



Intelligent Fault Diagnosis of Gears Based on Deep Learning Feature Extraction and Particle Swarm Support Vector Machine State Recognition

Ahmed N. Al-Masri¹, Hamam Mokayed²

¹ American University in the Emirates, Dubai, UAE

² LTU University of Technology, Sweden

Emails: ahmed.almasri@aue.ac ; Hamam.mokayed@ltu.se,

Abstract

Gear faults have always been a problem encountered in mechanical processing. For gear fault diagnosis, using mathematical statistical feature extraction methods, deep learning neural networks (DLNN), particle swarm algorithm (PSA), and support vector machines (SVM), etc. According to the feature extraction of deep learning and particle swarm SVM state recognition, the intelligent diagnosis model is established, and the reliability of the model is verified by experiments. The model uses the combination of spectral features extracted by deep learning adaptively and the time domain features extracted by mathematical statistics methods to form a joint feature vector, and then uses particle swarm SVM to diagnose the joint feature vector. After research, this paper draws a classification fitness curve combining the fault spectrum features extracted by DLNN and traditional time-domain statistical features. The classification result obtained by using this method is 95.3%. The reliability of the model is verified, and satisfactory diagnosis results are obtained. In addition, the application results also verify the effectiveness of adaptively extracting spectral features based on deep learning.

Key words: DLFE, SVM, Particle Swarm, Fault Diagnosis

1. Introduction

At present, in mechanical fault diagnosis, commonly used signal processing techniques include time domain statistical analysis, frequency domain statistical analysis, Fourier transform analysis, power spectrum analysis, cepstrum analysis, and envelope spectrum analysis. For non-linear and non-stationary signals, commonly used signal processing techniques include short-time Fourier transform, wavelet transform, wavelet packet transform, and empirical mode decomposition [1-2]. Traditional intelligent diagnosis methods are generally based on the "feature extraction + classifier" model, the core of which is the extraction of feature values and the design of the classifier. For different diagnosis objects, it is usually necessary to extract different fault feature values according to prior knowledge, which will definitely bring diagnostic errors to the final diagnosis result; At the same time, traditional classifiers generally use shallow models, which makes it difficult to describe the complex mapping relationship between signals and

equipment operating conditions. Based on this, many scholars have proposed the theory of deep learning, which has opened the wave of deep learning in academia and industry [3-4]. Deep learning can combine low-level features to form more abstract high-level feature representations, thereby discovering distributed feature representations of data. Deep learning has strong nonlinear expression ability and good discrimination ability, and has made breakthrough progress in the areas of speech recognition and image recognition. At present, there are also some applications of DNN in the fields of process monitoring / mechanical equipment fault diagnosis [5-6]. As the core equipment in modern industrial production, gear machinery is widely used in major equipment such as nuclear power plant units, large wind power equipment, aero engines, etc. [7-8]. Due to the actual needs of the project, gear mechanical equipment is generally operated under complex conditions such as high speed, high load, and large temperature difference. Under such conditions, the frequency of failures is much higher than other equipment. Therefore, as a key component of most equipment, the health of gear machinery is related to the safety of production. Once a fault occurs, it will cause production downtime, and threaten the life and property of people. Therefore, it is a strong actual demand for engineering to monitor its operating health [9-10]. In recent years, with the rapid development of fault monitoring and diagnosis technology, more and more comprehensive intelligent diagnostic technology can detect potential dangers before catastrophic equipment failures, improve equipment reliability and safety, at the same time, it can reduce the operation and maintenance cost of complex engineering system. Therefore, the health monitoring and fault diagnosis of gear machinery is of great significance [11-12].

Intelligent diagnostic methods have always been a research hotspot in the field of fault prediction and diagnosis of rotating machinery. Pei Cao proposed a DNN method for gear fault diagnosis based on stacked self-encoding (SAE) and softmax regression. Pei Cao first used SAE to extract features from the frequency spectrum of the vibration signal. Then use the learned features to train a softmax regression classifier to identify gear failures. The diagnosis results verify the feasibility of the method and can obtain better classification performance. As the number of DNN layers increases, the learning characteristics of each hidden layer become more robust to classification [13-14]. In the gearbox-based electromechanical system, some electrical signals, such as electromagnetic torque and motor current, can track the pulsation of the load torque. This allows the motor to be regarded as a non-destructive sensor for diagnosing gear failures. Jin Shoufeng first established the mechatronics model of the motor and gear transmission system. Then, Jin Shoufeng analyzed and simulated the theory of electromagnetic torque characteristic analysis (ETSA) and motor current characteristic analysis (MCSA) for fault diagnosis. It can be seen that the electromagnetic torque not affected by the main frequency can more directly reflect the fault information. The fault characteristics in the frequency domain are more obvious than the fault characteristics in the time domain. Finally, Jin Shoufeng experimentally compared the performance of the two methods under different speed and load torque conditions. Jin Shoufeng's research results show that both methods are affected by speed and load torque, and fault diagnosis is more effective at low speed and heavy load [15-16]. Vikas Sharma proposed a gear bearing fault identification method based on least squares SVM. The two energy selection criteria of maximum energy shadow speed and maximum relative energy are selected, and the appropriate microblog extraction is selected. The method of fault diagnosis consists of three stages. Please consider six basic microblogs first. According to the criteria of ripple selection, statistical features are extracted from the ripple coefficients of the original vibration signals. Shannon entropy benchmark Weibo considering the largest energy and the largest energy. Finally, Vikas Sharma uses these statistical features as input to LSSVM technology to classify gearbox failures. Based on the Shannon entropy ratio criterion for maximum energy, the optimal decomposition level of the wavelet is selected. In addition, Vikas Sharma takes the

energy of wavelet coefficients and Shannon entropy as two new features of the classifier, and uses other statistical parameters as inputs to the classifier. Vikas Sharma takes kernel functions and multi-kernel functions as a new method, and combines three strategies to multi-classify gearboxes. The fault classification results show that using multi-core and OAOT strategies, LSSVM can more accurately identify the fault category of the gearbox [17-18]. Oil gear operation plays an important role in the helicopter, which has a great impact on the flight safety of the helicopter. Zeng Ming has a dynamic impact on the active gear bearing system. Zeng Ming established the torsional vibration mode of a cracked gear transmission system, which considered the same nonlinear factors as the time-varying information. It is verified by Qiangsheng, bailasi and Ming that the crack of the gear frame affects the angular deformation of the car column. The deformation of the soaking bearing plate is studied under the static adaptability. When the existing dynamic model is used to predict the dynamic characteristics of the gears, the accuracy is very good. The defect features extracted from the model prediction show the consistency between the vibration characteristics of the meteor gearbox and the Kr horn conditions (length and position) [19-20].

In this paper, a diagnosis method based on deep learning of fault frequency domain features and time domain statistical features combined by PSO support vector machine for state recognition is proposed. Through the analysis and comparison of the test bench data, the superiority of this method is demonstrated: First, a DLNN is established by cascading a noise reduction autoencoder, and the fault features are adaptively extracted directly from the frequency domain signals, Based on the signal processing technology of diagnosis experience, the complex process of manual fault feature extraction is adopted. Secondly, deep learning is used to extract fault frequency domain features and manual methods are used to extract vibration time domain statistical features. In order to improve the accuracy and reliability of the fault diagnosis, we use the deep learning to extract the frequency-domain features of obstacles, and use the manual method to extract the time-domain statistical features of vibration, which are combined with the frequency-domain features.

2. Proposed Method

2.1 SVM Theory and Classification Strategy

(1) Machine learning theory

The purpose of machine learning is to find the relationship between the input and output of the system, so that the unknown output can be predicted most correctly [21]. The problem of machine learning can be expressed as: there is a certain unknown dependency relationship between the known variable y and the input x , that is, there is an unknown joint probability $F(x, y)$, which mainly focuses on finding rules from data samples, and using these obtained rules to detect data for effective prediction. The traditional statistical research institute satisfies the asymptotic theory when the number of samples approaches infinity, but in actual engineering, the number of samples we obtain is usually very small, which makes many learning methods in practical application better than those in theory [22-23].

(2) Basic ideas of statistical learning theory

For a certain type of machine failure, its feature vector is selected as training data. In this way, several sets of training data constitute a set of regions in n-dimensional space. Different fault types correspond to different areas. The so-called fault diagnosis becomes the interface to find these areas in the dimensional

space. The determination and expression of the interface must be able to be completed by training on the training data. The accuracy of fault diagnosis essentially becomes the accuracy of regional demarcation. To this end, the following three factors determine the accuracy of the classification:

1) Selection of feature vectors

The selected feature vector should characterize the corresponding fault most prominently, and contain the information with the most obvious difference from other fault features. Otherwise, the fault areas cross or overlap in the feature vector space. In this case, no matter what classification method is adopted, it is difficult to accurately determine the delimitation hyperplane, or even find the delimitation hyperplane. Therefore, the accuracy of the diagnosis will not be too high. This paper believes that the selection of feature vectors is as important as the selection of classification methods.

2) Choice of classification method

To be able to find the interface accurately, you need to choose a suitable classification algorithm. The classification algorithm can determine the parameters and expressions of the interface through the training of samples, and the training process has the characteristics of monotonic approximation. However, the existing classification algorithms have shortcomings. For example, for the fault diagnosis of rotating machinery, the optimal structure and structural parameters of the neural-network cannot be determined through sample training and depend entirely on personal experience.

3) Noise immunity training

The data contains noise interference and randomness. The interface determined after training is actually a boundary hyperplane with certain noise immunity (including randomness). During fault diagnosis, the input data is also noisy, and the influence of noise may make classification errors. Therefore, anti-noise ability is an issue that every classification algorithm must consider. In many cases, the fault feature space is not linearly separable, and sometimes even completely inseparable. SVMs use a simple method to solve this problem-upscaling. Through "dimensional improvement", the information contained in the data can be mined more deeply, which makes the low-dimensional linear inseparable problem rise to the high-dimensional and become linearly separable. The idea of upgrading dimension is often used in fault diagnosis, which can show the information hidden in the data, so as to better diagnose the fault.

(3) Support vector classification machine

1) Linear SVM for two types of linearly separable problems, the maximum interval can be transformed into a problem of optimizing the variables w and b finding the optimal value:

$$\min \frac{1}{2} \|w\|^2 \quad (1)$$

$$y_i \{(w \bullet x_i) + b\} \geq 1, \dots, n \quad (2)$$

For the two types of linear inseparable problems, we can solve it by introducing the relaxation variable ξ_i ($\xi_i \geq 0$), which weakens the constraints. The slack variable describes the degree of misalignment in the training set. Although we encounter linear inseparability problems, we always want to maximize the

hyperplane interval and the degree of misalignment reaches the minimum value. This introduces the penalty parameter C , which is used to weigh the classification intervals and mismatched samples. Thus the objective function becomes:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (3)$$

$$y_i \{(w \bullet x_i) + b\} \geq 1 - \xi_i, i = 1, \dots, n, \xi_i \geq 0 \quad (4)$$

2) Non-linear SVM

For non-linear classification, first use a non-linear mapping to map the data samples from the original space R to a high-dimensional feature space, and then find the optimal classification surface in the high-dimensional feature space. SVM is to first transform the input space to a high-dimensional space through a non-linear transformation defined by a kernel function, and then find the optimal classification surface in this space [24-25]. The form of SVM classification function is similar to that of neural-network, and its output is a linear combination of intermediate nodes. Each intermediate node, also known as a support vector network, corresponds to the inner product of the input sample and the support vector, as shown in Figure 1.

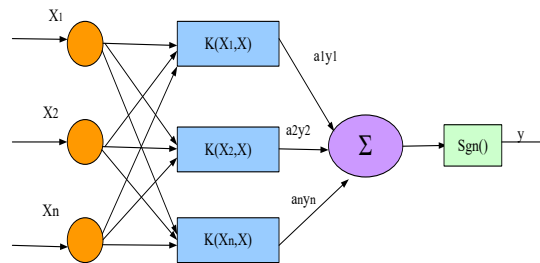


Figure 1. SVM

The schematic diagram of the SVM can be seen from FIG. 1. The high-dimensional feature space has a very large number of dimensions, and a dot product operation is required between the vectors. The huge computational workload will cause a "dimensional disaster" in the operation. According to the relevant theorem of the function, as long as the function $K(x_i, y_i)$ meets the Mercer condition, it corresponds to the dot product of a certain transformation space, so that a nonlinear classification problem can be converted into a quadratic programming problem to solve:

$$\min_{\alpha} \sum_{j=1}^l \alpha_j - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j k(x_i, x_j) \quad (5)$$

$$s.t. \sum_{i=1}^l y_i \alpha_i = 0, (0 \leq \alpha_i \leq C, i = 1, \dots, l) \quad (6)$$

The corresponding decision function is

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^I \alpha_i \cdot y_i (x_i \cdot x) + b \right\} \quad (7)$$

When constructing a classification function, the SVM first calculates the dot product in the input space, and then performs a non-linear transformation. Such a large amount of work is done in the input space, rather than in a high-dimensional space.

3) SVM classification strategy

Compared with the original two-class classification model of the SVM, it cannot meet the multi-class faults encountered in practice. The current multi-class processing methods mainly include: classic one-to-one classification (OVO), one-to-many classification (OVR) directed acyclic graph SVM (DAG -SVM), decision tree SVM (DT -SVM).

OVR method and OVO construction method adopt voting strategy, the algorithm is relatively simple, but there are problems of indivisible regions, which affects the effect of classification. Both DT and DAG are based on the decision tree construction strategy. Among them, DT algorithm greatly reduces the training SVM by establishing an effective decision tree, so it provides a guarantee for the fast and effective classification of mechanical faults. In the following, the classification effect and practicability of the decision tree SVM are further analyzed through two aspects of classification complexity and classification accuracy: First, the classification complexity comparison analysis, for fault data, assuming the fault category is K, the traditional SVM training "one-to-one" method with DAG requires the most SVMs to be trained, and it increases exponentially with the increase of the number of categories K; OVR uses the "one-to-one" method, which greatly reduces the complexity of training; (Decision tree SVM) Because of the tree structure, the complexity of training is minimal. This can effectively reduce the complexity of training and improve the efficiency of classification. Finally, the classification effect is compared and analyzed. Compared with the traditional SVM of OVO, OVR, DAG and DT (decision tree) classification strategies, the classification results are shown in Table 1:

Table 1. Comparison of classification accuracy of several SVMs

Classification strategy	OVR	OVO	DAG	DT
Accuracy	87%	85%	90%	90%

It can be seen from Table 1 that although the OVO method and OVR are relatively simple, the recognition efficiency is not high. The main reason is that the voting strategy uses the method of probability statistics. There are indivisible regions. When there are many classifications and the amount of classification data is relatively large, this problem will become more prominent. And the classification method of DT enables the SVM to actively establish the corresponding decision tree according to the actual fault, which effectively diagnoses the fault. However, the traditional DT algorithm generally adopts a fixed tree structure when constructing decision trees. The choice of decision nodes is arbitrary and easy to generate cumulative errors. The decision tree constructed in this way cannot adapt to the complex and diverse characteristics of equipment faults, so how to effectively construct a decision tree is a problem that needs to be studied. However, consulting related literatures, SVMs are mainly based on applications in the process of equipment fault diagnosis, and there are few related studies on the optimization and selection of decision-making.

2.2 DLFE

(1) DLFE

The purpose of deep plan is to simulate the learning process of brain, build deep model, combine a lot of training data and learn the hidden characteristics.

(2) Laminated noise reduction automatic encoder

Most of the noise eliminators overlap the automatic controllers to form a neural-network of hunger strike noise eliminators. The autoencoder is divided into three layers of new network without monitoring to encode the network and decode the network. The structure of noise to reduce automatic coder is Figure 2.

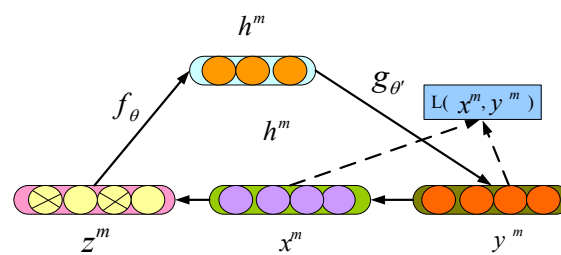


Figure 2. Schematic diagram of noise reduction autoencoder

From Figure 2, you can clearly see the circuit diagram of the noyz control controller. The coding network adds noyz with specific statistical characteristics to the sample data and encodes the sample. The coding network infers the original of the unimpeded sample based on the noise interference data.

(3) Pre-training and fine-tuning of DNN

DNN can effectively extract the field function of input signal by pre educating all levels without supervision. In order to optimize the basic performance of DNN, DNN is gradually adjusted in the map learning mode. The structure of automatic coder is Figure 3.

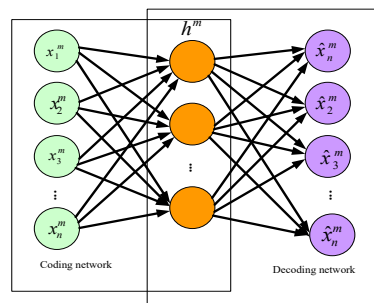


Figure 3. Structure of the automatic encoder

Figure 3 shows the structure of the autoencoder. Given an unlabeled training sample set $\{x^m | m = 1, 2, \dots, M\}$, m is the serial number of the input training samples and M is the total number of input training samples. The encoding network transforms each training sample x^m into an encoding vector h^m through the encoding function f_θ :

$$h^m = f_{\theta}(x^m) = s_f(\omega x^m + b) \quad (8)$$

In the formula, s_f is the activation function of the coding network; θ is the parameter set of the coding network, and $\theta = \{\omega, b\}$; ω and b are the connection weight and bias parameters of the coding network, respectively.

The decoding network uses the decoding function g'_{θ} to inversely transform the encoded vector h^m into a reconstructed representation \hat{x}^m of x^m :

$$\hat{x}^m = g'_{\theta}(h_m) = s_g(\omega' h_m + d) \quad (9)$$

DAE completes the training of the entire network by minimizing the reconstruction errors $L(x^m, \hat{x}^m)$ of x^m and \hat{x}^m , which is

$$L(x^m, \hat{x}^m) = \frac{1}{M} \sum_{m=1}^M \|x^m - \hat{x}^m\|^2 \quad (10)$$

Through labeled samples for training, the errors are transmitted from-top-to-bottom, and the deep learning network is fine-tuned (that is, top-down supervised learning). This process is a supervised training process.

Set the x^m health status type to d^m , and DNN to fine-tune by minimizing $\varphi_{DNN}(\Theta)$, which is

$$\varphi_{DNN}(\Theta) = \frac{1}{M} \sum_m L(y^m, d^m) \quad (11)$$

Where Θ is the parameter set of DNN and $\Theta = \{\theta_1, \theta_2, \dots, \theta_{N-1}\}$.

2.3 Intelligent Fault Diagnosis Model Based on DLFE and SVM State Recognition

The DNN-SVM model will receive the input signal of the gear vibration signal, and repeat the multi sparse noise to reduce the automatic encoder. The output of the first stage automatic recording output is the second level automatic compilation. The input of the second stage auto encoder will be used as the input of the 3 stage auto encoder. The output of the last-stage autoencoder is the adaptively extracted features of deep learning, which are combined with the artificially extracted time domain feature parameters to classify as the input of a particle swarm SVM, thereby completing fault diagnosis.

Deep learning-based gear fault diagnosis method includes the following steps:

(1) This method takes the gear's original vibration data X as the input sample, and performs a fast Fourier transform on it to obtain a new input sample spectrum signal X_1 .

(2) Normalize the vibration spectrum signal X_1 of the gear by linear normalization method to obtain the vibration spectrum signal X_2 . Assuming the X_1 data length of the gear vibration spectrum signal is n , then

$$\hat{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (12)$$

In the formula, \hat{x}_i is the i data point of the vibration spectrum signal $X_2, i = 1, 2, \dots, n$; x_i is the i data point of the vibration spectrum signal X_1 ; x_{\min} is the minimum value of the vibration spectrum signal X_1 ; x_{\max} is the maximum value of vibration spectrum signal X_1 .

(3) The vibration spectrum signal X_2 is input into a DLNN to perform deep learning on the gear spectrum characteristics.

(4) Combining the features automatically extracted by deep learning with artificially extracted time-domain statistical features, input SVMs for training, use PSA to optimize the parameters of SVMs, test the test samples, and complete the gear troubleshooting.

2.4 Particle Swarm

(1) Application of PSA

As an emerging swarm intelligence algorithm, particle swarm optimization is widely used in engineering fields such as engineering design and optimization, robot control, traffic dispatching, communication engineering, industrial production line optimization, and computers. Engineering design and optimization include neural-network optimization, fuzzy-neural-network rule extraction, circuit design, digital filter design, semiconductor device synthesis, layout optimization, control parameter, system identification and state optimization. In the field of power system, Party movement optimization is used to achieve power. Optimization, voltage control, power station reliability and most applicable composition. In robot control, particle swarm optimization is used in robot vibration suppression trajectory planning and mobile robot path planning. In the transportation field, particle swarm optimization is applied to the dynamic programming problem in the field of traffic grooming and path planning. In computer field, particle swarm optimization is applied to tasks such as task assignment, pattern recognition, image processing, and data mining. In the field of industrial production, particle swarm optimization is used to optimize raw material mixing and optimize computer control.

(2) Basic theory of PSA

Some scholars have proposed particle optimization algorithm based on the principle of swarm intelligence. This algorithm gets inspiration in the tide activities, shares information among individuals, and makes the group movement develop into evolutionism in the disordered process. Get the best solution.

If there is a particle swarm in D-dimensional space, and the particle swarm consists of m particles. $v_i(v_{i1}, v_{i2}, \dots, v_{iD})$ is the speed of the i particle, $x_i(x_{i1}, x_{i2}, \dots, x_{iD})$ is the position of the i particle,

$P_i = (P_{i1}, P_{i2}, \dots, P_{iD})$ is the optimal position currently found by the i particle, and $p_g = (p_{g1}, p_{g2}, \dots, p_{gD})$ is the optimal position found by the entire population. The update formula is as follows:

$$v_{id}(t+1) = \omega v_{id}(t) + c_1 r_1 (p_{id} - x_{id}(t)) + c_2 r_2 (p_{dg} - x_{id}(t)) \quad (13)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (14)$$

When $v_{id} > v_{\max}$, take $v_{id} = v_{\max}$;

When $v_{id} < -v_{\max}$, take $v_{id} = -v_{\max}$.

In the formula: $i = 1, 2, \dots, m$; $d = 1, 2, \dots, D$; r_1, r_2 is a random number between $[0, 1]$; t is the current number of iterations; ω is the inertia weight; c_1, c_2 is the acceleration constant.

3. Experiments

3.1 Experimental Environment

A fault simulation test was performed in a test bench of a multi-stage gear shifting system to verify the extraction effect of gear intelligent fault diagnosis method based on deep learning function extraction and particle group SVM state recognition. The test bench that you selected with this white paper can simulate a variety-of-gear-boxes such as gear-wear, tooth breakage, cutting, root cracking, and gear eccentricity.

3.2 Experimental Parameters

In this paper, the original vibration signal, the original vibration time-domain signal and the frequency-domain signal are measured under the six states of broken gear, gear cutting, tooth-root-crack, and eccentric gear. In each state, 100 samples were collected, of which 50 were used as training samples and the other 50 were used as test samples. The experimental parameters are shown in Table 2.

Table 2. Experimental parameters

Gear model	ER-16K
Number of teeth of large gear at medium speed end	100
Medium speed end large gear speed(r/min)	783
frequency of sampling(kHz)	12
sample length	700

3.3 Experimental Implementation

The DNN structure is set to 350-250-150-60-6 in this document. The new enabling function is the sigma function, the input sample is determined by the input sample, the output is in the pre training and fine

adjustment stage, and the number of model repetitions is 100. In addition, in order to improve the robustness of defect diagnosis, the coding network must add noise with specific statistical characteristics to the sample data. The adaptive method is suitable for DNN. Using artificial method, statistical features are extracted from frequency domain to detect whether DLFE and particle focus support vector modes are adaptive. The spectrum Frisbee using spectrum extraction is input to the classified warm support vector system.

4. Discussion

4.1 Feature Extraction and Analysis

The acceleration sensor is used to measure the original vibration signal, Original vibration time domain signal and frequency domain signal. The number of samples collected in each state is 100, of which 50 are training samples and the other 50 are test samples. See Table 3 for 6 operation states of medium speed shaft gear. Principal component analysis is carried out on the frequency domain statistical features extracted manually and the fault spectrum features of gear intelligent fault diagnosis based on DLFE and particle swarm SVM state recognition. The results are shown in Figure 4 and Figure 5.

Table 3. Six operation states of medium speed shaft big gear

Number of training samples n	Number of test samples n2	State type	Status mark
50	50	Status mark	1
50	50	gear tooth	2
50	50	Gear cutting tooth	3
50	50	Gear wear	4
50	50	Gear eccentricity	5
50	50	Normal gear	6

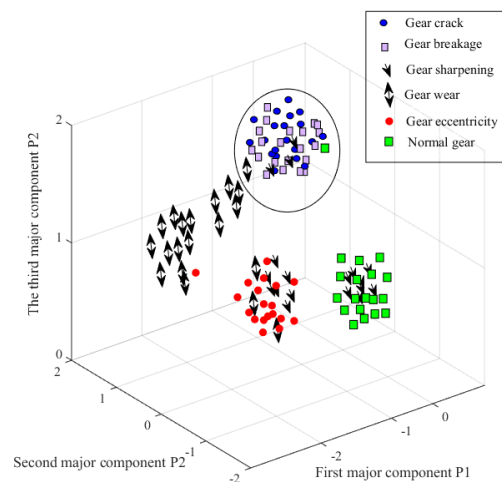


Figure 4. Main scatter charts of statistical characteristics of frequency domain extracted by artificial method

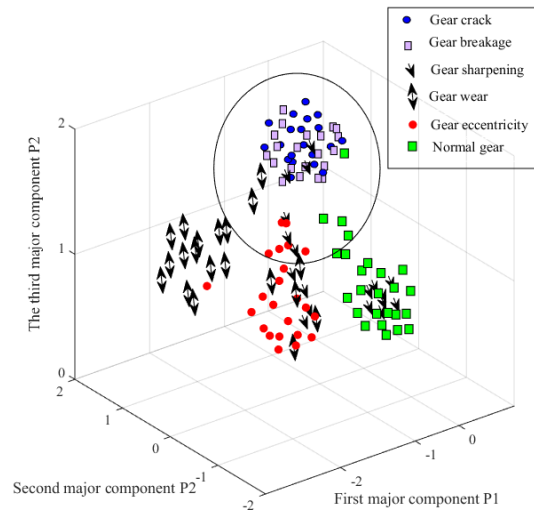


Figure 5. Principal component and scatter diagram of fault spectrum features extracted adaptively by DLNN

It can be known from Figure 4 and Figure 5 that compared with the fault features extracted by the artificial method, the fault features extracted by deep learning have fewer overlapping parts and have better separability. The results show that based on DLFE and particle swarm SVM state recognition, effective fault features in the frequency spectrum can be adaptively extracted, which avoids the complexity brought by the frequency domain feature extraction process by manual methods and saves a lot of time. Enhance the intelligence of the recognition process, and increase the accuracy of feature classification to a certain extent.

4.2 Frequency Domain Feature Classification Analysis

(1) Frequency domain feature classification of artificial method

According to the statistical feature of frequency domain extracted by artificial method and the defect of dlmm adaptive extraction, particle swarm SVM is input for classification. The classification results are shown in Figure 6 and Figure 7.

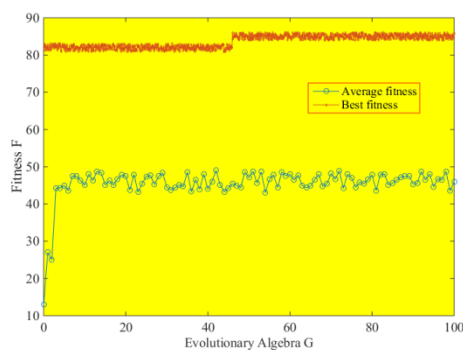


Figure 6. Fitness curve of particle swarm SVM for statistical feature classification in the frequency domain extracted by artificial methods

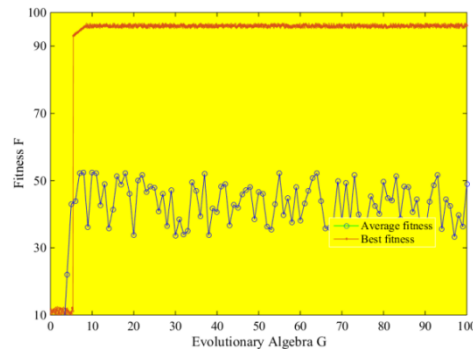


Figure 7. Fitness curve of particle swarm SVM for fault spectrum feature classification adaptive extraction of DLNN

Figure 6 shows an adaptive curve of a particle swarm support vector machine for statistical feature classification in the frequency domain extracted by an artificial method. Figure 7 is a fit curve of the fault spectrum feature classification of particle spectrum support vector machines. 6 and 7, the frequency domain statistical feature classification extracted by the artificial method was 84.6%, and the frequency domain statistical feature classification result extracted by the artificial method was adaptively extracted using the DLNN, and 84.6% of the feature of the obstacle spectrum occupied Yes.

(2) Feature classification in frequency domain

In order to verify the effectiveness of increasing the fault time domain features to improve the classification accuracy of the classifier, the fault spectrum features extracted by the DLNN are combined with the traditional time domain statistical features, and then input to the particle swarm SVM for classification. The results are shown in Figure 8.

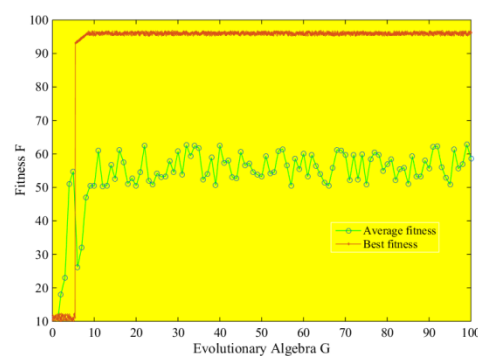


Figure 8. Classification fitness curve combining fault spectrum features extracted by DLNN and traditional time domain statistical features

It can be seen from Figure 8 that the classification fitness curve of the combination of the fault spectrum features extracted by the DLNN and the traditional time domain statistical features, the classification result is 95.3%. Comparing Figure 5 with Figure 6, it is found that adding traditional time-domain statistical characteristic parameters can improve the accuracy of classifiers in classifying faults. The combination of

deep learning to extract fault frequency domain features and manual method to extract vibration time domain statistical features, and the combination of instant domain features and frequency domain features improve the accuracy and reliability of fault diagnosis.

5. Conclusions

Gear faults have always been the top priority in mechanical processing problems. This paper proposes a method based on the use of mathematical statistical feature extraction methods, DLNN, PSA and SVM for gear fault diagnosis. DLNN adaptive extraction of fault spectrum diagnosis method.

In this paper, an intelligent diagnostic model is established based on the combination of DLFE and particle swarm SVM state recognition, and the reliability of the model is verified through experiments. The model uses the deep learning adaptive extraction of spectral features and mathematical statistics to extract the time-domain features are combined to form a joint feature vector, and then the particle swarm SVM is used to diagnose the joint feature vector. This model realized the reliable identification of different fault types of the large gears of the medium-speed shaft in the fault diagnosis of the multi-stage gear transmission system test bench, and obtained satisfactory diagnosis results. The application results also verify the effectiveness of adaptively extracting spectral features based on deep learning.

Compared with the traditional diagnostic method of statistical features in the frequency domain, the method in this paper gets rid of the reliance on a large amount of signal processing knowledge and diagnostic engineering experience, saves a lot of time, and achieves higher monitoring and diagnostic accuracy. In addition, the time-domain statistical feature parameters and the fault spectrum features extracted by the DLNN are fault features extracted from different angles. The combination of the two can effectively improve the classification accuracy of the classifier.

References

- [1] Gui Y, Han Q K, Chu F L. A vibration model for fault diagnosis of planetary gearboxes with localized planet bearing defects[J]. *Journal of Mechanical Science & Technology*, 2016, 30(9):4109-4119.
- [2] Qingyang Xu. Minimal Structural ART Neural Network and Fault Diagnosis Application of Gas Turbine[J]. *Open Mechanical Engineering Journal*, 2016, 10(1):13-22.
- [3] Kuniaki Noda, Yuki Yamaguchi, Kazuhiro Nakadai. Audio-visual speech recognition using deep learning[J]. *Applied Intelligence*, 2015, 42(4):722-737.
- [4] Zhenlong Yuan, Yongqiang Lu, Yibo Xue. Droiddetector: Android malware characterization and detection using deep learning[J]. *Tsinghua Science and Technology*, 2016, 21(1):114-123.
- [5] Duyu Tang, Bing Qin, Ting Liu. Deep learning for sentiment analysis: Successful approaches and future challenges[J]. *Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery*, 2015, 5(6):292-303.
- [6] Xin Lu, Zhe Lin, Hailin Jin. Rating Pictorial Aesthetics Using Deep Learning[J]. *IEEE Transactions on Multimedia*, 2015, 17(11):1-1.
- [7] Nantian Huang, Huaijin Chen, Guowei Cai. Mechanical Fault Diagnosis of High Voltage Circuit Breakers Based on Variational Mode Decomposition and Multi-Layer Classifier[J]. *Sensors*, 2016, 16(11):1887.
- [8] Shu-fa Yan, Biao Ma, Chang-song Zheng. Weighted Evidential Fusion Method for Fault Diagnosis of Mechanical Transmission Based on Oil Analysis Data[J]. *International Journal of Automotive Technology*, 2019, 20(5):989-996.
- [9] Guo-Qian Jiang, Ping Xie, Xiao Wang. Intelligent Fault Diagnosis of Rotary Machinery Based on Unsupervised Multiscale Representation Learning[J]. *Chinese Journal of Mechanical Engineering*, 2017,

30(6):1314-1324.

- [10] Nan Pan, Xing Wu, Yu Guo. Bearing compound fault acoustic diagnosis based on improved blind deconvolution algorithm[J]. Transactions of the Canadian Society for Mechanical Engineering, 2015, 39(3):657-667.
- [11] Zhou Q, Yan P, Liu H, et al. A hybrid fault diagnosis method for mechanical components based on ontology and signal analysis[J]. Journal of Intelligent Manufacturing, 2017(6):1-23.
- [12] Chen Li, Hu Niansu. Fault Diagnosis for Steam-Flow Exciting Vibration of Ultra Supercritical 1000MW Steam Turbine[J]. Open Mechanical Engineering Journal, 2015, 9(1):1067-1075.
- [13] Pei Cao, Shengli Zhang, Jiong Tang. Pre-Processing-Free Gear Fault Diagnosis Using Small Datasets with Deep Convolutional Neural Network-Based Transfer Learning[J]. IEEE Access, 2017, PP(99):26241-26253.
- [14] Yang, H.-Y. Pan, Y.-G. Li. A novel incremental semi-supervised VPMCD gear fault on-line diagnosis method[J]. Journal of Vibration & Shock, 2015, 34(8):49-54.
- [15] Jin Shoufeng, Fan Di, Malekian Reza. An image recognition method for gear fault diagnosis in the manufacturing line of short filament fibres[J]. Insight: Non-Destructive Testing and Condition Monitoring, 2018, 60(5):270-275.
- [16] M Frini, A Soualhi, M El Badaoui. Gear fault detection using the geometric properties of electrical currents in three-phase induction motor-based systems[J]. International Journal of Condition Monitoring, 2017, 7(2):47-52.
- [17] Vikas Sharma. Case study on the effectiveness of gear fault diagnosis technique for gear tooth defects under fluctuating speed[J]. Iet Renewable Power Generation, 2017, 11(14):1841-1849.
- [18] Li, P. Zhang, B. Li. Application of universal quantum gate neural network in gear fault diagnosis[J]. China Mechanical Engineering, 2015, 26(6):773-777.
- [19] ZENG Ming, YANG Yu, ZHENG Jinde. μ -SVD Based Denoising Method and Its Application to Gear Fault Diagnosis[J]. Journal of Mechanical Engineering, 2015, 51(3):95.
- [20] XU Yong-gang, ZHAO Guo-liang, MA Chao-yong. Application of gear fault diagnosis method based on dual-tree complex wavelet transform and local projective method[J]. Journal of Vibration Engineering, 2015, 28(4):650-656.
- [21] Andreas Holzinger. Interactive Machine Learning (iML)[J]. Informatik Spektrum, 2015, 39(1):1-5.
- [22] Faber, Felix, Lindmaa, Alexander, von Lilienfeld, O. Anatole. Machine Learning Energies of 2 M Elpasolite (ABC₂SD₆) Crystals[J]. International Journal of Quantum Chemistry, 2015, 115(16):1094–1101.
- [23] Jonathan Carifio, James Halverson, Dmitri Krioukov. Machine Learning in the String Landscape[J]. Journal of High Energy Physics, 2017, 2017(9):157.
- [24] Yue Cui, Min Zhang, Danshi Wang. Bit-based support vector machine nonlinear detector for millimeter-wave radio-over-fiber mobile fronthaul systems[J]. Optics Express, 2017, 25(21):26186.
- [25] Elias Giacomidis, Son T. Le, Mohammad Ghanbarisabagh. Fiber nonlinearity-induced penalty reduction in CO-OFDM by ANN-based nonlinear equalization[J]. Optics Letters, 2015, 40(21):5113-6.