2018 Prognostics and System Health Management Conference

Unsupervised Gear Fault Diagnosis Using Raw Vibration Signal Based on Deep Learning

Xueyi Li and Zhendong Liu
College of Mechanical Engineering and Automation
Northeastern University
Shenyang, China

Yongzhi Qu
School of Mechanical and Electronic Engineering
Wuhan University of Technology
Wuhan, China

David He
College of Mechanical Engineering and Automation
Northeastern University
Shenyang, China

Department of Mechanical and Industrial Engineering
University of Illinois at Chicago
Chicago, USA
davidhe@uic.edu

***Abstract*—Gears are the most common parts of a mechanical transmission system. Gear wearing faults could cause the transmission system to crash and give rise to the economic loss. It is always a challenging problem to diagnose the gear wearing condition directly through the raw signal of vibration. In this paper, a novel method named augmented deep sparse autoencoder (ADSAE) is proposed. The method can be used to diagnose the gear wearing fault with relatively few raw vibration signal data. This method is mainly based on the theory of wearing fault diagnosis, through creatively combining with both data augmentation ideology and the deep sparse autoencoder algorithm for the fault diagnosis of gear wear. The effectiveness of the proposed method is verified by experiments of six types of gear wearing conditions. The results show that the ADSAE method can effectively increase the network generalization ability and robustness with very high accuracy. This method can effectively diagnose different gear wearing conditions and show the obvious trend according to the severity of gear wear faults. This paper provides an important insight into the field of gear fault diagnosis based on deep learning and has a potential practical application value.**

***Keywords*—autoencoder; augmentation; gear fault diagnosis; unsupervised learning; raw vibration signal**

I. INTRODUCTION

Gears are one of the most widely used components in rotary machinery including the automobile, the airplane, and the manufacturing equipment. So it is important to monitor and diagnose wear conditions of the gears. Timely detection and control of gear wear faults can effectively avoid further loss. Therefore, effective earlier fault detection and diagnosis are vital to the normal operation of machinery. The diagnosis of gear wear faults and identification of the damaging grade are needed in engineering practice. With the development of fault diagnosis technology, mounting number of technologies are applied to gear wear fault diagnosis. The wear condition of gear teeth in gear fault diagnosis is an important index to show whether the machine can run continuously. Currently, gear fault diagnosis can depend on needless prior knowledge accumulation and fault diagnosis experience.

The vibration signal of gears is widely used in fault diagnosis of gears. In recent years, quite a number of methods for gear fault detection have been proposed. Some research are using fuzzy and artificial neural network for the fault diagnosis of a spur gear. And using the decision tree to manually extracts some statistical functions such as skewness, standard deviation [1]. In order to extract the features of the different sub-bands frequencies in the vibration signal, wavelet packet transform and principal component analysis are always used [2]. Also, there are some individuals through attribute clustering to hierarchical propose feature [3]. Li *et al*. [4] proposed using wavelet packet and two support vector machine models recognize faults. On the basis of the Heidari's study, Wang [5] added *K*-nearest neighbors method to practice. There are many other ways to analyze vibration data, e.g., using multiscale weighted permutation entropy to extract the hidden fault-related features over multiple temporal scales [6]. Hilbert transform and Euclidean distance technique were developed for fault diagnosis and damage level identification [7]. These methods can be used to extract features. Nevertheless, the artificial feature extraction requires a lot of time and labor costs. In the aspect of automatic feature extraction, deep learning methods represent good choices.

In addition, wavelet packet transform and principal component analysis are used to extract features. Normally, monitoring procedure for a gear transmission system uses artificial neural networks and support vector machines [8].

2166-5656/18/$31.00 ©2018 IEEE
DOI 10.1109/PHM-Chongqing.2018.00182

Zhang *et al*. [9] used wavelet packet decomposition. And after that operation signals are transformed to frequency domain using fast Fourier transform. Some papers show that high-pass filtering improved the success rates remarkably in the case of Directed acyclic graph support vector machine (DAG-SVM) [10]. In addition, some papers propose a new gear fault identification method based on Hilbert-Huang transform (HHT) and self-organization map (SOM) neural network [11]. Gan *et al*. proposed a novel multiple-domain manifold method to achieve representative features based on singular value decomposition and manifold learning, as well as the improved Hilbert-Huang spectrum of the signal [12]. These methods have also achieved favorable results. However, these approaches would make more use of existing prior knowledge and involve a lot of personal factors. In the face of different faults, it is easy to cause serious bias. Moreover, these methods mainly deal with the problem of dichotomy, which can be very troublesome for multi-classification problems such as twenty-classification or more.

The vast majority of processing methods depend on transformation of the time domain signals into frequency domain signals and then use the frequency domain signals to classify gear wearing faults. There is a rarely way to deal with the time domain signal directly. Reference [13] was the first one using unsupervised deep learning method for diagnosis of gear pitting fault. Now, the majority of the methods are supervised learning, which require a large number of labeled samples. It usually needs a lot of data for deep learning for training. However to obtain large amounts of data is not easy. To overcome the challenge of obtaining large amount of data for gear wearing fault diagnosis, in this paper, an augmented deep sparse autoencoder algorithm (ADSAE) is proposed. The proposed method can used for gear wearing fault diagnosis with only limited raw vibration signal data.

The remainder of this paper is organized as follows: Section 2 elaborates ADSAE algorithm proposed in the paper. In Section 3, 6 experiments were designed to verify whether the ADSAE algorithm is effective in classifying gear wearing faults. Section 4 analyses and discusses the experimental results and future research direction. The current research conclusions are summarized in Section 5.

II. THE METHODOLOGY

A. Sparse Autoencoder

Autoencoder (AE) is an unsupervised neural network. AE is widely used in many aspects. The encoder part has a good effect on image dimensionality reduction, while the decoder part has a good performance in the field of generating antagonism network. At present, there are two main applications. The first one is data denoising and the second one is dimensionality visualization. With appropriate net size and sparse constraints, AE can learn more interesting features. If the decoded data can be easily restored to the original data by AE, it is believed that the artificial neural network has retained the data information better. The AE includes the input layer, hidden layer, and output layer. The input layer is usually at the beginning of the network, and the output layer is generally at the end. The hidden layer can be more than one layer. The network structure is shown in Fig. 1. In Fig. 1， Layer 1 is the input layer, Layer 2 to Layer 4 are the hidden layers, and Layer 5 is the output layer. Layer 1 to Layer 3 are the encoder parts of the AE. Layer 3 to Layer 5 are the decoder parts of the AE. The encoder parts are mainly used to compress the features, and the decoder parts are used to recover compressed features.

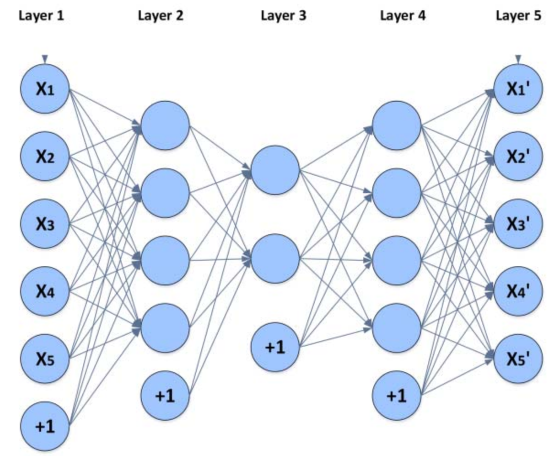

Fig.1 The structure of autoencoder.

Suppose that the AE input feature sample set is $\mathrm{X} = \{X_1, X_2, X_3, X_4, X_5\}$. Then function f_θ converts each training sample X_n to the next layer as defined in Eq. (1). In the Eq. (1), s is on behalf of the activation function. The parameter matrix of the AE network is the value $\theta = \{W, b\}$. W is the weight matrix of the first layer to the second layer. b is the bias coefficient, which is the circle "+1" in the Fig. 1.

$$h_{w,b}(X) = f_\theta(X_n) = s(WX_n + b) \quad (1)$$

The last layer output vector $\mathrm{X}' = \{X_1', X_2', X_3', X_4', X_5'\}$ is reconstructed with decoding function Eq. (2) by the decoder network. The parameter matrix of the decoder network is generated. In the Eq. (2), W' is the weight matrix from the hidden layer to the output layer. b' is the bias coefficient from the hidden layer to the output layer.

$$X' = g_{\theta'}(h_{w,b}(X)) = f_\theta(X_n) = s(W' h_{w,b}(X) + b') \quad (2)$$

AE uses the gradient descent algorithm to adjust the network weights. A lot of iterations reduce error function values $J(W, b)$ by Eq. (3). Finally, the output vector X' realized. In the Eq. (3), *nl* is network layer number, *sl* is the number of units in the corresponding layer, λ Is the weight attenuation coefficient, and m is the number of samples. Equation (3) is shown that the hidden layer vector $h_{w,b}(X)$ retains the key characteristics of high dimensional input data.

$$J(W, b) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}|X - X'|^2\right) + \frac{\lambda}{2}\sum_{l=1}^{nl} \sum_{i=1}^{sl} \sum_{j=1}^{sl-1}\left(W_{ji}^{(1)}\right)^2 \quad (3)$$

In order to improve the practicability of AE, it is necessary to limit the learning ability of AE. So we put forward the concept of the sparse autoencoder (SAE) [14]. SAE puts AE network contacting with sparse restrictions through making part of the hidden layer neurons in the state of inhibition. SAE can make the connection more varied. The correlation between neurons can enhance the ability of network for feature extraction.

In order to satisfy the sparse limit, penalty factor in Eq. (4) is needed for the layers. Making the average activation p of hidden layer neurons take close to zero value. In this paper, we choose $K_L(p||p_j')$, hidden layer neurons j, as punishment factor.

$$K_L(p||p_j') = \sum_{j=1}^{s_2} p\log\frac{p}{p_j'} + (1-p)\log\frac{1-p}{1-p_j'} \tag{4}$$

The total cost function of the system can be expressed as Eq. (5). In the Eq. (5), s_2 is the number of hidden layer neurons, and β is the weight that controls the size of the penalty factor. With the increasing difference between p and $p_j{}'$, the value of the penalty factor rises sharply.

$$J_{All}(W,b) = J(W,b) + \beta\sum_{j=1}^{s_2} K_L(p||\,p_j') \tag{5}$$

Deep sparse autoencoder (DSAE) is composed of multiple SAEs. Stacked artificial neural network can be used as a classifier to identify the characteristics of the training set. Therefore, DSAE algorithm based on SAE and stack theory can extract feature factors of high-dimensional data. DSAE can enhance the robustness of the SAE algorithm and extract nonlinear features.

For gear wear fault diagnosis, the six-layer DSAE network consisting of 5 hidden layers and 1 classifier is designed. As shown in Fig. 1, DSAE puts the superior hidden layer features to the junior hidden layer as the input vector. And then using gradient descent method trains each of SAE network layers step by step. And through many iterations it updates parameter matrix of each of network layer to extract the feature information of the input signals. The updating formula of parameter matrix is Eq. (6). In Eq. (6), μ represents the learning rate. Back-propagation algorithm is used to fine-tune the weight and bias of the whole network.

$$\begin{cases} W_{ij}^{(l)} = W_{ij}^{(l)} - \mu\frac{\partial}{\partial W_{ij}^{(l)}}J_{All}(W,b) \\ b_{ij}^{(l)} = b_{ij}^{(l)} - \mu\frac{\partial}{\partial b_{ij}^{(l)}}J_{All}(W,b) \end{cases} \tag{6}$$

B. The Data Augmentation Algorithm

In this paper, two kinds of data enhancement algorithms are proposed. The augmented algorithms include augmentation of stack denoising autoencoders and raw vibration data shifting. The former mainly improves the generalization ability of the model, while the latter mainly improves the robustness of the model.

1) Augmentation of stack denoising autoencoders.

In addition to the low dimensional feature extractions from the original data, AE can also reduce the original data noises. This ability enables the decoded features from the AE reflect input data natural characteristics of the original signal more precisely. Therefore, the signal decoded by the AE can be used as a more precise expression of the original signals. That is the theoretical foundation of augmentation of stack denoising autoencoders. The actual measured vibration signals contains a lot of noise. Reducing the noise of the original signals can effectively improve the robustness of data. The original signals X is encoded into a lower dimensions. And the compressed signals are restored through the decoder process. The characteristics of the decoded signal can reflect the characteristics better than the original signal and having stronger robustness. In the process of compression and decompression, AE actually makes the original signal more pure. This method enables the training model to have a stronger generalization ability.

The augmentation of denoising autoencoders algorithm is shown in the Fig. 2. At first, original vibration signals X are put into Autoencoder I to get signals X' through the process of noise reduction. X' has better robustness and can reflect the core characteristics of the original signals. The original signals X and decoded signal X' are simultaneously put as an input signal to Autoencoder II. In this way, data that are 4 times the size of the original signals can be generated. So the generalization ability of the model is further enhanced. Input signals of Autoencoder III are the sum of the input signals and decoded output signals of Autoencoder II. This gives the input data 8 times the size of the original signals. More data samples will be helpful to improve the generalization ability of the AE.

2) Augmentation of data shifting.

AE network is the fully connected network. Thus, it can be regarded as a disorder. So the effect of the start point of the sample data is not obvious, which gives the possibility of augmentation of the raw vibration data shifting translation. This paper applies this data enhancement thought to the gear raw vibration data. The main way of raw vibration data shifting augmentation is to choose different starting position, and intercepting a certain length of sample data from its starting point. In this way, more samples can be obtained through the limited data length. In this way, we can enlarge the number of samples and increase the robustness of the training model.

The total number of sample points of the collected data is $X = \{X_1, X_2, X_3, X_4, X_5, \dots, X_R\}$. In a period of gear rotation, the sample characteristic number of the sensor extracted is N. Sample characteristics of a total length of the original data are R. The starting position is ε. k is a decimal between zero and one. So the initial position ε, in the Eq. (7), is the product of the total length R and the coefficient k. All the points in this data set are X_i In the Eq. (7).

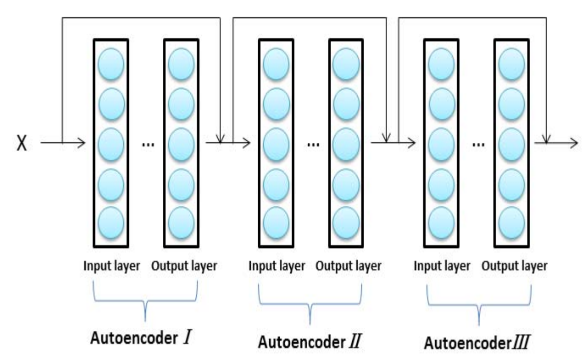

Fig2. The principle of augmentation of stack denoising autoencoders.

$$\varepsilon = k\mathrm{R}, \quad 0 < k \le \frac{R-N}{R} \tag{7}$$

$$X_i = \{X_{\varepsilon,}\, X_{\varepsilon+1,} \dots\dots X_{\varepsilon+N}\} \tag{8}$$

In order to have the augmented gear raw vibration data with better robustness, the data of each group are generally selected in the same interval so that there is no unbalance between the data of different categories. The unbalanced data extraction is not conducive to the training of the AE to obtain the desired low dimensional features.

3) The general procedure of the proposed method

In this paper, gear wear fault detections take advantage of the raw vibration signal by ADSAE. The flowchart is shown in Fig.3. The procedure consists of the following steps:

Step 1. Collect the raw vibration signal from the gearbox through the acceleration sensor.

Step 2. The raw vibration signals are divided into three parts: training set, validation set, and testing set.

Step 3. Train artificial neural network.

3.1 Use augmented algorithm to training data in order to improve the generalization ability and robustness of the model.

3.2 Deal with the augmented sample data to make all the data in the sensitive region of the activation function.

3.3 Train the training data in the first layer of the AE, using Eq. (2) to calculate the output of the layer.

3.4 Use Eq. (4) to calculate the loss function and adjust the value of the parameters with error back propagation algorithm.

3.5 Train the second hidden layer using the same method as training the first hidden layer.

3.6 Train the third hidden layer and stack sequentially all the layers together.

Step 4. To connect the most middle hidden layer to the softmax layer, and fine-tuning the parameters.

Step 5. Verifying ADSAE algorithm with gear raw vibration data.

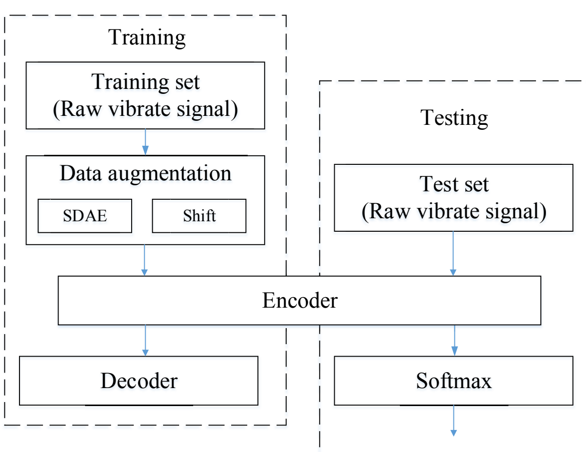

Fig.3 The general procedure of the proposed method.

III. GEAR TEST EXPERIMENTAL VERIFICATION

To verify the effectiveness of the ADSAE using an experiment. The experiment was performed on a gearbox test rig. Raw vibration signals with 6 different gear wear states were collected during the experiment. The gearbox test rig is shown in Fig. 4. It consists of two 45KW Siemens servomotors, one of which is the driving motor and the other motor is the load motor. An acceleration sensor was installed on the surface of the gearbox housing. The main parameters of the gearbox are in Table 1.

The gear speed was set as 800 RPM and 400 Nm torque was used in this experiment. Table 2 shows 6 gear wear states. Condition 1 was a normal gear. In Condition 2, the middle tooth had about 10% worn, and the other two teeth were normal. Condition 3 had about 30% worn in the middle gear tooth, and the other two teeth were also normal. In Condition 4, the middle gear was about 50% worn, with about 10% wear on the upper tooth, and the other tooth was normal. About 50% of the middle gear in Condition 5 was worn, and about 10% of the two teeth to it were worn. In Condition 6, about 50 % of the middle gear was worn, about 30% of the upper tooth were worn, and about 10% of the below tooth.

The raw vibration signals extracted by the acceleration sensor, and as shown in the Fig. 5. From the top to the bottom the figures respectively correspond to the raw vibration signal amplitudes of condition1, 2, 3, 4, 5, and 6 in Table 2. It can be seen from the Fig. 5 that condition 3 has relatively obvious convex points, and condition 4 has relatively larger amplitude than others. The rest of the conditions are not significantly different.

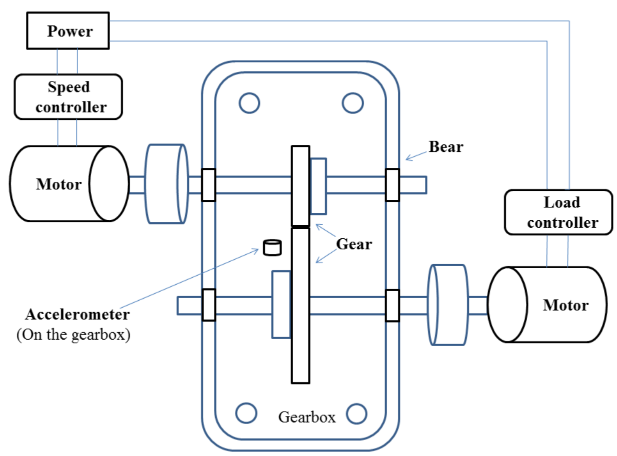

Fig.4 Schematic representation of the gearbox test rig.

TABLE 1. THE MAJOR PARAMETERS OF GEARBOX

Gear Parameter	Driving Gear	Driving Gear
Tooth number	72	40
Module	3mm	3mm
Pitch diameter	120mm	120mm
Base circle diameter	202.974mm	112.763mm
Pressure angle	20°	20°
Tooth width	85mm	85mm

TABLE 2. THE APPROXIMATE PERCENTAGE OF WEARING AREA

Wearing percentage	Upper tooth	Middle tooth	Lower tooth
Condition1	Normal	Normal	Normal
Condition2	Normal	10%	Normal
Condition3	Normal	30%	Normal
Condition4	10%	50%	Normal
Condition5	10%	50%	10%
Condition6	30%	50%	10%

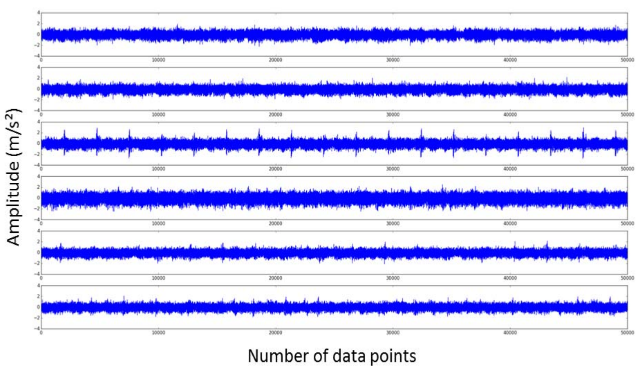

Fig.5 Raw vibration signals of gear wear fault conditions.

Each data sample set had a total of 768 data points. The structure of the ADSAE model designed for the training data sets was: 768-384-192-96-192-384-768. Training set had 1.12 millions samples, validation set had 17.5 thousands samples, testing set had 17.5 thousands samples. The optimizer used stochastic gradient descent (SGD) algorithm, and loss used categorical cross-entropy. The initial value of learning rate was set as 0.01, decay as 1e-6, momentum as 0.9, and activation as relu.

IV. RESULTS AND DISCUSSIONS

Fig. 6 shows the comparison of accuracy and loss between the ADSAE algorithm and the DSAE algorithm for gear wear fault diagnosis. As one can see in Fig. 6 (a), the accuracy of the blue ADSAE algorithm is closes to 100% for gear wear fault diagnosis. Its fluctuation is small, and more stable. The accuracy of the DSAE curve accuracy is about 72%. As seen in Fig. 6 (b), the loss curve of the ADSAE algorithm is close to zero, which is significantly lower and more stable than that of the DSAE algorithm. To sum up, ADSAE algorithm is better than DSAE algorithm for gear wear faults detection. ADSAE can effectively improve the generalization ability and robustness of gear fault wear diagnosis model.

Table 3 provides the accuracy comparison between the ADSAE algorithm and the DSAE algorithm's for training set, verification set, and testing set. As seen in Table 3, the accuracy of the ADSAE can reach up to 99.31% for training. The accuracy for verification and testing set is close to that for training. This result indicates that the ADSAE algorithm can achieve good gear wear fault diagnosis results. On the contrary, the accuracy of the DSAE algorithm for verification and testing is more different than that for training. Therefore, ADSAE's gear wear fault diagnosis results are more accurate, with better generalization ability and robustness using raw vibration signals.

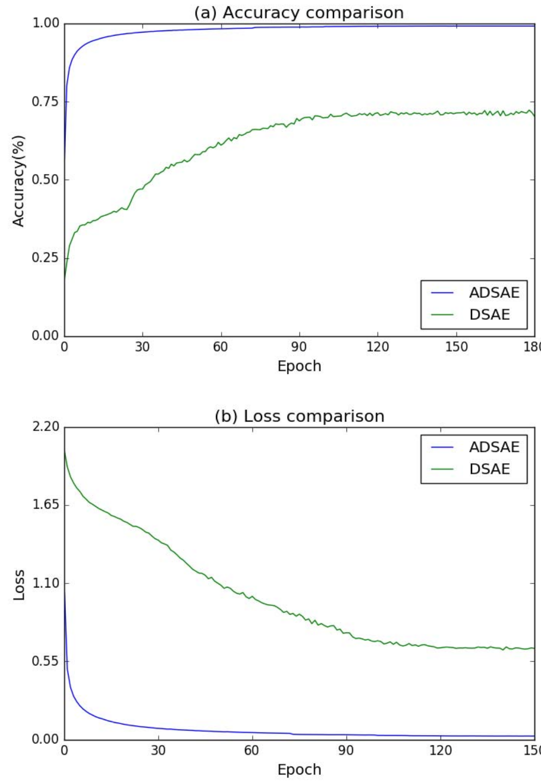

Fig.6 The comparison of accuracy and loss between the ADSAE algorithm and the DSAE algorithm for gear wear fault diagnosis.

TABLE 3. THE ACCURACY COMPARISON BETWEEN THE ADSAE ALGORITHM AND THE DSAE ALGORITHM'S FOR DATA SETS

Accuracy	Training set	Validation set	Testing set
ADSAE	0.9931	0.9925	0.9914
DSAE	0.7230	0.6638	0.6642

Fig. 7 shows the confusion matrix by the ADSAE algorithm for the testing set. As seen in Fig. 7, condition 4 got all the correct classifications, other condition classification results are all above 98%. The majority of predicted false gear wear faults are very close to the actual gear faults, and the probable cause is that the original vibrate signals of gears are too close. For the sample size of the testing set is very large, the original vibration signal data cannot be detected absolutely correct. But the results still showed that ADSAE algorithm has a good ability of fault diagnosis with the gear raw vibration signal.

The ADSAE model obtained from the previous training was used to reduce the feature dimension to 3 dimensions and drawing the Fig. 8. In order to be able to identify the gear wear faults corresponding the color, the corresponding gear wear condition number is marked on the corresponding color. As can be seen from Fig. 8, with the increasing severity of the gear wear faults, the features of the faults have obvious trend. It is obvious that with the fault degree increasing, the features along

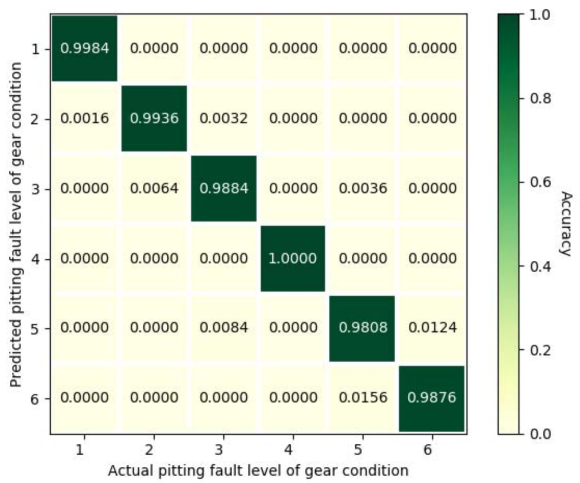

Fig.7 The confusion matrix by the ADSAE algorithm for the testing set.

the *x*-axis direction gradually decreases, along with the *y*-axis direction gradually decrease, and along the direction of the *z*-axis gradually increases. Condition 1, e.g., is a normal gear, and its feature points mainly focus on the small values of the *x*-axis and *y*-axis. Condition 7 is the gear with the most serious fault degree, and its feature points mainly focus on the large value of *x*-axis and small value the *y*-axis. The intermediate gear wear fault features are also concentrated in specific areas according to this trend. All the conditions in the *z*-axis show a trend that can be recognized. As can be seen from the Fig. 8, the gear wear faults with large *z*-axis values are more serious than those with smaller *z*-axis values.

These results provide substantial evidence that the ADSAE algorithm used for gear wearing fault diagnosis with raw vibration signals is effective. It should be noted that this method has verified with only six conditions of gear wear faults. Future research will extend the current work into different types of gear wear faults.

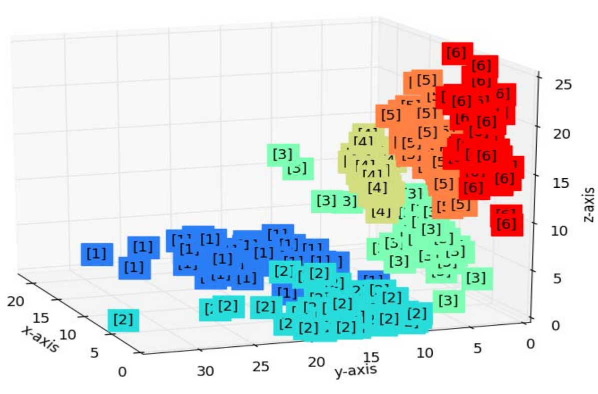

Fig.8 The visualization of three-dimensional features of gear wear conditions.

V. Conclusions

In this paper, a novel method ADSAE for gear wear fault diagnosis using limited raw vibration data was presented. The method has three main advantages. First and foremost, the method only needs relatively limited sample data. The second benefit is that the method does not need to use extra preprocessing and only takes advantage of raw vibration data. This benefit can effectively improve the fault diagnosis efficiency. Additionally, the method mainly uses unsupervised learning. The experimental results have shown that the ADSAE algorithm can successful diagnose the gear wear faults and can show the obvious trend according to the severity of gear wear faults.

Acknowledgement

Sponsor: NSFC grant no. 51675089.

References

[1] K. Aharamuthu, and E. Ayyasamy. "Gear fault diagnosis using vibration signals based on decision tree assisted intelligent controllers." Journal of Vibroengineering 15.4, 2013, pp.1826-1837.

[2] I. Attoui, A. Boudiaf, N.Fergani, B. Oudjani, N. Boutasseta, and Deliou. "Vibration-based gearbox fault diagnosis by DWPT and PCA approaches and an adaptive neuro-fuzzy inference system." International Conference on Sciences and Techniques of Automatic Control and Computer Engineering IEEE, 2016, pp.234-239.

[3] M. Cerrada, F. Pacheco, D. Cabrera, G. Zurita, & C. Li, "Hierarchical feature selection based on relative dependency for gear fault diagnosis." Applied Intelligence 44.3, 2015, pp.1-17.

[4] N. Li, C. Liu, C. He, Y. Li, and X. Zha. "Gear fault detection based on adaptive wavelet packet feature extraction and relevance vector machine." ARCHIVE Proceedings of the Institution of Mechanical Engineers Part C Journal of Mechanical Engineering Science 1989-1996 (vols 203-210) 225.11, 2011, pp.2727-2738.

[5] D. Wang. "K -nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited." Mechanical Systems & Signal Processing s 70–71, 2016, pp.201-208.

[6] G. Jiang, P. Xie, S. Du, Y. Guo, and Q. He. "A new fault diagnosis model for rotary machines based on MWPE and ELM." Insight - Non-Destructive Testing and Condition Monitoring 59.12, 2017, pp.644-652.

[7] H. Li, J. Zhao, X. Zhang, and H. Teng. "Gear fault diagnosis and damage level identification based on Hilbert transform and Euclidean distance technique." Journal of Vibroengineering 16.8, 2014, pp.4137-4151.

[8] W. Shang, X. Zhou, and J. Yuan. "An intelligent fault diagnosis system for newly assembled transmission." Expert Systems with Applications 41.9, 2014,: pp.4060-4072.

[9] Z. Zhang, Y. Wang, and K. Wang. Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. Springer-Verlag New York, Inc. 2013.

[10] W. Zhu, L. Yang, J. Guo, Y. Zhou, C. Lu, W. Zhu, et al. "A study on crack fault diagnosis of wind turbine simulation system." International Conference on Reliability, Maintainability and Safety IEEE, 2015, pp.53-57.

[11] G. Cheng, Y. Cheng, L. Shen, J. Qiu, and S. Zhang. "Gear fault identification based on Hilbert–Huang transform and SOM neural network." Measurement 46.3, 2013, pp.1137-1146.

[12] G. Meng, C. Wang, and C. Zhu. "Multiple-domain manifold for feature extraction in machinery fault diagnosis." Measurement 75, 2015, pp.76-91.

[13] Y. Qu, M. He, J. Deutsch, and D. He, "Detection of Pitting in Gears Using a Deep Sparse Autoencoder." Applied Sciences 7.5, 2017, pp.515.

[14] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, and J. Tang et al. "Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images." IEEE Trans Med Imaging 35.1,2016, pp.119-1