the need of deep web structure. In this network, the basic unit is an AE. Using the network as much as...
Mapping the hidden layer is $y \in [0, 1]^d$. In the formula, $\theta = \{W, b\}$, $W$ is the weight matrix of $d \times d$ and $b$ is the bias vector of $d$

$$s(x) = \frac{1}{1 + e^x} \quad (2)$$

This function is the activation function of the deep network SAE neural network.

Step 2: Since it is SAE, it is going to get $y$ at the end of the equation with the mapping function

$$z = g_d(y) = s(Wy + b) \quad (3)$$

In the above formula, it maps to the next hidden layer $z \in [0, 1]^d$, $\theta = \{W, b\}$ and by setting $W = W^T$ to obtain the optimal inverse mapping weight matrix $W$.

Step 3: Repeat step 1 and step 2, each input $y^0$ mapping can get the corresponding $y^0$, and then the corresponding $z^0$ is reconstructed, and the optimal parameters of the whole model can be obtained by minimising the reconstruction error each time, as follows [8]:

$$\theta^*, \theta^* = \arg \min \frac{1}{n} \sum_{i=1}^{n} L(x^{(i)}, y^{(i)}) \quad (4)$$

In (4), $L(x)$ is the loss function. In this function, each component of the vectors $x$ and $y$ is the Bernoulli distribution, and the spacing between vectors $x$ and $y$ is measured using cross-moisture. The loss function is as follows:

$$L(x, z) = -\sum_{k} (x_k \log z_k + (1 - x_k) \log(1 - z_k)) \quad (5)$$

In (5), $x_k$ is the input data of the $k$-layer network and $z_k$ is the output data of the $k$-layer deep network.

### 2.3 Modified activation function

In depth network training and application in fault diagnosis, the activation function has great influence on the deep network processing signal. According to (2), the commonly used activation function is

$$s(x) = \frac{1}{1 + e^x}$$

This function is the sigmoid function and the function image is shown in Fig. 3. According to the curve change found, when the input $x$ is larger or smaller, the output value gradually tends to stabilise the value, which is used in network training, when the input $x$ is larger or smaller, will affect the network learning which cannot analyse the complete data.

Therefore, we adjust the activation function to

$$s(x) = \begin{cases} x, & \text{when } x \geq 0 \\ ax, & \text{when } x < 0 \end{cases} \quad (6)$$

Equation (6) is the leaky rectified linear unit (ReLU) function, where $a \in (0, 1), U, 1 < U$ and $U \in [0, 1]$. The adjustment of activation function is very good to solve the problem that the sigmoid function training should input larger or smaller, which can also enable the activation function to converge and improve the network performance.
3 Gear fault feature extraction method based on SAE

In the diagnosis of gear faults, the first is to use wavelet preprocessing, and the second is to use deep network for feature advance and fault classification. The first step is to use wavelet preprocessing to improve the convergence speed of deep networks. Here, is its detailed diagnostic procedure. Fig. 4 shows the design of the diagnostic procedure.

3.1 Wavelet preprocessing implementation

Step 1: The benchmarking process is performed by collecting data from the points. The reason for this is mainly to remove the influence of factors such as the environment on the vibration measurement sensor.

Step 2: Select sym3 as the basic wavelet function, determine the signal decomposition into five layers, and implement wavelet decomposition transform.

Step 3: The signal in the frequency domain is obtained after the processing in step 2, and this signal is processed through a threshold using wavelet coefficients. Thresholds are obtained using the improved threshold calculation method in Section 3.2.

Step 4: After the signal threshold is processed, the signal is reconstructed by the wavelet inverse transform. This completes the denoising of the signal.

Step 5: The denoising signal is prepared for the next step in entering the deep network.

3.2 Deep network training of deep learning

According to the above study, a deep learning SAE was selected in this system. SAE forms a deep network by laminating automatic encoders. For this network training, it is necessary to input the original signal to carry out signal data, which is very good for the automatic encoder to learn to extract the signal characteristics, and ultimately realize fault diagnosis for the real signal. This requires the encoder to better improve the stability of the expression in the raw signal data. The gradient descent algorithm was chosen during the training of the automatic encoder. For the actual application of the deep network in this system, it can be seen that the automatic encoder used is not only a single hidden layer network. The SAE deep network is based on greedy layer-by-layer training proposed by Hinton et al., that is, by alternately training each layer of the deep network. Calculate the loss function; the main definition is as follows [9]: (see (7)).

\[ S_j \text{ is the number of hidden layer nodes and } L(W, W', b, b'; X) \text{ is a loss function.} \]

\[ KL(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \]  

(8)

There is data loss between encoding and decoding. The loss processing formula is as follows:

\[ X_C = \text{rand(size, corrupted}_\text{level}) \times X \]  

(9)

In the above formula, rand() is the same matrix as \( X_C \). The matrix takes probability \( 0 - \text{corrupted}_\text{level} \) as 0, probability \( 1 - \text{corrupted}_\text{level} \) as 1 and the encoding stage is as follows [10]:

\[ h = f(WX_C + b) \]  

(10)

The decoding stage is as follows:

\[ Y = g(W'h + b) \]  

(11)

The specific training steps are:

Step 1: Using formula (9) to perform loss processing on the original data set, a loss data set \( X_C \) can be obtained.

Step 2: Take \( X_C \) as input and encode according to formula (10) (that is, forward propagation processing). Then according to formula (11) to decode, get the output \( Y \).

Step 3: Use formula (7) to calculate the loss value.

Step 4: Then, according to the modified activation function in Section 2.2, the gradient descent algorithm is trained to obtain the optimised data.

Step 5: Repeat step 2–step 5 until the loss function converges.

3.3 Deep learning fault feature extraction

The data preprocessed by the wavelet is output to a deep network that has been trained, and the fault feature can be obtained by processing in the deep network.

4 Experimental verification

To verify the effect of the SAE on the fault feature extraction of the worn gear, the experimental sample data was collected on the experimental bench of the fan drive simulation system, as shown in Fig. 5. The drive simulation system used in the test bench during the collection process zero load, the sampling frequency is 20 kHz, the acquisition duration is used for data collection every 4 s, and the data acquisition has four channel signals. They are pulse speed, of which 50 sets were used for training the network and 50 sets were used for testing. According to the different locations in Fig. 5, the signals are divided into three groups including high-
speed stages C2, C3 and low-speed stage C4, where the acquisition point C1 is used to measure the rotation speed. Fig. 6 shows a schematic diagram of the structure of the test stand planetary gear box. The wear gear in Fig. 6 shows a B planetary gear. The time-domain and frequency-domain diagrams of the acquired signals are shown in Fig. 6. It can be seen from this figure that the complexity of the noise signal occupies different frequency bands, which brings certain difficulty to the extraction of fault features. Amplifying the signal on the one hand leads to a slower convergence in the training of the SAE machine, and on the other hand, the use of the sigmod function as an activation function results in an overly large input affecting the comprehensiveness of the network learning. Therefore, the wavelet denoising preprocessing is used to improve the network convergence rate. At the same time, the stacking automatic coder of leaky ReLU function activation function was used for feature extraction.

First, the training dataset is preprocessed by wavelet denoising, then the spectrum of the denoising signal is obtained through Fourier transform, and then the acquired spectrum data is used for deep learning network training input data. The 100-layer deep learning network was also established during training. The training set samples were divided into ten groups of data for input. To improve the accuracy of the network, five trainings were performed for each group of data, and the convergence of five training sessions was used as shown in Table 1. When the collected data is used for feature extraction through the training network, the data is first preprocessed, and then the preprocessed data is input into the deep learning network where the training is completed. To verify the strengths and weaknesses of deep networks with different activation functions, the preprocessed data are input into the training network under the sigmod activation function and the leaky ReLU activation function, respectively, and the comparison of processing the same set of data is shown in Fig. 7. It can be seen in this figure that the characteristics of the sigmod activation function in the SAE are influenced by the activation. Function training is done so that the characteristic peaks are processed and cannot be resolved. By testing 50 sets of data, the correctness of feature extraction for the two different networks under sigmod activation function and leaky ReLU activation function is compared, as shown in Table 2. From Table 2, it can be concluded that leaky ReLU activation function network can exhibit greater advantages in feature processing of data for feature extraction under large and small environments. With the same data, the leaky ReLU activation function network is 54% more accurate than the sigmod activation function network. This also illustrates the advantages of the SAE of the leaky ReLU activation function for gear fault feature extraction.

### Table 1  Training convergence data

| No |  | Sigmoid activation function network |  | Leaky ReLU activation function network |  |
|----|---|------------------------------------|---|---------------------------------------|---|
|    | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| time, s | 77 | 76 | 74 | 76 | 58 | 73 | 70 | 72 | 74 | 75 |
| performance, ×10⁻³ | 102 | 2.9 | 6.07 | 0.293 | 0.896 | 246 | 2.18 | 8.94 | 0.234 | 0.43 |
| gradient, ×10⁻⁵ | 2.65 | 0.499 | 0.095 | 98.8 | 0.096 | 0.496 | 0.883 | 0.424 | 0.184 | 1.00 |
| iterations, step | 1000 | 1000 | 980 | 1000 | 762 | 1000 | 905 | 1000 | 1000 | 1000 |

This is an open access article published by the IET under the Creative Commons Attribution License

(http://creativecommons.org/licenses/by/3.0/)
Conclusion

An improved method for extracting fault frequency spectrum features from an SAE is proposed. Compared to the traditional feature extraction methods, the signal processing knowledge is eliminated, and at the same time, more comprehensive data feature analysis is performed based on big data, which greatly improves the accuracy.

Acknowledgments

This work was supported by the Beijing Scholars Plan (2015–025) and the National Natural Science Foundation of China (51575055).

References

[1] Shi, P.M., Liang, K., Zhao, N., et al.: ‘Gear based intelligent fault diagnosis based on deep learning feature extraction and particle swarm optimization based on support vector machine state recognition’, China Mech. Eng., 2017, 28, (9), pp. 1056–1061 + 1068

[2] Ren, H., Qu, J.F., Chai, Y., et al.: ‘Research status and challenges of deep learning in fault diagnosis’, Control Decis., 2017, 32, (8), pp. 1345–1358

[3] Yu, L., Qu, J.L., Gao, F., et al.: ‘Feature extraction algorithm for weak vibration signals based on improved sparse coding’, Chin. J. Sci. Instrum., 2017, 38, (3), pp. 711–717

[4] Miao, Z.H.H., Zhou, G.X., Liu, H.N., et al.: ‘Feature extraction algorithm and experimental study of vibration signal based on sparse coding’, J. Vib. Shock, 2014, 33, (15), pp. 76–81 + 118

[5] Lee, H., Kim, Y., Chang, O.K.: ‘A deep learning model for robust wafer fault monitoring with sensor measurement noise’, IEEE Trans. Semicond. Manuf., 2017, 30, (1), pp. 23–31

[6] Sun, W., Shao, S., Zhao, R.: ‘A sparse auto-encoder-based deep neural network approach for induction motor faults classification’, Measurement, 2016, 89, (ISFA), pp. 171–178

[7] Gupta, L., Samaka, M., Jain, R., et al.: ‘Fault and performance management in multi-cloud based NFV using shallow and deep predictive structures’, J. Reliab. Intell. Environ., 2017, 3, (4), pp. 221–231

[8] Mijumbi, R., Serrat, J., Gorricho, J.L., et al.: ‘Management and orchestration challenges in network functions virtualization’, IEEE Commun. Mag., 2016, 54, (1), pp. 98–105

[9] Zembutsu, H., Miyata, Y., Yoshimura, Y.: ‘Network function virtualization management orchestration device, method, and program’, WO/2017/188387[P], 2017

[10] Casellas, R., Martinez, R., Vilalta, R., et al.: ‘Control, management, and orchestration of optical networks: evolution, trends, and challenges’, J. Lightwave Technol., 2018, 36, (7), pp. 1390–1402