



# Establishment of a deep learning network based on feature extraction and its application in gearbox fault diagnosis

QingE Wu<sup>1</sup> · Yinghui Guo<sup>1</sup> · Hu Chen<sup>1</sup> · Xiaoliang Qiang<sup>1</sup> · Wei Wang<sup>1</sup>

Published online: 12 April 2019  
© Springer Nature B.V. 2019

## Abstract

Gearbox is an important part of mechanical equipment. If a fault cannot be timely detected, it will cause significant economic losses. In order to solve the problem of early fault diagnosis quickly and accurately, this paper proposes a feature extraction method by the decomposition of feature value to the waveform of signal, and inputs the extracted feature into the deep learning network established by this paper. Firstly, the input signal is reconstructed and the feature value is decomposed. Secondly, the extracted features are input into the established deep learning network as the deep learning signals to carry out fault diagnosis. Finally, the fault is identified by the established deep learning network. In a number of experiments, to compare with the existing some fault diagnosis methods, such as support vector machine, classical neural network, lifting wavelet and logical regression, the experimental results show that the average accurate recognition rate of the proposed method by established deep learning network based on feature value decomposition to fault diagnosis is 96.65%, its variance is 0.36 and the diagnostic speed is 0.612 s. However, the average accuracy of the best diagnostic method at present is 93.52%, the variance is 0.47 and the diagnostic speed is 0.826 s. It indicates that the proposed method has a good accuracy, stability and fastness.

**Keywords** Gearbox · Fault diagnosis · Feature value decomposition · Deep learning network

## 1 Introduction

At present, the gearbox is an important part of some mechanical equipment, and has been widely used in automotive transmissions, heavy machinery, ship building and other equipment. As a key component of mechanical equipment, once the fault cannot be found in time, the consequences caused by fault are not only mechanical damage, more likely to cause sig-

---

✉ QingE Wu  
wqe969699@163.com

✉ Wei Wang  
wwang@zzuli.edu.cn

<sup>1</sup> School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China

nificant casualties and significant economic losses, so timely monitoring and fault diagnosis to the gearbox are of great significance.

The original signal of the gearbox is a non-stationary signal, because the original signal collected are the signal that the gearbox is running and the signal that suddenly changes over time. The effective feature information is the key to the fault diagnosis. Gearbox fault diagnosis had experienced a long time from the early manual detection to the subsequent computer detection until today's intelligent detection. Now, there are a lot of artificial intelligence detection methods. However, the detection methods that are most widely used are the support vector machine (SVM) and neural networks. From development to today, domestic and foreign researchers in the gearbox fault diagnosis had proposed artificial neural network, support vector machine and other scientific effective identification methods. Recently, Bruna and Mallat (2013) used a wavelet to carry out quantitative analysis for the deep network structure, which was an important exploration in this direction. Wang et al. (2016) discussed the characteristics of the convolutional neural network for image processing, outlined two typical saliency detection algorithms by the use of convolutional neural networks, and analyzed the future development of deep learning. Kankar et al. (2011) used a combination of multiple logical regression and wavelet packet transform (WPT) to diagnose faults. Huang et al. (2011) proposed a multi-fault diagnosis method based on improved SVM, which used a radial basis kernel function and empirical mode decomposition method. Wang et al. (2015) used the neural network to achieve a quantitative accurate identification for the crack fault of rotor. Jiang and Yuan (2014) improved the traditional wavelet neural network model, which could effectively diagnose the gearbox fault. Deep belief networks (DBN) were proposed by Hinton et al. (2006), which was constructed on the basis of the restricted Boltzmann machine (RBM) (Roux and Bengio 2008). It is theoretically proved that RBM can fit any discretely distributed signal as long as there were enough network units. In this algorithm, a number of RBMs were stacked to form a DBN. In the DBN, the upper-level network could extract a higher-order feature from the lower-level network, which was an unsupervised deep learning algorithm. The training algorithm to RBM is based on using CD (Zhang et al. 2015) algorithm. The deep model obtained by training large-scale data (de Mendivil 2018; Yang et al. 2015) had been also studied for better fault diagnosis. The above several methods can effectively diagnose the fault, but these methods have their own disadvantages in the multi-classification or precision.

Signals collected by different means or affected by various other factors may mix with some noise. Feature value decomposition can well extract the active components of the signals, and can effectively filter out part noises of the original signals. Therefore, this paper proposes a feature extraction method by feature value decomposition (FVD) based on the characteristics of the signals, and also plays a role of dimension reduction. Then, the features that are extracted by FVD are input into deep learning network (DLN) as the deep learning signals. Finally, the fault classification and identification are achieved by the intelligent diagnostic method, which is called the FVD-DLN method. Experiments show that this method is effective, accurate and fast for fault diagnosis of gearbox of automobile transmission. For more clearly displaying idea in this paper, the flow frame of the study is shown in Fig. 1.

The novelty of the presented work mainly includes three aspects, i.e., First, a feature extraction method based on the singular value decomposition of the signal is proposed. Second, the learning method of grading, layering and classification based on the characteristics of the signal is given. Third, the design of neural network structure based on above given deep learning method is performed, namely, connection, depth, calculator, learning training of network and other design.

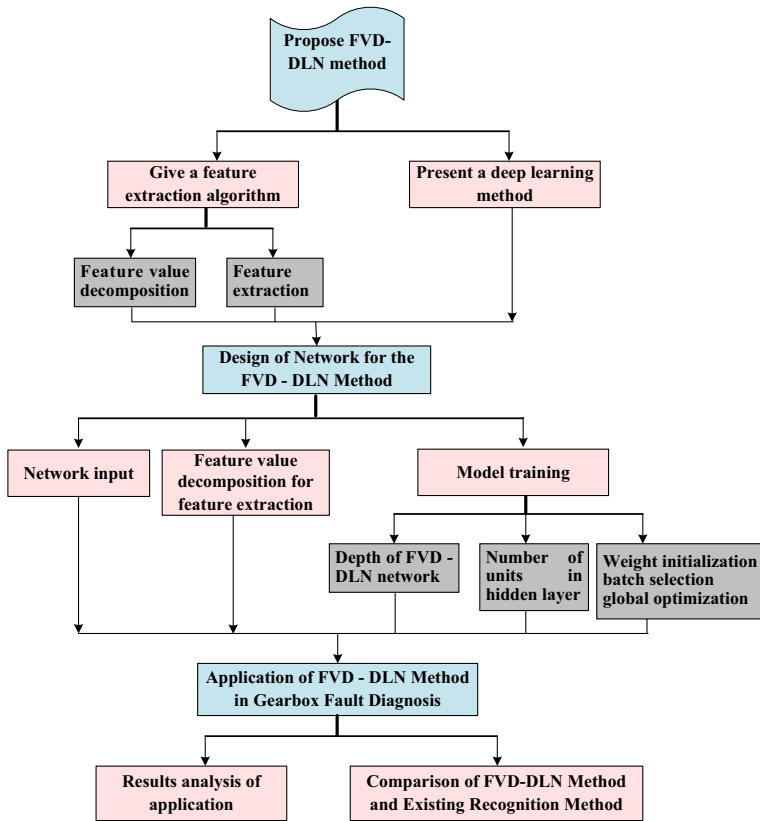


Fig. 1 The flow frame of idea

## 2 FVD-DLN method proposed

### 2.1 Feature extraction algorithm

The first novelty of this paper is: In previous literatures (Jiang and Zhang 2018; Wu et al. 2015, 2016; Li et al. 2018; Cho and Yu 2018), the feature extraction to target was achieved based on the relationship between signals or other features of the signal, such as variance, error, amplitude, and so on. However, this paper proposes a feature extraction method based on the singular value decomposition of the signal for the characteristics of the signal itself, and the decomposition of each parameter or decomposition of coefficient of parameter is implemented. Such feature extraction method can achieve signal denoising and dimensional reduction of large size data.

#### 1. Feature value decomposition

In this section, the Hankel matrix is used, so here it is simply introduced as follows:

Hankel matrix is a  $m \times n$  symmetrical matrix, and the total different number of elements of the matrix is  $m + n - 1$ . Moreover, two numbers on the connected line obtained by connecting between the second element of the first row and the second element of the first column in the matrix are the same, and the numbers on the same lines that parallel to this connected

Fig. 2 A 5 × 5 Hankel matrix

red	orange	yellow	green	cyan
orange	yellow	green	cyan	blue
yellow	green	cyan	blue	indigo
green	cyan	blue	indigo	purple
cyan	blue	indigo	purple	pink

line are the same, i.e., this matrix can be intuitively shown in Fig. 2. In Fig. 2, the squares of the same color represent the same number. For example, a 4 × 5 Hankel matrix is given as follows:

$$D = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 5 & 6 & 7 \\ 4 & 5 & 6 & 7 & 8 \end{pmatrix} \tag{1}$$

The general form of Hankel matrix is shown in the formula (2).

Assume the sampling sequence of the sample data for the running gearbox is  $y = \{x_1, x_2, \dots, x_k\}$  and its length of the sequence  $y$  is  $k$ .  $x_i (i = 1, 2, \dots, k)$  is the sample signal data. A  $m \times n$  matrix  $D$  can be constructed by  $y$ , and then the special Hankel matrix is used to carry out a construction for the discrete signal  $y$ . The constructed matrix  $D$  is given as follows:

$$D = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_{n+1} \\ \vdots & \vdots & & \vdots \\ x_m & x_{m+1} & \cdots & x_k \end{pmatrix} \tag{2}$$

where  $D$  is a constructed matrix with a  $m \times n$  dimension.  $x_i (i = 1, 2, \dots, k) (i = 1, 2, \dots, k)$  is input signal. According to the characteristics of Hankel matrix, there is  $k = m + n - 1$ . The feature value decomposition is implemented to  $D$  as follows:

$$D = USQ^T \tag{3}$$

where  $U, Q$  is an orthogonal matrix, and  $s$  is a diagonal matrix. For any real symmetric matrix  $D$ , there are always two orthogonal matrices  $U, Q$ , and then  $D$  is turned into a diagonal matrix, i.e., formula (3). Their expressions are as follows:

$$\begin{aligned} U &= [u_1, u_2, \dots, u_m] \\ Q &= [q_1, q_2, \dots, q_n] \\ D &= [\text{diag}[\lambda_1, \lambda_2, \dots, \lambda_r], 0], \quad r = \min\{m, n\} \end{aligned} \tag{4}$$

where  $U$  is a  $m \times m$  matrix;  $Q$  is a  $n \times n$  matrix;  $\lambda_1, \lambda_2, \dots, \lambda_r$  are feature values obtained by decomposing the matrix  $D$ , and their relationship is  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$ . These feature values obtained by decomposition can be used as extracted features.

## 2. Fault feature extraction of gearbox

For the operation signal  $y = f(x)$  of the mechanical gearbox, the sequence that consists of signals  $x_1, x_2, \dots, x_k$  can be obtained by sampling and its length is  $k$ . A special Hankel matrix is structured for the sequence. The feature value vector  $z = (\lambda_1, \lambda_2, \dots, \lambda_r)$  is obtained by feature value decomposition, as is the extracted features of operation signal. According to the characteristics of feature value decomposition, the larger value of the feature is the effective part of signal that is the reflection of the fault, and the small part of the feature value is the embodiment of the noise of signal. So, the feature values are arranged from large to small and the larger feature value is taken as the fault feature.

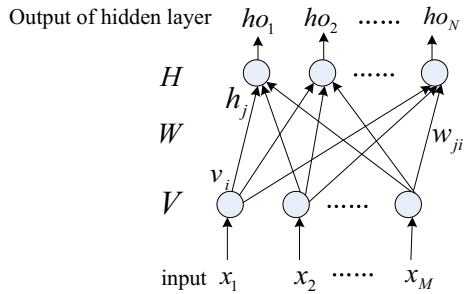
### 2.2 Deep learning method

The second novelty of this paper is: the learning method of grading, layering and classification based on the characteristics of the signal is given. That is, the pre-stage achieves the input according to the characteristics of the signal, and the normalization processing such as unified units, input by dimension or semantic field is implemented, etc.; The intermediate processing stage is a hidden layer. The hidden layer has several layers, and the learners of each layer are different. For example, the first layer performs redundancy removal processing on the data, and the second layer performs feature detection, extraction and labeling, and so on; The output stage is the design of detector and recognizer of the output layer, and the output result. For example, the design of detector and recognizer device is some calculations of maximum membership, recognition rate, variance, error comparison, etc. The output results are the effective values of the target or target parameters, and the output form is data, tables or images, and so on.

The deep learning algorithm is composed of multiple non-linear monolayer networks superimposed, which has a strong mathematical expression. Deep belief networks (DBN) were proposed by Hinton et al. (2006), which was constructed on the basis of the restricted Boltzmann machine (RBM) (Roux and Bengio 2008). It is theoretically proved that RBM can fit any discretely distributed signal as long as there were enough network units. In this algorithm, a number of RBMs were stacked to form a DBN. In the DBN, the upper-level network could extract a higher-order feature from the lower-level network, which was an unsupervised deep learning algorithm. RBM was composed of a visual layer  $V$  and a hidden layer  $H$ , which its structure was characterized by interconnection between layers and is independent interior layers. The structure of RBM is shown in Fig. 3. The bottom neurons  $V = (v_i)_{M \times 1}$  in the graph are a second order visual layer.  $W = (w_{ji})_{N \times M}$  are the connected weights between two layers of the interconnect layer. The upper neurons  $H = (h_j)_{N \times 1}$  is a hidden layer. Where, assuming the number of neurons in the hidden layer of RBM model is  $N$ , and the number of neurons in the visual layer is  $M$ .

RBM is a probability model. The relation between the two layers is completely independent and interconnected. Therefore, in the case when the state is known in visual layer, the state of each neuron in the hidden layer can be obtained by the conditional probability  $p(h|v) = \prod_{j=1}^N p(h_j|v)$ . Similarly, in the case when the state is known in the hidden layer, the state of visual layer can be obtained by the conditional probability  $p(v|h) = \prod_{j=1}^M p(v_j|h)$ . Where,  $h, h_j \in H, v, v_i \in V$ . In RBM, the number of neurons in the visual layer is set as the characteristic dimension of the training data. The number of neurons in the hidden layer needs to be set in advance. The connection weight  $W$  that needs to be trained is a  $N \times M$  matrix. The biased vectors are set as  $a$  that is a  $M$  dimension column vector in the visual

Fig. 3 RBM model



layer and  $b$  that is a  $N$  dimension column vector in the hidden layer, respectively. The training algorithm to RBM is based on using CD (Zhang et al. 2015) algorithm, the detailed steps are as follows:

- P1 Initialize the connection weight  $W$ . According to the training rules for the network, define the incentive function is  $h = W * x + a$  to determine the input value of the hidden layer, where  $x = (x_1, x_2, \dots, x_M)$  is the network input signal;
- P2 The incentive function  $h$  obtained by the first step is the input to the hidden layer. The open state function of the hidden layer neuron, i.e., the probability of the open state is defined by  $p(h_j) = \frac{1}{1+e^{-h_j}}$  based on the network stability requirement. A logistic function is used as the incentive function  $ho_j$  of output of the hidden layer. The probability value of each hidden neuron in the open state is calculated the by the above calculation formula. 0 indicates the probability of closing state, and 1 indicates the probability of opening state. Because each neuron is binary, the specific switch state of each neuron can be determined. It is necessary to compare the open probability of each neuron obtained with a randomly selected value  $c = UI(0, 1)$  in the uniform distribution from 0 to 1, and the comparative formula is given as follows

$$ho_j = \begin{cases} 1, & p(h_j = 1) \geq c \\ 0, & p(h_j = 1) < c \end{cases} \quad (5)$$

- P3 Determine the opening state of the neurons in visual layer. The calculation method is similar to the second step.

The probability distribution of the training samples obtained by the CD algorithm is the optimal distribution. The greedy algorithm is used in training process of DBN, that is, each RBM is trained to layer by layer, and thus the best results of each layer are obtained for each RBM. However, the global results obtained are suboptimal. The next step requires a global fine tuning, then the global optimal results can be gotten. The specific training process (Ma et al. 2015) of DBN is as follows:

- P1 Training the first RBM model;
- P2 Fixed the weight and the biased vectors obtained by the first step, and let the output as the input of next RBM;
- P3 Repeat the corresponding first two steps according to the number of hidden layers designed;
- P4 Carry out a global optimal tuning.

### 3 Design of network for the FVD-DLN method

The third novelty of this paper is: the design of neural network structure based on above given deep learning method is performed, namely, connection, depth, calculator, learning training of network and other design. The design of the connection structure between the layers, for example, single connection, multiple interconnection and sparse connection; Design of each stage and the number of layers at each stage, such as the design of the input layer, several hidden layers, output layers, etc.; The design of calculator in each layer, for example, preprocessing, feature extraction, maximum membership, recognition rate, variance, error comparison, etc.; Learning training is designed according to feedback, feedforward, and so on.

#### 3.1 Network input

The input value of the network is the original signal, which is ready to be processed. In here, the input signal is the fault data of gearbox. It needs to collect discrete signals. The fault data of the gearbox used by experimental verification is provided by a car repair shop. These data come from a test drive system.

We carried out some experiments on fault. These faults that were some single point faults were manually determined by electrical discharge machining. Perform the statistics on the frequency of gearbox faults. Statistics show that 60% of the gearbox fault is caused by the gear fault, so the fault diagnosis of gear is only discussed in here. For the fault of the gear, several features here are selected in the frequency domain. The more obvious gear faults in the frequency domain are at the edge band at the meshing frequency. Therefore, in the selection of the characteristic signal in the frequency domain, the amplitude  $A_{i,j1}$ ,  $A_{i,j2}$  and  $A_{i,j3}$  at the edge frequency band  $f_s \pm n f_z$  of the 1, 2, 3 axes at 2, 4, 6 gears are extracted. Where  $f_s$  is the meshing frequency of the gear;  $f_z$  is the transfer frequency of the axis;  $n = 1, 2, 3$ ,  $i = 2, 4, 6$  indicate the position of gear;  $j = 1, 2, 3$  denotes the number of axis. Because there are two pairs of gear meshing at the 2nd axis and the 3rd axis, 1 and 2 represent two meshing frequencies, respectively. In this way, the input of network is a 15-dimensional vector. These data have different units and orders of magnitude, so they should be normalized before being inputted the FVD-DLN network. In this section, the number of samples provided is nine sets of data and each data is a 15-dimensional vector. Here, the nine sets of data take from the rotational speeds of three different gears, which are 6120.09 rpm, 8670 rpm and 1705.5 rpm, respectively, and the number of teeth of three gears is 17, 12 and 61 correspondingly. Table 1 shows the nine sets of data of input vector, which are the normalized sample data.

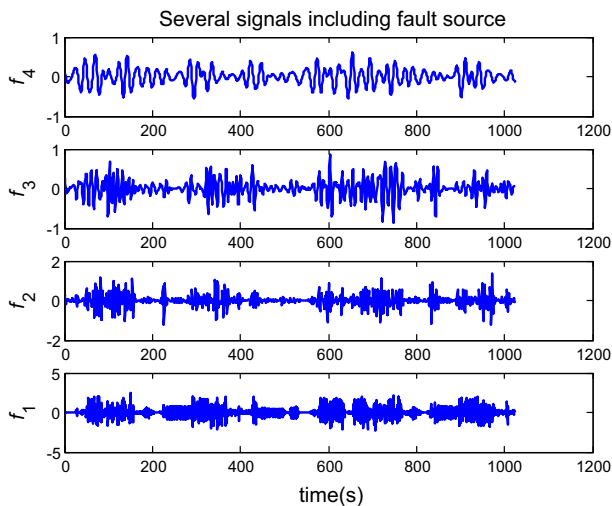
The fault is a given single point fault processed by human through electrical discharge machining. The sampling frequency is 5 s, 1.4 s, 1 s, respectively. Fault data are no fault, tooth root's crack, fracture of gear in turn, as shown in Fig. 4.

In Fig. 4,  $f_1$  denotes the signal with no fault in the case of the above three rotational speeds.  $f_2$  and  $f_3$  denote the signals with tooth root's crack at rotational speeds of 6120.09 rpm and 8670 rpm, respectively.  $f_4$  denotes the signal with fracture of gear in the case of the above three rotational speeds. The points of much higher amplitude of wave curve are fault points for signals  $f_2$ ,  $f_3$  and  $f_4$ .

1000 points that appear in Fig. 4 are explained as follows: According to collected signal data, there are several periods about the vibration signal of gearbox. The 1000 time points in Fig. 4 are a period inside the signal.

**Table 1** Sample data of state of gearbox

Serial number of data	Feature samples	State of gear
1	0.2286 0.1292 0.0720 0.1592 0.1335 0.0733 0.1159 0.0940 0.0522 0.1345 0.0090 0.1260 0.3619 0.0690 0.1828	No fault
2	0.2090 0.0947 0.1393 0.1387 0.2558 0.0900 0.0771 0.0882 0.0393 0.1430 0.0126 0.1670 0.2450 0.0508 0.1328	No fault
3	0.0442 0.0880 0.1147 0.0563 0.3347 0.1150 0.1453 0.0429 0.1818 0.0378 0.0092 0.2251 0.1516 0.0858 0.0670	No fault
4	0.2603 0.1715 0.0702 0.2711 0.1491 0.1330 0.0968 0.1911 0.2545 0.0871 0.0060 0.1793 0.1002 0.0789 0.0909	Tooth root's crack
5	0.3690 0.2222 0.0562 0.5157 0.1872 0.1614 0.1425 0.1506 0.1310 0.0500 0.0078 0.0348 0.0451 0.0707 0.0880	Tooth root's crack
6	0.0359 0.1149 0.1230 0.5460 0.1977 0.1248 0.0624 0.0832 0.1640 0.1002 0.0059 0.1503 0.1837 0.1295 0.0700	Tooth root's crack
7	0.1759 0.2347 0.1829 0.1811 0.2922 0.0655 0.0774 0.0227 0.2056 0.0925 0.0078 0.1852 0.3501 0.1680 0.2668	Fracture of gear
8	0.0724 0.1909 0.1340 0.2409 0.2842 0.0450 0.0824 0.1064 0.1909 0.1586 0.0116 0.1698 0.3644 0.2718 0.2494	Fracture of gear
9	0.2634 0.2258 0.1165 0.1154 0.1074 0.0657 0.0610 0.2623 0.2588 0.1155 0.0050 0.0978 0.1511 0.2273 0.3220	Fracture of gear

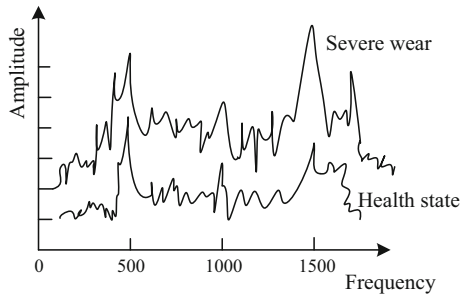
**Fig. 4** Vibration signal of gearbox with fault source

In addition, the relationship between the rotational frequency and the teeth number of each wheel is discussed as follows:

The meshing frequency of gear is equal to the rotational frequency of the gear (rotation per second or Hz) multiplied by its number of teeth. The meshing frequencies of the two gears that mesh with each other are equal. For example, a gear with 17 teeth has a rotational speed of 6120.09 rpm (rotation per minute), and the number of teeth of gear meshed with it is 61, then the latter has a rotational speed of  $(17/61) \times 6120.09 = 1705.5$  rpm. The rotational



**Fig. 5** Meshing frequency and its second harmonic in severe wear



frequency of the two is 102.0 Hz and 28.4 Hz, respectively. It can be verified that the meshing frequency of the two is  $102.0 \times 17 = 28.4 \times 61 = 1734.0$  Hz.

When the gear is working, there are some kinematic process is, i.e., First is to transfer the movement; Second is to change the speed of the movement; Third is to change the direction of the movement, for example, change parallel to vertical or any other angle.

The spectrum characteristics of the meshing frequency of gear and its harmonics when the gear fault occurs are introduced as follows:

1. In the initial state of health, the amplitude of the meshing frequency is the highest, and the amplitude of each harmonic is sequentially decreased.
2. As the gear wear increases, the tooth profile of involute is gradually damaged, and the gear vibration is intensified. At this time, the amplitude of the meshing frequency and each harmonic are gradually increased, and the increase of the amplitude of each harmonic is much faster than that of the meshing frequency.
3. When the wear is severe, the amplitude of the second harmonic exceeds the amplitude of meshing frequency, as shown in Fig. 5.

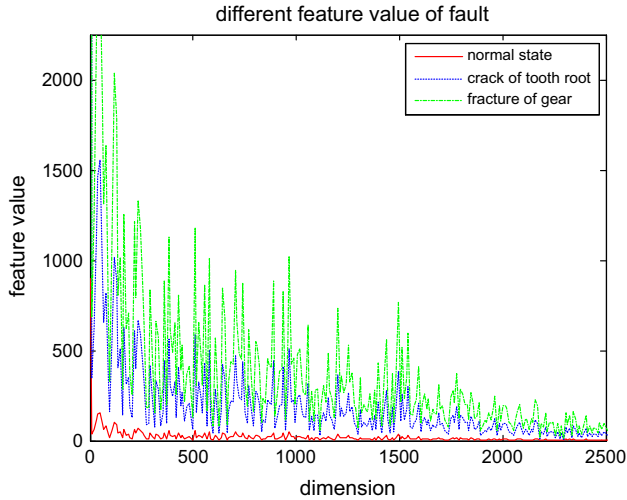
From the above Fig. 5, the degree of wear of the gear can be judged based on the increment of amplitude of the meshing frequency and its harmonic.

### 3.2 Feature value decomposition for feature extraction

After the processed data is obtained, these data are decomposed to be the feature value and the input signal sequence is  $y = \{x_1, x_2, \dots, x_{10000}\}$ .  $x_i (i = 1, 2, \dots, 10000)$  is the  $i$ th data point. Moreover, the range of length of signal  $x_i$  is chosen to be  $[0, 2500]$  in the actual experiment for quick implementation and obvious results through a large number of experiments. Using the Hankel matrix for  $y$  to construct a matrix  $A$ . According to the characteristics of a  $m \times n$  Hankel matrix, since the dimension of the input signal sequence  $y$  is 10,000 in here, and when the number  $n$  of column of the constructed matrix  $A$  is chosen to be 2500, the number  $m$  of row of the constructed matrix  $A$  must be 7501. The dimension of Hankel matrix  $A$  is  $7501 \times 2500$ . The matrix  $A$  is given as follows:

$$A = \begin{pmatrix} x_1 & \cdots & x_{2500} \\ \vdots & \ddots & \vdots \\ x_{7501} & \cdots & x_{10000} \end{pmatrix} \quad (6)$$

Because of the differences between the different signals, the feature values obtained by the formula (6) for feature value decomposition are also different. Figure 6 shows the feature values obtained by different feature decomposition. The feature values of different signal



**Fig. 6** Feature value distribution curve

$x_i$  are different within a same range of length. From Fig. 6, the feature values with faults at normal state are small that is almost zero, however, the feature values with faults at fault states are larger. It can be seen from the Figure that the feature values have obvious differences within the 2000 dimensions of the feature values. So, the 2000 dimensions here are chosen as the feature vectors.

### 3.3 Model training

After the feature value decomposition, the obtained feature vector is normalized and used as the input of DLN. DLN is a probabilistic model that cannot be used directly as a fault diagnosis model, which needs to be designed accordingly. In order to be able to do more classification, the ordinary neural network classifier  $yo_i (i = 1, 2, \dots, O)$  need to be added to the output layer of the model. Here, assume the value of  $yo_i$  is  $v_i$ . If  $v_k = \max_{i=1,2,\dots,O} \{v_i\}$ , i.e.,  $v_k$  is maximum, then the corresponding classifier  $yo_k$  is chosen and marked as max, as a maximum membership classifier. DLN model is shown in Fig. 7. When the model carries out the identification and classification for target, the feature vector is inputted from the input layer, and it is passed along layer by layer from the input layer to the hidden layer, and finally passed to the classifier layer. The classification results are obtained by the classifier layer.

#### 3.3.1 Depth of FVD-DLN network

The number of hidden layers and the number of neurons in each layer are four kinds of structure in the following, i.e., (a), (b), (c) and (d), as shown in Fig. 8.

The data in the box in Fig. 8 is the number of neurons in the layer. Through the four DLN models with different depth, these given input data are classified. By the simulation results, the classification results are different for inputting the same sample data. The classification accuracy of the four different classifiers of DLN with different depth on the test set is shown in Table 2.

Fig. 7 DLN classifier structure

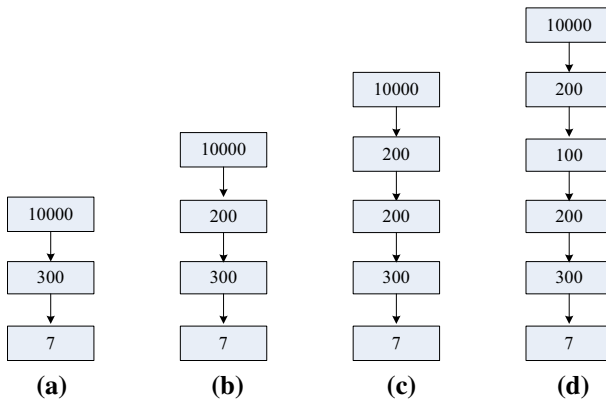
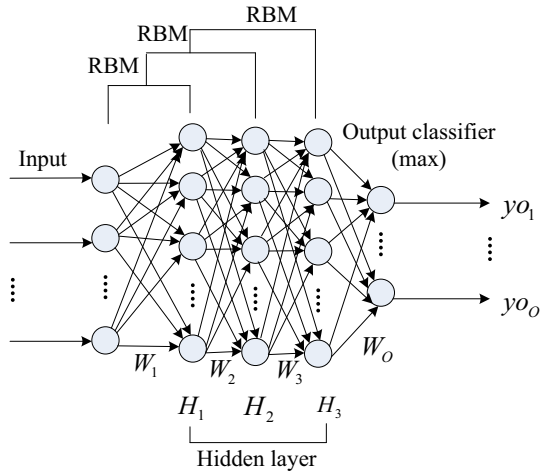


Fig. 8 Four different depth classifiers of DLN

Table 2 Classification accuracy of DLN at different depths

Different classifiers of DLN with different depth	Average classification accuracy (%)
<i>a</i>	90.82
<i>b</i>	92.53
<i>c</i>	98.71
<i>d</i>	96.56

It can be seen from Table 2, the accuracy of the designed DLN classifier is improved when its hidden layer increases from one layer to 3 layers. The accuracy is the highest when the number of layer reaches the third layer. If the number of hidden layers is more than 3 layers, the accuracy will gradually reduce, so the number of hidden layers here is chosen as 3 layers.

### 3.3.2 Number of units in hidden layer of FVD-DLN network

In Hinton (2012), Hinton pointed out that the number of bits required to characterize the category of the sample was usually equal to the number of constraints applied to a parameter by a training sample in the training process of the model, which could be calculated by using the following expression:

$$\text{bits} = -\log_2 n \quad (7)$$

The  $n$  in the formula represents the number of category of sample. After determining the number of bits, let it multiply the capacity of the training set, and select a value less than one order of magnitude as the number of hidden neurons (Zhang et al. 2015). If the training set is highly redundant, the fewer number of hidden neurons can be used. The number of hidden neurons can be calculated by using the following expression:

$$N_p = N_s * \text{bits}/10 \quad (8)$$

where  $N_p$  denotes the number of hidden neurons;  $N_s$  indicates the number of training samples. The numbers of neurons of hidden layers that can be set are 200, 200 and 300, respectively.

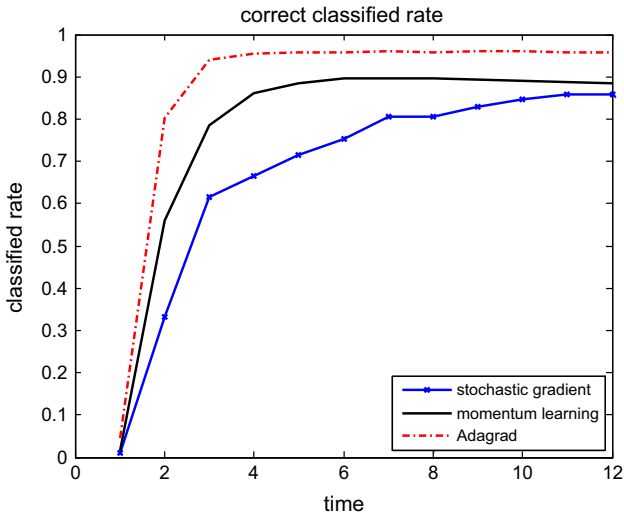
### 3.3.3 Weight initialization, batch selection and global optimization

Some parameters will be trained in the designed structure of DLN in the following. The training is divided into the following three steps:

1. Weight initialization. The training is carried out layer by layer, which is a greedy training. First initialize each parameter, including the weight and bias of each layer. In general, the weights  $W$  between each layer are assigned to the values generated by the normal distribution  $N(0, 0.01)$  (Hinton 2012), and all biases are set as 0.1 learning rate should be from 0 to 1. According to the training experience for sample, that the learning rate is selected as 0.1 is more appropriate in the course of the experiment, so the learning rate here is selected as 0.1.
2. Batch selection of sample. In the course of updating all parameters, select the small batch training data. Although a training sample can also be used to update, the amount of calculation is too large. In the case of small batch training, the choice of capacity of batch is critical. If the batch selected is too small, the training is not sufficient; if the capacity of batch selected is too large, the optimal point can be easily missed. Hinton (2012) suggested that if the categories of data in the data set are equal probability, the desired size of the batch should be equal to the number of categories of samples, and each category contains a sample as far as possible. For categories of non-equal probability data set, choose 10–100 as the size of batch by experience. The experimental data set here is equal probability category, so the size of batch is selected by 7. That is, every 7 samples are selected as a small batch of training set to carry out the training in batches. In the use of small batch data, the parameters also need to make the appropriate changes in the process of updating the parameters (Fischer and Igel 2013). The average gradient is usually used, the expression is as follows:

$$\theta^{(t+1)} = \theta^{(t)} + \varepsilon \left( \frac{1}{B} \sum_{t'=B*t+1}^{B*(t+1)} \frac{\partial \log p(v_{t'}|\theta)}{\partial \theta} \right) \quad (9)$$

where  $B$  is the capacity of small batch;  $p(v_{t'}|\theta)$  is the conditional probability;  $\theta^{(t)}$  is the value of parameter of sample at time  $t$ .



**Fig. 9** Accuracy of global training

3. Global optimization algorithm. Using the supervision algorithm to carry out the whole fine tune and classify. The parameter adjustment is achieved for the whole network by using the training methods to ordinary neural network. The stochastic gradient method (SG), the momentum learning rate (MLR) (Hinton 2012) and the Adagrad variable learning rate (He and Li 2016; Tuerxun and Dai 2015; Lu and Zhang 2016; Su et al. 2017; Su and Zhang 2017; Zhang et al. 2017) are used respectively in the selection of methods. The training results are shown in Fig. 9.

The abscissa in the graph is the number of iterations and the ordinate is the classification accuracy of the parameters. It can be concluded from simulation that the classification accuracy of the parameters is highest when the Adagrad variable learning rate is used for global training. So the Adagrad variable learning rate is chosen for classification.

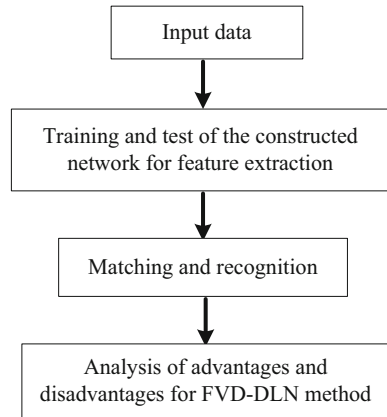
## 4 Application of FVD-DLN Method in Gearbox Fault Diagnosis

According to the deep learning network constructed by the above section, the basic process of fault diagnosis mainly consists of three steps, i.e., the first step is to implement the training and test of the FVD-DLN network for classification and feature extraction for parameters, the second step is to carry out the matching and recognition of fault, the third step is to perform the analysis of advantages and disadvantages of FVD-DLN method for fault diagnosis, as shown in Fig. 10.

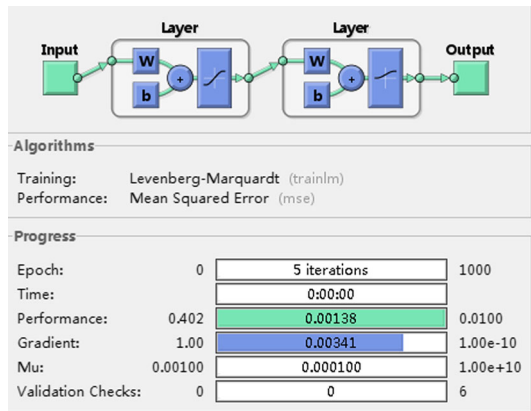
### 4.1 Training and test of the constructed network for actual data

In each test, all nine sets of data given above are input into FVD-DLN network, i.e., the data given in Table 1 in Sect. 3.1. The above gearbox is diagnosed by using the FVD-DLN method. On the basis of the above input samples and test samples, the code is performed and

**Fig. 10** The process of fault diagnosis based on the constructed FVD-DLN network



**Fig. 11** Training process of FVD-DLN network



simulated. The test is repeated three times. The output results of fault type are as follows:

output = [1, 2, 3; 1, 2, 3; 1, 2, 3] is a  $3 \times 3$  matrix, i.e., 
$$\text{output} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}.$$

The numbers 1, 2, 3 in the above matrix represent three different faults, i.e., the number 1 indicates no fault, the number 2 indicates the tooth root’s crack, and the number 3 indicates the fracture of gear. Here, the number of rows of the matrix indicates the number of tests, and the number of columns indicates the category of the fault type. The matrix has 3 rows indicating that the test is repeated three times, and 3 columns indicating that there are three types of faults. It can be seen from this matrix of output result that the output of each test can correctly detect these three types of faults.

The simulation results by FVD-DLN method show that the detection result of this network to fault is correct. That is, the network has successfully diagnosed these three kinds of faults, so the network is valid for gearbox fault diagnosis. The effect of training process for the network is shown in Fig. 11. Thus, when the maximum number of training times is reached, training stops.

In Fig. 11 there are an input layer, two hidden layer and an output layer. The first hidden layer is a preliminary feature extraction layer, i.e., data classification layer, the second hid-

den layer is a further feature extraction layer, and the output layer is a fault matching and recognition layer.

In the hidden layer, the input data are classified firstly, i.e., preliminary feature extraction is implemented, then the further feature extraction is performed based on the above feature extraction algorithm. In the output layer, the fault matching and recognition are carried out. The matching and recognition method is discussed in the following.

### 4.2 Matching and recognition method

In section, the fault recognition discussed denotes the fault diagnosis. These neurons in the output matching layer of FVD-DLN network are some matching filters, such as, matching operator, matching rule, etc. By making use of these matching filters, the degree of similarity in a certain parameter between the identified fault and the known fault is computed. For better identifying fault, it requires to adjust to the weights  $0 \leq w_{pq} \leq 1$  in the output layer. The adjustment for  $w_{pq}$  is: From the display of data obtained in neurons of this layer, if the degree of similarity is greater than the threshold value, the value of weight  $w_{pq}$  is increased, otherwise, its value is decreased, where  $p, q = 1, 2, \dots, M$ . The input  $F_q^t$  in the output matching layer is  $F_q^t = \sum_p w_{pq} E_p^t$  at time  $t$ , where  $E_p^t$  is the output data in the hidden layer.

Based on the extracted characteristic vector of fault, the fault recognition is performed. Since some faults are complicated or uncertain, the characteristic parameters that the characteristic vectors consist of uncertain faults have uncertain characteristic to some extent. Thus, the known characteristic parameters can be thought to be uncertain number, and then both the known characteristic vectors and the diagnosed characteristic vectors are uncertain number vectors. So, the uncertainty fault recognition can be performed well by the constructed FVD-DLN network. The characteristic vector of unknown fault with that of known classificatory fault are compared in output matching layer. Based on the maximum membership principle, if and only if the degree of similarity between the characteristic vector of unknown fault and that of given  $i_0$ th category fault is maximal, the unknown fault is judged to belong to the  $i_0$ th category. The recognition method is given as follows:

#### 1. Calculation for degree of similarity

Assume there are  $n$  categories of fault. The characteristic vector of every fault consists of  $k$  characteristic parameters. Assume the  $i$ th category fault has  $n_{ij}$  values on the  $j$ th characteristic parameter,  $X_{ij}^m$  denotes the  $m$ th number of the  $i$ th category fault on the  $j$ th characteristic parameter, and  $\theta_{ij}^m$  is its mean;  $\tilde{X}_j$  is the observation of the unknown fault on the  $j$ th characteristic parameter, and  $x_j$  is its mean, where  $i = 1, 2, \dots, n, j = 1, 2, \dots, k, m = 1, \dots, n_{ij}$ .

Here, the fault recognition denotes the observation vector which consists of the observed number  $\tilde{X}_j$  of an unknown fault is classified a characteristic vector of a known fault category, i.e., the known characteristic vector are the most similar to the observation vector of an unknown fault.

Assume that  $\mu_{X_{ij}^m}(u)$  and  $\mu_{\tilde{X}_j}(u)$  are the membership function of  $X_{ij}^m$  and  $\tilde{X}_j$  respectively. The Cauchy membership function is chosen based on experience. There is

$$\mu_{X_{ij}^m}(u) = \frac{\sigma_{ij}^2}{\sigma_{ij}^2 + (u - \theta_{ij}^m)^2}$$

and

$$\mu_{\tilde{X}_j}(u) = \frac{\sigma_j^2}{\sigma_j^2 + (u - x_j)^2}$$

where  $u$  is a variable to correspond to  $X_{ij}^m$  or  $\tilde{X}_j$ ;  $\sigma_{ij}$  and  $\sigma_j$  are the ductility degree of  $\mu_{X_{ij}^m}(u)$  and  $\mu_{\tilde{X}_j}(u)$ , respectively.

In order to make sure the type of the unknown fault, the degree of similarity  $d_{ij}^m$  between  $\tilde{X}_j$  and  $X_{ij}^m$  needs to be calculated, i.e.,

$$d_{ij}^m = \tilde{X}_j \circ X_{ij}^m$$

where  $\circ$  represents a synthetic operation, which is the supremum of intersection of  $\mu_{\tilde{X}_j}(u)$  and  $\mu_{X_{ij}^m}(u)$ , that is, it is the height of intersection of two distribution curves between  $\tilde{X}_j$  and  $X_{ij}^m$ , therefore, there is

$$\frac{\sigma_{ij}^2}{\sigma_{ij}^2 + (u - \theta_{ij}^m)^2} = \frac{\sigma_j^2}{\sigma_j^2 + (u - x_j)^2}$$

Then, there is

$$u = \frac{\sigma_{ij}x_j + \sigma_j\theta_{ij}^m}{\sigma_j + \sigma_{ij}}$$

Therefore, there is

$$d_{ij}^m = \frac{\sigma_{ij}^2}{\sigma_{ij}^2 + (u - \theta_{ij}^m)^2} \Big|_{u = \frac{\sigma_{ij}x_j + \sigma_j\theta_{ij}^m}{\sigma_j + \sigma_{ij}}} = \frac{(\sigma_j + \sigma_{ij})^2}{(\sigma_j + \sigma_{ij})^2 + (x_j - \theta_{ij}^m)^2}$$

For  $m$ , take the upper bound value of  $d_{ij}^m$ , so there is

$$d_{ij} = \vee_m d_{ij}^m$$

So, for each known  $i$ th data category, there is a degree of similarity between it and the  $j$ th identified unknown fault, then for any  $j$ , the similarity matrix can be obtained between the unknown fault vector and the known  $i$ th category fault vector, that is,

$$D_i = [d_{i1}, d_{i2}, \dots, d_{ik}]^T$$

## 2. Matching and recognition principle

The method for implementing the matching is given as follows: First, according to the above solving method for degree of similarity, carry out the calculation of degree of similarity between the element in matrix of feature parameters of the identified unknown fault and the corresponding element in matrix of feature parameters of the known standard template library, so that the similarity matrix is obtained. Second, solve the quadratic norm of this matrix. Finally, the type of fault is judged according to the value of matrix norm and the principle of maximum membership, that is, the parameter type corresponding to the maximum norm is the type of fault which belongs to parameter type of the known standard template.



The maximum membership principle is: if  $\exists i_0 \in B$ , it can satisfy the following formula

$$i_0 = \arg \max_i \{\|D_i\|\}$$

where  $\|\cdot\|$  is the matrix norm. Based on the maximum membership principle, the unknown fault is judged to belong to the  $i_0$ th category fault. Therefore, a fault is diagnosed.

### 4.3 Analysis of advantages and disadvantages for FVD-DLN diagnosis method

#### 1. Establish a standard template database

The fault database used in experimental test and the sample database are established as follows:

The gear raw data in the mechanical equipment is an analog signal, so the signal needs to be discretely sampled. The gear fault data used in the experimental verification is collected by our staff of mechanical laboratory. The data comes from a test drive system. The main components of the experimental instrument are a motor, a torque sensor and a dynamometer. The main fault point of the discussion is the crack of tooth root and the fracture of gear.

The rotation speed of the gear is 1705.5 rpm, and about 1000 data points are collected according to the rotation frequency that the gear rotates one cycle, so every 1000 sampling points are used as one sampling result. In order to reflect the data more comprehensively, 10 sampling points are moved backwards on the basis of the first sampling, as the second sampling data, and so on, the number of sampling points is 1000 until the gear rotates one cycle. Thus, 100 sets of sampled data can be obtained.

The history data of gear are inputted to the above constructed depth network and is used as a standard template library that includes multiple types of data, such as data types at health, minor, medium, and severe fault states, etc. For the data at each state, the same number of parameters are discussed, such as variance, error, amplitude, meshing frequency and its amplitude, amplitude of its primary and secondary harmonics, increment of amplitude, frequency and its increment, etc. In this experiment, 10 parameters are chosen. The data collected for each parameter is 1000. Then, each parameter is decomposed according to the feature value decomposition formula (2) so as to constitute a feature matrix.

The specific establishment process for sample library is given as follows:

1. 100 different samples for health and fault data are chosen randomly from health and fault database, and 6 data are chosen randomly from 10 data of each health or fault. Thus, a total of 600 data are obtained.
  2. Five data are chosen randomly from 6 data of each health or fault as a training data set, thus, 500 training data based on 100 different samples are composed for experimental health and fault database. Since the total number of samples is 600 samples, the remaining 100 data compose the test samples database.
  3. The established 500 training samples are carried out the feature extraction and fault recognition. The 10 kinds of features for each fault are acquired, respectively. The training vector for samples contains 5000 characteristic parameters.
2. Classification and extraction of data by feature extraction algorithm

After the fault data to be identified is input into the network, the feature value decomposition is performed. The discrete sequence  $x_i (i = 1, 2, \dots, 1000)$  is the  $i$ th sampling point. 10 parameters are chosen as a feature sequence to carry out the matching. Therefore, according to the formula (2), the feature matrix  $A$  of the signal  $y = \{x_1, x_2, \dots, x_{1000}\}$  is constructed with  $991 \times 10$  dimension, i.e.,

$$A = \begin{pmatrix} x_1 & \cdots & x_{10} \\ \vdots & \ddots & \vdots \\ x_{991} & \cdots & x_{1000} \end{pmatrix}$$

Due to the difference between different signals, there are also the differences among features calculated by formula (2). In first hidden layer of network, the different amplitudes are extracted by using the formula (2) so as to obtain the classification result for fault. In second hidden layer of network, the increments of amplitude are extracted by formula (2) so as to obtain detailed features.

### 3. Matching and recognition

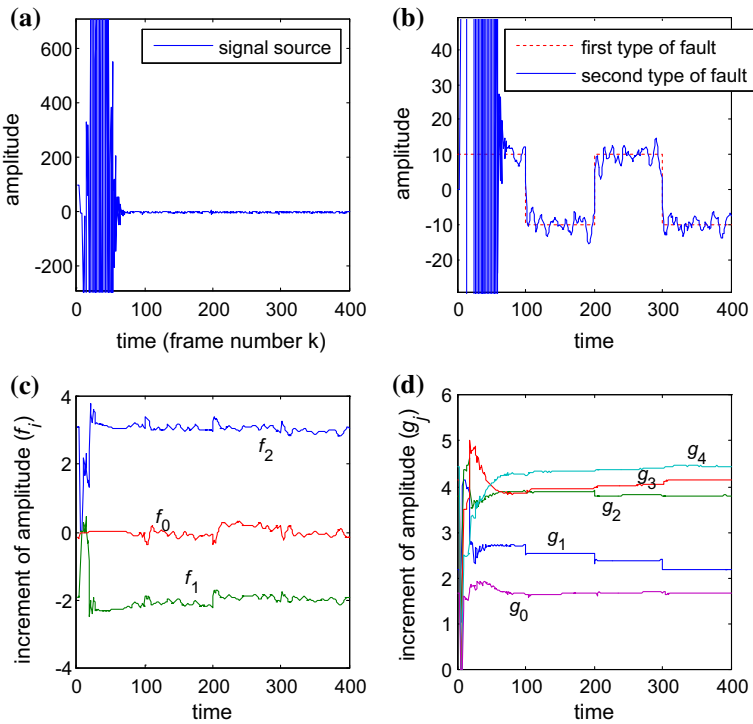
The basic step of the fault matching recognition is as follows:

- P1 The fault in test database are carried out the feature extraction in hidden layer of network  
 P2 The unknown identified characteristic vector with all known characteristic vectors in fault archive is matched in output layer. The random selections and matching recognitions in 100 times are implemented based on the recognition principle. The most similar feature matrix is chosen as the recognition result which has the maximum degree of similarity based on the Sect. 4.2  
 P3 For each fault in the test sample database, after 100 times recognition are done according to the steps P1 ~ P2, the times of the correct recognition and the error recognition are recorded respectively, and the correct recognition rates are obtained by the following calculation formula (11)

Next, the FVD-DLN network is tested by using health and fault data other than the training samples. There are a set of two different types of 100 health and fault samples with 10 kinds of characteristic parameters. Then the FVD-DLN network is tested by using this set of data to detect whether or not the FVD-DLN network can successfully identify different fault parameters.

According to the characteristic parameter type of the gearbox fault, the number of state of the last layer of FVD-DLN network here is set to 10. The 100 health and fault samples that are the signal source are identified by using the FVD-DLN network. The amplitude of various signal source inputted is shown in Fig. 12a. In the first hidden layer of network that is shown Fig. 11, the FVD-DLN performs the classification for the fault signals, and the classification effect is shown in Fig. 12b. In Fig. 12a, b, the negative amplitude indicates the magnitude in the opposite direction. In the second hidden layer of network, further, the FVD-DLN carries out the feature extraction for classified signals, and the extraction effect is shown in Fig. 12c, d, respectively. In the output layer of network, the matching and recognition for various fault signals are performed, and the response of these 10 states is identified as the ten characteristic parameters of the fault. The final recognition results for the gearbox fault are ten types of characteristic parameters. The ten types of fault characteristic parameters are the amplitude of two kinds of faults that are the crack of tooth root and the fracture of gear, as well as the increment of their amplitude at minor, middle and severe fault states, at different rotational speeds, respectively. It can be seen that the FVD-DLN network can better distinguish these seven parameters of the gearbox. This shows that the hidden layer of the designed FVD-DLN network have a good distinguishing ability to feature extraction of signals.

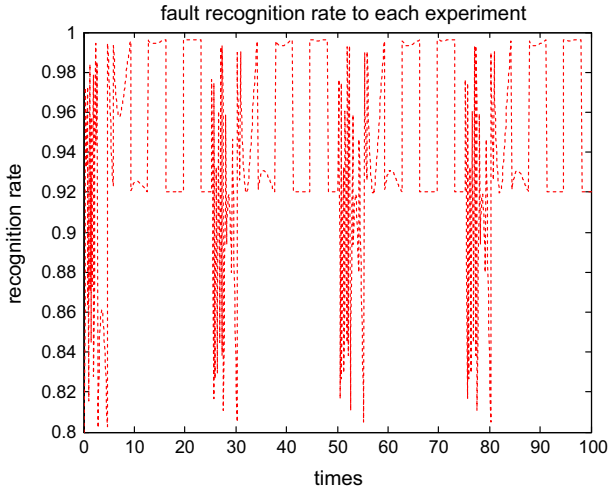
In Fig. 12, the horizontal axis is the time axis that is the frame number  $k$ , and the vertical axis denotes different measures, which it denotes the amplitude of signal in Fig. 12a, b, but it denotes the increment of amplitude of fault signal in Fig. 12c, d.  $f_0$  indicates the increment of amplitude of the crack of tooth root at minor fault state, and its value is small. The same,



**Fig. 12** Recognition effects to fault data of gearbox by the FVD-DLN method

$f_1$  and  $f_2$  indicate the increment of that at middle and severe fault states, respectively, and their values are bigger.  $g_0$ ,  $g_1$ ,  $g_2$ ,  $g_3$  and  $g_4$  indicate the increments of amplitudes of the fracture of gear at minor and middle fault states, as well as those at severe fault state and at rotational speeds of 1705.5 rpm, 6120.09 rpm and 8670 rpm, respectively, and the value  $g_0$  is minimal and the value  $g_4$  is maximum among five parameters.

Figure 12a shows the input signal source that includes health and various fault data, and they will be classified and identified by the FVD-DLN network. Figure 12b shows that the first hidden layer of the FVD-DLN network classifies the 100 data samples according to the nature of the fault parameters, and divides two types of parameters of different natures, i.e., the first type of parameter is the crack of tooth root and the second type of parameter is the fracture of gear. Then, the second hidden layer of the network further subdivides the two types of parameters. That is, the first type of parameter is further divided into three different parameters, which are the increment of amplitude of the crack of tooth root at minor, middle and severe fault states, respectively, as shown in Fig. 12c. The second type of parameter is further divided into five different parameters, which are the increments of amplitudes of the fracture of gear at minor and middle fault states, as well as those at severe fault state and at three different rotational speeds, respectively, as shown in Fig. 12d. At the moment  $k$ , the responses of the 10 states in the final output layer of the FVD-DLN network are the values of ten category of parameters, respectively. There are two states, which are two amplitudes of the crack of tooth root and the fracture of gear, respectively. There are three states, which are three increments of amplitude of the crack of tooth root at minor, middle and severe fault states, respectively. Similarly, other five states are five increments of amplitudes of the



**Fig. 13** Recognition rates of fault at each times in 100 experiments

fracture of gear at minor and middle fault states, as well as those at severe fault state and at three different rotational speeds, respectively.

Assume that the total number of samples per experiment is  $\mathbb{N}$ , and the number of experiments is  $\mathbb{Z}$ . The fault sample is identified by the FVD-DLN method, and the number of correctly identified samples is recorded as  $n_i (i = 1, 2, \dots, \mathbb{Z})$ , then the recognition rate is defined as follows:

$$R_i = \frac{n_i}{\mathbb{N}}, \quad i = 1, 2, \dots, \mathbb{Z} \quad (10)$$

where  $i = 1, 2, \dots, \mathbb{Z}$  denotes the number of experiments.  $R_i$  indicates the correct recognition rate at the  $i$ th times;  $n_i$  denotes the number of samples by recognition correctly at the  $i$ th times;  $\mathbb{N}$  indicates the total number of samples at each times.

In this experiment, the cross validation method for gearbox fault data set is used. The training samples and test samples are randomly selected from the data set at each time. The ratio of the training samples and the test samples selected is 5: 1. The 100 experiments are carried out. The recognition rates of fault at each of the times in the 100 experiments are shown in Fig. 13.

After 100 experiments, the accuracy of recognition by FVD-DLN method to gearbox fault data at each times is different. So, the mean value of accuracy of 100 times is calculated by 96.65% and the variance is obtained by 0.36. It is shown that the FVD-DLN method given in this paper is effective and feasible.

#### 4. Analysis of advantages and disadvantages

From experimental results, the fault diagnosis method based on the FVD-DLN can be concluded the following advantages:

1. The FVD-DLN method not only can deal with the simple fault based on the known reason, but also deal with the uncertain fault caused by uncertainty reason.
2. The FVD-DLN method can continuously reduce the complexity of data set by relying on the features of attribute. Moreover, through calculating the similar degree between unknown fault and known fault by the matching principle, the FVD-DLN method can educe the decision-making results with the higher accuracy by the reduction of attributes.

3. This method based on FVD-DLN can more accurately perform the feature extraction for multi-dimensional parameters by extracting important attributes, and decrease the computing time.
4. The FVD-DLN method does not need to provide any prior information in addition to data processing.

The main disadvantages of FVD-DLN method is:

1. Since the calculation of similar degrees are defined or given by the experts or users, the subjectivity to processing problem is correspondingly stronger.
2. In a longer chain of fault diagnosis, the system needs the more parameters setup, so the discussion to problem and its attributes are on the increase accordingly.

#### 4.4 Comparison of FVD-DLN and existing recognition methods

In order to verify the superiority of FVD-DLN method, this paper also compares it with the classical neural network (CNN) (Wang et al. 2015, 2016), promoted wavelet (PW) (Bruna and Mallat 2013; Jiang and Yuan 2014), support vector machine (SVM) (Huang et al. 2011), logical regression method (LR) (Kankar et al. 2011) to carry out the fault diagnosis.

$\mathbb{Z}$  experiments were performed, and the number of samples  $j$  was used from the minimum number  $S_{\min}$  to the maximum number  $S_{\max}$ , that is, the number  $j$  of samples was  $S_{\min} \leq j \leq S_{\max}$ . According to the above formula (10), there is a correct recognition rate  $R_i$  for each experiment. Then, for each determined number  $j$  of samples, the recognition rate is calculated as follows:

$$C_j = \frac{1}{\mathbb{Z}} \sum_{i=1}^{\mathbb{Z}} R_i, \quad S_{\min} \leq j \leq S_{\max} \quad (11)$$

where  $C_j$  indicates the correct recognition rate for each determined number  $j$  of samples;  $R_i$  can be calculated by the above formula (10).

In the experiment in this section, for each determined sample number  $j$ , using the above various methods, 100 experiments are performed, that is,  $\mathbb{Z} = 100$ , then all the sample numbers  $10 \leq j \leq 1000$  are used for the experiments. The experimental result is the fault recognition rate under the condition of each determined number of samples, as shown in Fig. 14.

Further, the mean of the correct recognition rate can be calculated by averaging the correct recognition rate from the minimum number  $S_{\min}$  of samples to the maximum number  $S_{\max}$  of samples, the calculation is as follows:

$$\bar{C} = \frac{1}{S_{\max} - S_{\min} + 1} \sum_{j=S_{\min}}^{S_{\max}} C_j \quad (12)$$

where  $\bar{C}$  indicates the average correct recognition rate;  $C_j$  can be calculated by the above formula (11).

In this experiment, the number of experimental samples used was from minimum number 10 to maximum number 1000, and the repeating number of experiments is 100. The results are shown in Fig. 14.

In Fig. 14, the horizontal axis is the time axis, and the vertical axis denotes the correct recognition rate. From Fig. 14, the correct recognition rate increases with the increase of the number of samples. When the number of samples reaches 400 or more, the correct recognition

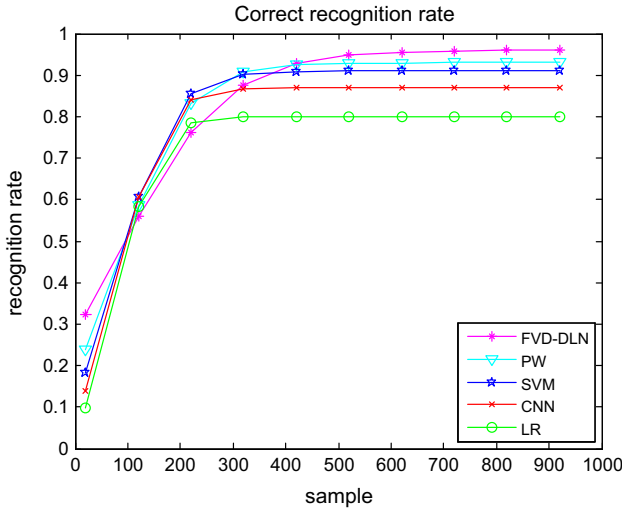


Fig. 14 Comparison of FVD-DLN and four existing methods in gearbox fault identification

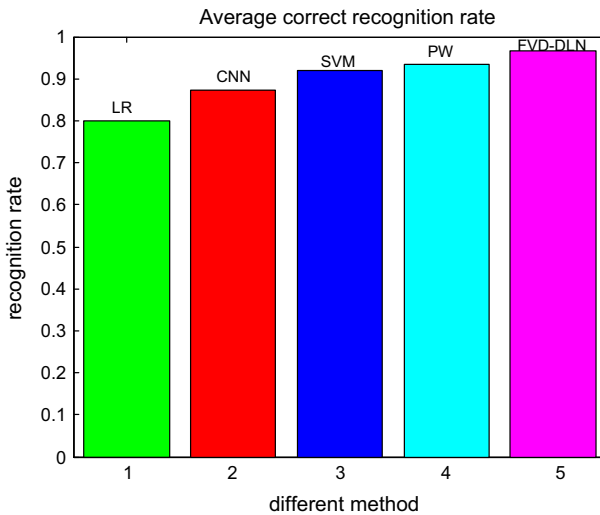


Fig. 15 Average correct recognition rate of several recognition methods

rate of various methods tends to be stable, and the correct recognition rate of the FVD-DLN method proposed in this paper is the highest. However, when the number of samples is less than 400, the advantages of the FVD-DLN method cannot be exhibited, so the FVD-DLN method is suitable for the case where the dimension of samples is relatively large.

In order to visually verify the accuracy of various methods for fault diagnosis, the simulation is further carried out based on the formula (12). The simulation results show that the average accuracy of these methods for fault diagnosis. At the same time, the average correct recognition rate is also shown in Fig. 15.

From Fig. 15, the number of experimental samples was from 10 to 1000, after 100 experiments, the average correct recognition accuracy of the FVD-DLN method to the fault

**Table 3** Comprehensive comparison of FVD-DLN and several existing recognition methods

Recognition methods	Average recognition accuracy (%)	Recognition speed (s)
Logical regression	79.98	0.896
Classical neural network	87.36	0.938
Support vector machine	91.87	1.087
Promoted wavelet	93.52	0.826
FVD-DLN	96.65	0.612

diagnosis of gearbox is 96.65%, and the variance is 0.36. This shows that the FVD-DLN method presented in this paper is effective and feasible.

From Figs. 14 and 15, in a large number of data fault diagnosis, when the number of samples reaches a certain value, the recognition rate of FVD-DLN method for gearbox fault diagnosis is higher than that of the several existing diagnosis methods. It can be seen that the FVD-DLN method has a good diagnostic capability for gearbox fault.

The recognition effect of FVD-DLN method given in this paper is superior to that of several existing methods for the gearbox fault, because the FVD-DLN method not only has higher recognition accuracy, but also has faster recognition speed. The comprehensive comparison results are shown in Table 3.

By comparing the above four methods, it can be seen that the accuracy obtained by FVD-DLN method is the highest and the average accuracy rate is 96.65%, and the average recognition speed is 0.612 s. The average accuracy obtained by existing methods are shown as follows: those of promoted wavelet, support vector machine, classical neural network, logical regression are 93.52%, 91.87%, 87.36% and 79.98%, respectively, and the average recognition speed is 0.826 s, 1.087 s, 0.938 s, and 0.896 s in turn.

The further comparison between FVD-DLN method and other existing methods is summarized as follows:

1. From the point of view to dealing with the uncertainty, the existing methods use the probability or evidence to express the uncertainty, however, the FVD-DLN method not only can use the probability, but also can use the degree of similarity to express the uncertainty.
2. From the complexity of the calculation, the existing methods have the complexity of index information and index time, however, the FVD-DLN method can decrease the computing time, and reduce the storage capacity by the reduction of attributes.
3. The existing methods cannot distinguish correctly the fault reason between uncertainty and ignorance, however, the FVD-DLN method is able to more intuitively to distinguish between the two by using feature extraction algorithm in hidden layer.
4. The existing methods require assuming the priori probability, conditional probability, or independent evidences, however the FVD-DLN method does not need to provide any prior information besides data processing.

In addition to the advantages of the FVD-DLN method mentioned above, its greatest disadvantage is that it requires setting the system parameters, however, the existing methods do not need those.

## 5 Conclusion

This paper proposes a FVD-DLN fault diagnosis method to classify and identify for faults of gearbox. The method firstly performs spatial reconstruction and singular value decomposition for the input signal, extracts features and gives feature extraction methods; Secondly, the singular value as a feature signal is input into the DLN network for learning and training; Thirdly, a FVD-DLN network is constructed based on the feature extraction of the signal and deep learning method. Finally, the trained FVD-DLN network is applied to the fault diagnosis of the automotive transmission gearbox. The application process includes four aspects, i.e., training and test of the constructed network for actual data, matching and recognition, analysis of advantages and disadvantages for FVD-DLN diagnosis method, and the comparison of it and existing recognition methods. To compare with existing referred fault diagnosis methods, the greatest advantage of the FVD-DLN method is that it is not only has faster processing speed, but also applicable to dense signal categories environment, more conversion and transmission error. However, its biggest disadvantage is that the system parameters setup is complicated. For example, some parameters in weights adjustment need great deal simulations to determine, and these parameters are related with the choice of the threshold value. These studies show that the method proposed in this paper has good accuracy, stability and rapidity.

Although the deep learning had achieved great success in practice (Zeng et al. 2016), the deep model obtained by training large-scale data (Ren et al. 2017) is still much work to be done in the future for the theoretical analysis behind it. Moreover, joint deep learning and multi-stage deep learning are two examples, but there is more work in this area in the future.

**Acknowledgements** This work is supported by Center Plain Science and Technology Innovation Talents (194200510016); Science and Technology Innovation Team Project of Henan Province University (19IRT-STHN013); Key Science and Technology Program of Henan Province (172102410063), respectively.

## References

- Bruna J, Mallat S (2013) Invariant scattering convolution networks. *IEEE Trans Pattern Anal Mach Intell* 35(8):1872–1886
- Cho W, Yu NY (2018) Secure communications with asymptotically Gaussian compressed encryption. *IEEE Signal Process Lett* 25(1):80–84
- de Mendivil JRG (2018) Conditions for minimal fuzzy deterministic finite automata via Brzozowski's procedure. *IEEE Trans Fuzzy Syst* 26(4):2409–2420
- Fischer A, Igel C (2013) Training restricted Boltzmann machines: an introduction. *Pattern Recognit* 47(1):25–39
- He Y, Li B (2016) A combined learning strategy for deep learning model. *J Autom* 42(6):953–958
- Hinton G (2012) A practical guide to training restricted Boltzmann machines. In: Montavon G, Orr GB, Müller K-R (eds) *Neural networks: tricks of the trade*. Springer, Berlin, pp 599–619
- Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554
- Huang J, Hu X, Yang F (2011) Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker. *Measurement* 44(6):1018–1027
- Jiang T, Yuan S (2014) Diagnosis of rolling bearing based on improved wavelet neural network. *J Huazhong Agric Univ* 33(1):131–136
- Jiang H, Zhang H (2018) Iterative ADP learning algorithms for discrete-time multi-player games. *Artif Intell Rev* 50(1):75–91
- Kankar PK, Sharma SC, Harsha SP (2011) Rolling element bearing fault diagnosis using wavelet transform. *Neurocomputing* 74(10):1638–1645



- Li Y, Wu QE, Peng L (2018) Simultaneous event-triggered fault detection and estimation for stochastic systems subject to deception attacks. *Sensors* 18(2):321–345
- Lu H, Zhang Q (2016) Applications of deep convolutional neural network in computer vision. *J Data Acquis Process* 31(1):1–17
- Ma S, Shen T, Wang R, Lai H, Yu Z (2015) Identification of terahertz spectroscopy based on deep belief networks. *Spectrosc Spectr Anal* 35(12):3325–3325
- Ren J, Qi Y, Dai Y, Xuan Y, Shi Y (2017) nOSV: a lightweight nested-virtualization VMM for hosting high performance computing on cloud. *J Syst Softw* 124(2):137–152
- Roux NL, Bengio Y (2008) Representational power of restricted Boltzmann machines and deep belief networks. *Neural Comput* 20(6):1631–1649
- Su H, Zhang W (2017) A combined backstepping and dynamic surface control to adaptive fuzzy state-feedback control. *Int J Adapt Control Signal Process* 31(11):1666–1685
- Su H, Zhang T, Zhang W (2017) Fuzzy adaptive control for SISO nonlinear uncertain systems based on backstepping and small-gain approach. *Neurocomputing* 238(17):212–226
- Tuerxun T, Dai L (2015) Deep neural network based uyghur large vocabulary continuous speech recognition. *J Data Acquis Process* 30(2):365–371
- Wang X, Wang L, Li X (2015) Fault diagnosis of crack of motor rotor based on LM-BP Neural network. *Micro-motor* 43(4):18–20
- Wang W, Cao T, Zeng Y, Li F (2016) Deep learning in saliency detection. *J Mil Commun Technol* 37(2):92–97
- Wu QE, Yang W, Chen Z, Zhang P (2015) Research of semantic understanding on target region of interest for fuzzy image. *Eng Appl Artif Intell* 37(1):135–144
- Wu QE, Wang J, Yang C, Cui G, Yang W (2016) Target recognition by texture segmentation algorithm. *Expert Syst Appl* 46(1):394–404
- Yang L, Qi Y, Han J, Wang C, Liu Y (2015) Shelving interference and joint identification in large-scale RFID systems. *IEEE Trans Parallel Distrib Syst* 26(11):3149–3159
- Zeng X, Shu L, Huang G (2016) Fluctuating interval number series forecasting based on gm (1,1) and svm. *J Grey Syst* 28:1–14
- Zhang C, Ji N, Wang G (2015) Restricted Boltzmann machine. *J Eng Math* 32(2):1005–3085
- Zhang X, Liu X, Li Y (2017) Adaptive fuzzy tracking control for nonlinear strict-feedback systems with unmodeled dynamics via backstepping technique. *Neurocomputing* 235(26):182–191

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.