Small Fault Diagnosis of Front-end Speed Controlled Wind Generator Based on Deep Learning

Hai-Ying DONG, Li-Xia YANG, Hong-Wei LI
School of Automation & Electrical Engineering
Lanzhou JiaoTong University
Lanzhou Gansu 730070
CHINA
yanglixialzjtu@163.com

Abstract: - In view of the difficulty in diagnosing the early small faults of front-end controlled wind generator (FSCWG), this paper proposes a small fault diagnosis methods based on deep learning. The method adopts a deep learning method, uses vibration data under several different small fault patterns of FSCWG as input of the model and gets deep learning diagnosis model by learning complicated implicit layer structure and training. Then using the trained network to extract feature of FSCWG from original vibration data by layer-wise, and fully excavate the associations among the data and form a more abstract executive property categories or characteristics, to improve the diagnosis accuracy. The results show that compared with the traditional fault diagnosis method of neural network (NN) and support vector machine (SVM) method, the small fault diagnosis method based on deep learning enhances the small fault diagnosis accuracy in the process of generator operation.

Key-Words: - front-end speed controlled wind generator; small fault; deep learning; convolutional neural network

1 Introduction

The front-end speed controlled wind turbine is put forward by German VOITH company in recent years, it uses flexible hydrodynamic transmission to replace the traditional rigidity transmission mode, and the brushless electric excitation synchronous generator directly coupled with power grid, can produce electricity just like conventional hydropower and thermal power using synchronous generator. As the key part of front-end speed controlled wind turbine in energy conversion, the operation conditions of wind generator directly affect the power quality and generation unit performance. Because of the strong coupling and strong nonlinear characteristics, a very tiny exceptions in FSCWG are likely to lead to big swings of another signal which is wide apart in the space and time, and result in the deterioration of generator running state, even worse, it will cause serious accidents. Although in some cases, the initial small fault will not cause damage in a short period of time, however, as the generator operation, failure accumulation and dissemination will have an adverse effects on the generator. Therefore, to study the early small fault diagnosis, track generator operation process in time, diagnose and deal with tiny fault as soon as possible, not only can effectively prevent and reduce the occurrence of major accidents, at the same time, it is also a key means to ensure the safe operation of the generator, reduce the cost of production and management, has important practical significance.

In general, small fault [1] has the characteristics of the early small amplitude, gradually occur and not easy to find. The small fault of control system has two aspects meaning: on the one hand, it is an early stage, weak degree, no obvious symptoms or potential failures; on the other hand, the small fault is a physical significance of another fault early stages, the characteristics of small faults led to the small fault diagnosis difficult [2]. Because of the high cost, when serious fault occurs, wind turbines will be forced to stop, therefore, wind turbine diagnosis generally is given priority to with early fault diagnosis, under the condition that the unit continues to run without affected, to avoid major accident and cause economic losses. At present, for front-end speed controlled wind turbine, the mainly research is the design of the unit and control technology, its small fault diagnosis research is still in its starting stage, but for other types of wind turbines, the early fault diagnosis method has a certain research at home and abroad, these methods can be roughly divided into the following kinds:

signal analysis based diagnosis, model-based diagnosis, statistical analysis based approach, artificial intelligence based diagnosis and synthesis of various methods [3]. Signal analysis uses signal processing algorithm to analyse the characteristics of the collected signal, extracts variables which can reflect the fault characteristics, and then, you need to combine with fault mechanism and field practical experience, artificially to determine whether a fault occurs, the typical methods have observer method [4,5], filter method [6] and so on; the biggest shortage of model based method is the accuracy of diagnostic results significantly depends on the accurate mathematical model, and the method needs to provide a large number of parameters about generator structure, which will lead to huge quantity of calculation. The statistical analysis methods need to artificially set the control limit subjectivity to determine whether fails occur, which increases the possibility of error check and lacking check [7]. Because the artificial neural network has strong self-learning, self-organizing, adaptive and nonlinear mapping ability, it is suitable for uncertain classification problems with complex causality [8], so in recent years, it has been extensively used in wind turbine fault diagnosis. Literature [9] aims at difficulty to extract the stator current characteristics when some early faults occur in wind power generator, and puts forward a new method for wind power generator fault diagnosis that combining the wavelet transform improved by single-band reconstruction and BP neural network; literature [10] uses wavelet to extract the fault characteristics, and artificial neural network for fault intelligent diagnosis, in addition, a signal analysis method combines wavelet analysis with fourier transform is used for traditional fault diagnosis, to verify the validity of the intelligent diagnosis; literatures [11-13] use SVM method for wind turbine fault diagnosis. These methods have made certain achievements, but these methods firstly need to extract fault features from the original data, then fault features as network inputs, through the network learning to realize the nonlinear mapping of characteristics to the fault classification, and training network data gained from the simulation test, which cannot reflect the actual complicated work environment of the wind power generator, in addition, because of the complexity of the wind power generator, and its failure characteristics are multiplicity and cross with each other, it is difficult to obtain reliable and correct diagnosis, and the data get from simulation experiment is bad for wind power generator small fault diagnosis.

According to above situations, this article uses a deep learning approach to small fault diagnosis of FSCWG. As a kind of neural network learning algorithm, deep learning has a broad application prospect, it already has preliminary results in image recognition and speech recognition, it mainly by learning low-level features to form the abstract executive representation, and to find distributed feature of the data. Just like image process and speech recognition, fault diagnosis also belongs to the category of pattern recognition, so this article apply the deep convolutional neural network (DCNN) model to the early small fault diagnosis of FSCWG for the first time, explore the effectiveness of the model under new field and then,simulation analysis was conducted to verify the feasibility and validity of this method as well as make a beneficial attempt for the new fault diagnosis research direction of wind power generator.

2 Small Fault of FSCWG

Because wind power generation system is under complex natural wind, and work in the wild, and bad environment with exposure and thunderstorm for a long time, all these make it prone to various failure. Generator as a core component, it directly impacts on the performance, efficiency and power supply quality of the whole system, it is equally a part in the system prone to failure [14]. Failure types of FSCWG include mechanical failure and electrical failure and there often accompanied by mechanical, magnetic, acoustic, electric, and signs of the insulation system change before a failure occurs. From the mechanical point of view, in running time, the generator inevitably shaking, cycle or intermittent running of the motor will cause winding looseness, bearing wear, and so on; from the perspective of the electric, voltage imbalance or uneven motor winding potential distribution will cause insulation damage; in environmental terms, high temperature and dirt will directly or indirectly cause motor overheating, they are important reasons to accelerate the insulation aging and reduce the insulation performance [15].

When the generator failure occurs, the vibration, noise and temperature of different parts will change, but the vibration signal is the most sensitive about the influence of the fault [16], therefore this article mainly aims at the vibration fault diagnosis research for FSCWG. The vibration fault of FSCWG usually occurs mainly in the rotor, bearing and foundation part, it's common small fault [17,18] is shown in Table 1.

Table 1 Common small fault of FSCWG

Small fault type	Possible cause of the failure	Vibration characteristics
Rotor imbalance	Rotor imbalance fault can be reflected in the rotor quality unbalance, initial bending, hot state imbalance, parts fall off scale, imbalance of coupling, etc. All the imbalance can be considered as rotor quality eccentricity.	(1) The rotor quality unbalance cause larger vibration of the entire unit bearing, and the vibration amplitude increased obviously when the rotor through critical speed; (2) The vibration frequency is consistent with the rotor speed, mainly with one time frequency amplitude, while other harmonic amplitude is small; (3) The axis trajectory is oval.
Rotor contact-rubbing	Rotor contact-rubbing fault mainly refers to friction of the rotor and static, it is caused by the eccentricity of rotor and stator, poor rotor alignment, big rotor dynamic deflection and other reasons.	(1) Except the power frequency, there are abundant higher harmonic components, such as 2×, 3×,...;(2) Axis trajectory shows diffusion and disorderly phenomena; (3) The waveform in the time domain has obviously chopped phenomenon.
Rotor base loose	Loose phenomenon is caused by the bolts not tighten, the base loose or large bearing clearance, etc. Loose will result in severe vibration of the rotor.	(1) Vibration direction often shows the direction of up and down;(2) Except the rotating fundamental vibration frequency, higher harmonic (2×, 3×,...)ingredients will occur, and also fractional harmonic (1/2×, 1/3×,...) and resonance; (3) Because of the nonlinear existence under loose circumstances, when the speed changes, vibration will increase or decrease suddenly.
Shaft misalignment	Shaft misalignment means the tilt and pan between two adjacent rotor axis and bearing centerline.	(1) The axial-radial appears twice frequency vibration, Give priority to once and twice frequency component, the more serious the misalignment, the greater proportion the double frequency is, and likely to appear higher harmonic; (2) When misalignment degree is small the trajectory is elliptic, and be a banana shape or figure 8 shape when moderate and serious degree.
Shaft crack	Shaft crack equivalent to destroy the symmetry of axis interface, in the circumferential direction will exist maximum and minimum bending stiffness.	Produce two times the speed frequency vibration, the larger the crack, the bigger the of the double frequency vibration is, in addition, through the first order critical speed, the vibration of the one time frequency peak also increases, at the same time, the actual critical speed of the rotor system will be reduced accordingly.
Airflow exciting-vibration	When the rotor movement is disturbed by various reasons, there will exit airflow uneven circumferential pressure in the labyrinth cavity, forming unstable exciting force acting on the rotor, mechanical vibration caused by the exciting-vibration is often shown as asynchronous vibration which work below the working speed.	(1) General exciting-vibration frequency is 0.6 to 0.9 times the power frequency;(2) The vibration of the axis trajectory showed changes from small to large, and will reach a saturated state due to the bearing limitation, but because of the existence of the unstable exciting force, axis path turn to change from large to small.

3 Small Fault Diagnosis of FSCWG Based on Deep Learning

3.1 Structure of Deep Convolution Neural Network

As a new classification model, the biggest different of deep learning to artificial neural network is that the neural network mostly only has one or two hidden layer, while the deep learning contains multiple layers, and the connection forms between the neurons of hidden layers are not completely consistent, each neuron only connects with local receptive field of the previous layer. Through deep learning, the data information hidden in the original input data can be extracted and abstracted, the deeper the layer, the deeper the data concept is to express data characteristic, and the more conducive for the final classification, it is what the shallow structure cannot to express and get [19].

The typical structure of DCNN [20] is shown in Fig.1.

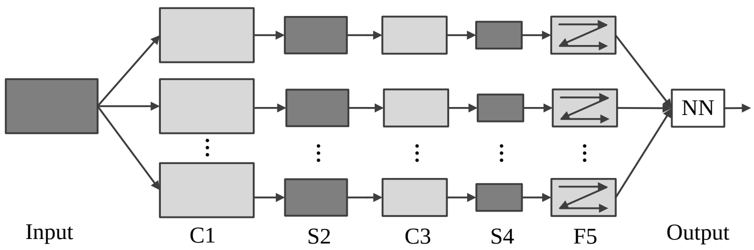

Fig.1 The structure of DCNN

The first layer of DCNN is the input layer, and followed by a number of convolution layer and down sample layer, the last is a classifier, such as softmax classifier, tanh function, clustering algorithm, etc., and the corresponding classification results output by the classifier. Usually, behind each convolution layer has followed a down sample layer. In the convolution layer, local connections and weights sharing can reduce the training parameters of the network and the complexity of calculation. Convolution computation results as input of sigmoid activation function, and the output as feature data of this layer, it is just as the input of the next layer. In convolution layer, each feature data is not necessarily connected with all the feature data of the previous layer, but connected to several among them. In down sample layer, by pooling operation(such as average-pooling or max-pooling), adjacent several features of the corresponding feature data in previous layer is merged into one, to reduce the resolution ratio of characteristics, so as to realize the translation, scaling and distortion invariance. The input vibration data can be directly transmitted to the first convolution layer, and learn the characteristics by layer-wise, finally, using the labeled sample data input to softmax classifier and obtaining the output, and the parameters of the entire network fine-tuned through feedback [21].

3.2 Basic Layers and The Layers Learning of DCNN

For the DCNN, using a way of layer-wise learning from anterior to later. After studying all convolution layers and the down sample layers, using the labeled data input to network, and learned characteristics as classifier input, through BP algorithm fine-tuning trained parameters of the entire network.

DCNN involves three layers and learning of each layer are briefly introduced as follows:

1) Convolution layer

Convolution layer is mainly to extract features through convolution kernels. Assuming that the input of a convolution layer is the output of the previous layer x_i, and do convolution operation with the different subregion of the input x_i and j different convolution kernels W_{ij}. Then through an activation function f, such as the sigmoid function, etc., each time the output of the convolution is mapped to the output characteristics of the layer area, using the formula 1 to get the appropriate output characteristics.

$$h_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * W_{ij}^l + b_j^l\right) \qquad (1)$$

Where, l represents the number of layers, W is the convolution kernel, * is 2-dimensional convolution, b is the bias and M_j represents the selected set of input features.

2) Down sample layer

The down sample layer is mainly for pooling operation, its number of features is the same as the previous adjacent convolution layer, by pooling operation to reduce the spatial resolution of the convolution layer. There are two purposes, one is to reduce the feature dimension; another is that it can strengthen the tolerance of characteristics to small distortion and small rotation. In the input of down sample layer, each size of $r \times r$ overlap area as a pool, and the output can be obtained by commonly used average-pooling operation or max-pooling operation. Output of average-pooling operation is the average value of input matrix, while output of max-pooling operation is the maximum of input matrix [22].

The output characteristics can be calculated by formula 2.

$$h_j^l = f(down(h_j^{l-1}) \cdot \omega_j^l + b_j^l) \qquad (2)$$

Where, ω is the corresponding weights, b is the corresponding bias and *down*( ) is the corresponding down sample function [23]. If the averaging or maximum values of $r \times r$ block eigenvalues of the input characteristic data is calculated, the output characteristics of the data in two dimensions will narrow r times.

3) The full connection layer

The full connection layer is essentially a standard single hidden layer neural network. It vectorizes the input matrix x and mapped to vector y through the weight matrix W, bias vector b and function f.

$$y = f(b + W \cdot x) \quad (3)$$

Where, f is the commonly used tanh function or softmax function. Here is the multiple classification problem, so the softmax classifier is adopted.

For a given test input x, we want to use hypothesis function for each category j to estimate the probability value $p(y=j|x)$, namely, we want to estimate the probability of each classification results of x. Therefore, our hypothesis function will output a k-dimension vector (the sum of vector elements is 1) to represent the k estimated probability value. Specifically, we assume that the function $h_\theta(x^{(i)})$ is as follows

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^{\mathrm{T}} x^{(i)}}} \begin{bmatrix} e^{\theta_1^{\mathrm{T}} x^{(i)}} \\ e^{\theta_2^{\mathrm{T}} x^{(i)}} \\ \vdots \\ e^{\theta_k^{\mathrm{T}} x^{(i)}} \end{bmatrix}$$

Where, $\theta_1, \theta_2, \cdots, \theta_k \in \mathrm{R}^{n+1}$ are the parameters of the model, $\frac{1}{\sum_{j=1}^{k} e^{\theta_j^{\mathrm{T}} x^{(i)}}}$ is to normalize the probability distribution, and the sum of all probability is 1.

Parameter θ is no longer a column vector, but a matrix, and each row of the matrix can be seen as a category that corresponds to the parameters of the classifier. There were k lines in all. So matrix θ can be written as the following form

$$\theta = \begin{bmatrix} \theta_1^{\mathrm{T}} \\ \theta_2^{\mathrm{T}} \\ \vdots \\ \theta_k^{\mathrm{T}} \end{bmatrix}$$

Then the system loss function is

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^{\mathrm{T}} x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^{\mathrm{T}} x^{(i)}}}\right] + \frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2$$

Where, 1{.} is an indicative function, that is, when the values in the curly braces are true, the result is 1, otherwise the result is 0. $\frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2$ is a weight decay, to punish the too big values of the parameters.

For the minimization problem of $J(\theta)$, the parameters of the system is obtained by gradient descent method in this paper, we must take the partial derivatives of loss function. In softmax, the partial derivatives of loss function are as follows

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[x^{(i)}(1\{y^{(i)} = j\} - p(y^{(i)} = j | x^{(i)}; \theta))\right] + \lambda\theta_j$$

Note in the formula, $\frac{\partial J(\theta)}{\partial \theta_{jl}}$ is a vector, denote that it is obtained for the case of category i, so the above formula is partial derivatives formula of one category, we need to request the partial derivative formula for all categories. $\nabla_{\theta_j} J(\theta)$ represents the partial derivatives of loss function for the first parameters of j category.

In the standard implementation of the gradient descent method, in each iteration, model parameters need to be updated as follows $\theta_j := \theta_j - \alpha\nabla_{\theta_j} J(\theta)$, ($j$=1,2,…,$k$). α is learning factor.

3.3 Small Fault Diagnosis of Generator Based on The Deep Learning

3.3.1 Basic Principle

The basic principle of fault diagnosis using deep learning is: the input shaft and output shaft vibration data of generator as input, the diagnosis results of deep learning is the network output. Firstly, using existing fault symptom and the corresponding diagnosis to train deep learning network, and then through deep learning to train hidden layers and extract features by layer-wise, design the corresponding relationship between network fault symptom and diagnosis, get the weights and bias parameters of each layer, and finally, the vibration data input to the network, you can use the after trained deep learning network for fault diagnosis, and the corresponding diagnosis results are obtained.

Input layer: receiving the vibration data of input shaft and output shaft from the monitoring object, and getting processing normalized data $X=(x_1,x_2,\ldots,x_n)$;

Middle layer: by internal learning and processing, the input information is translated into a targeted solution, the middle layer contains two convolution layers, two down sample layers and a full connection layer, it connects with the input layer through numerical W_{ij} and connects with output layer via bias b, sigmoid function is chosen as the activation function to realize the nonlinear mapping from input mode to the output mode.

Output layer: through the comparison of output neurons and bias, the diagnosis results are obtained. Output layer nodes are the fault mode summation. If the output of *j*th mode is (0 0... 0 1 0...0 0), that is, the output of *j*th node is 1, the rest of the output is 0, it represents the *j*th fault occurs (0 represents fault-free mode), if all the outputs are 0, it indicates no failure occurs [24].

3.3.2 Diagnostic Procedure

The main steps of FSWTG small fault diagnosis based on DCNN are as follows:

Step 1 Preprocess the input vibration data x using whiting method, and substitute the input vibration data x with low redundancy expression $\tilde{x}$.

Step 2 With preprocessed data $\tilde{x}$ input to first convolution layer (C1), optimal weights are obtained by convolution autoencoder training, and then will get the characteristics area h1 of the first convolution layer through formula 1.

Step 3 Do down sample operation for characteristic area of the first convolution layer, that is, take the maximum value of every adjacent $r\times r$ points as characteristic data after sampling, therefore, eigenvalues h2 of the first down sample layer (S2) is obtained.

Step 4 The eigenvalues h2 of first down sample layer (S2) as input of the second convolution layer (C3), continue to train by using convolution autoencoder, substitute trained weights into formula 1 will get characteristic value h3 of the second convolution layer.

Step 5 Do down sample operation for the second convolution layer (C3) and get the eigenvalues h3, take maximum value of every adjacent $r\times r$ points as eigenvalues h4 of the second down sample layer (S4).

Step 6 Transform the eigenvalues h4 of second down sample layer (S4) to get the corresponding characteristic matrix, each row corresponding to one input sample data, as the input of the full connection layer, through softmax classifier to obtain the final classification result.

4 Simulated Analysis

This article uses the DCNN with a total of 6 layers, considering both the ability of sample to reflect the continuous change of generator operation process and the amount of calculation, the first input layer select 784×784 input nodes, horizontal coordinate denotes vibration data of generator input shaft, while the vertical coordinate denotes vibration data of generator output shaft. The second and the fourth layer are convolution layers. The first convolution layer has 12 convolution kernels, and the second convolution layer has 24 convolution kernels, the size of convolution kernels is 10×10; the third and fifth layer are down sample layers, the down sampling method is maximum pooling operation, sixth layer is full connection layer, its hidden layer contains 200 neurons, the manifestation layer directly connected to the last pooling layer. The output of the full connection layer contains 6 neurons, and uses softmax function as output, the DCNN structure used in this paper is shown as below

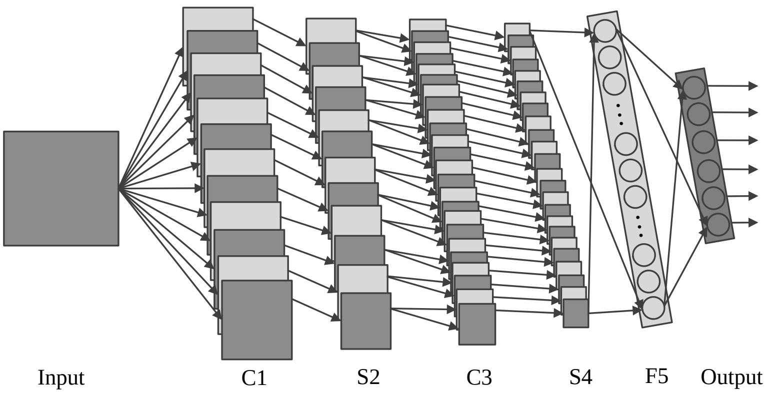

Fig.2 DCNN structure used in this paper

In this experiment, 823202 groups of data are used as training data, and 54880 groups of data as testing data, the rotation speed of FSWCG is 1500 r/min, all data collected from Changma farm in Gansu province.

4.1 Structure of Deep Convolution Neural Network

Because the DCNN contains more than one hidden layer, training a DCNN requires a large amount of calculation. Thus the fast convergence is critical for training algorithm. Learning rate decides weights changes in every training circuit. Big learning rate can lead to system instability, while small learning rate will lead to training for a long time, and the convergence speed is slow, but can guarantee the network error values don't jump out of the trough on the error surface and finally reach a minimum error value. So in general, tend to choose small learning rate in order to ensure the stability of the system. In this article, the influence of different learning rate of network training as showed in the figure below.

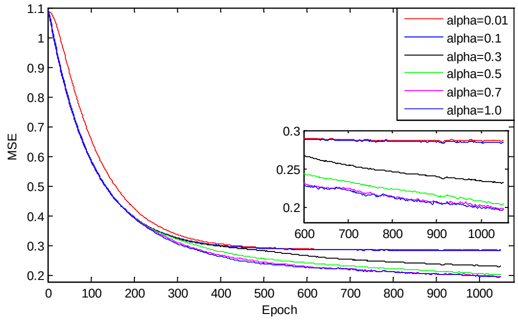

Fig.3 The mean square error (MSE) curve under different learning rate and its partial enlarged detail

The network running time under different learning rate is shown in the following table.

Table 2 Network running time under different learning rate

Learning rate α	Running time/s
0.01	126.646828
0.1	130.829943
0.3	126.850576
0.5	129.362913
0.7	126.361134
1.0	126.641551

Considering the network training time and the speed of convergence, according to the Fig.2 and Table 2, the learning rate in this paper is 1, namely, do not use learning rate.

4.2 Structure of Deep Convolution Neural Network

Comparing the deep structure network proposed in this paper and traditional shallow structure NN and SVM methods, the partial results are shown in the following table.

Table 3 The patial learning results of traditional method and the method proposed in this paper

Failure mode	Output coding	Methods	The output after training					
Rotor imbalance	1 0 0 0 0 0	NN	0.764412	0.069565	0.161324	0.451745	0.092232	0.041408
		SVM	0. 890856	0.078948	0.143041	0.130542	0.136351	0.056936
		DCNN	0.9819846	0.0000315	0.0035247	0.0050306	0.0057029	0.0037257
Rotor contact-rubbing	0 1 0 0 0 0	NN	0.189785	0.869481	0.128067	0.125053	0.096247	0.104098
		SVM	0.095063	0. 889792	0.082127	0.114703	0.094461	0.069591
		DCNN	0.0000229	0.9827393	0.0045028	0.0071321	0.0008748	0.0047281
Rotor base loose	0 0 1 0 0 0	NN	0.040171	0.396251	0. 890489	0.081319	0.026836	0.087655
		SVM	0.050272	0.267395	0. 893683	0.039469	0.026241	0.101482
		DCNN	0.0090210	0.0072575	0.970583	0.0014913	0.0071397	0.0045085
Shaft misalign-ment	0 0 0 1 0 0	NN	0.223838	0.095311	0.046468	0.747907	0.036082	0.208351
		SVM	0.408692	0.073284	0.030068	0.878032	0.018628	0.154915
		DCNN	0.0064562	0.0130410	0.0063563	0.9739125	0.0000243	0.0002097
Shaft crack	0 0 0 0 1 0	NN	0.280351	0.039488	0.215483	0.034531	0.898126	0.038098
		SVM	0.323685	0.016781	0.157567	0.027565	0.898616	0.031258
		DCNN	0.0000243	0.0000525	0.0094179	0.0000040	0.9905008	0.0000006
Airflow excitingvibration	0 0 0 0 0 1	NN	0.210935	0.038465	0.116727	0.059382	0.024503	0.792738
		SVM	0.173154	0.034654	0.084174	0.053941	0.041026	0.869585
		DCNN	0.0000324	0.0000320	0.0050633	0.0000000	0.0007369	0.9941354

Table 4 Comparision between the traditional methods and DCNN

Methods	Running time/s	Accuracy rate
NN	73.63	71.92%
SVM	64.27	79.38%
DCNN	126.64	92.74%

It can be seen from Table 3 and Table 4, the fault diagnosis accuracy of deep learning method is obviously improved comparing to NN and SVM methods, verified the effectiveness of the proposed method, but compared to traditional methods, deep learning have more parameters need to learn and adjust, so the operation time is longer than the traditional methods, but it can effectively solve the reliable and accurate diagnosis problem of FSCWG, which is caused by multiplicity and overlapping of fault characteristics.

5 Conclusion

Due to the complexity, time-varying and non-stationary of FSCWG fault signal, early fault is vulnerable to be eliminated by noise and the resonance signal components, it is difficult for early small fault diagnosis to use the general methods, therefore, to extract the effective classification characteristics from original vibration signal of FSCWG is very important. In this paper, a FSCWG small fault diagnosis method based on DCNN is proposed, the method built multi hidden layer learning model, and through the layer-wise initialization and fine-tuning training mechanism, will be able to get more useful characteristics in the original vibration signal which was collected from wind farm, thus to improve the accuracy of fault diagnosis and classification. With the measured data under different operating conditions of wind farms which can be used to verify the model, can truly reflect the small fault of the wind turbine, has more practical application value. The results show that compared with the traditional NN and SVM method, the proposed approach significantly improves the recognition rate of fault diagnosis, to verify the effectiveness of the proposed method. In further research, we will continue to study a variety of deep learning model and model parameters optimization, to furtherly enhance the accuracy of fault diagnosis.

6 ACKNOWLEDGEMENTS

This work is supported by Natural Science Foundation of Gansu Province (145RJZA136), Natural Science Foundation of Gansu province (joint) project (213244) and Youth Science Foundation of LanZhou JiaoTong University (2013040).

References:

[1] LI Juan, ZHOU Dong-hua, SI Xiao-sheng, CHEN Mao-yin, XU Chun-hong. Review of incipient fault diagnosis methods, *Control Theory & Applications*, Vol.29, No.12, 2012, pp. 1517-1529.

[2] LIU Chun, JIANG Bin, ZHANG Ke, WU Yun-kai. Incipient Fault Detection of Linear System with Disturbance, *Safe process China'2015*, Shanghai, July, 2015, pp. 70-77.

[3] PENG Hua-dong, CHEN Xiao-qing, REN Ming, YANG Dai-yong, DONG Ming. Intelligent Fault Diagnosis Technology and System for Wind Turbines, *Power System and Clean Energy*, Vol.27, No.2, 2011, pp. 62-70.

[4] Freire, N.M.A. , Estima, J.O. , Cardoso, A.J.M.. A New Approach for Current Sensor Fault Diagnosis in PMSG Drives for Wind Energy Conversion Systems, *IEEE Transactions on Industry Applications*, Vol.50, Issue2, 2014, pp. 1206-1214.

[5] Sanchez,H., Escobet,T., Puig,V., Odgaard,P.F.. Fault Diagnosis of an Advanced Wind Turbine Benchmark Using Interval-Based ARRs and Observers, *IEEE Transactions on Industrial Electronics*, Vol.62, Issue6, 2015, pp. 3783–3793.

[6] Dey,S., Pisu,P., Ayalew,B.. A Comparative Study of Three Fault Diagnosis Schemes for Wind Turbines, *IEEE Transactions on Control Systems Technology*, Vol.23, Issue5, 2015, pp.1853–1868.

[7] Wang Tian-Zhen, Xu Man, Tang Tian-Hao, Christophe Claramunt. A fault detection method based on dynamic peak-valley limit under the non-steady conditions, *WSEAS Transactions on Circuits and Systems*, Vol.12, No.4, 2013, pp. 118-128.

[8] MI Jiang, JI Guoyi. Application of improved BP neural network in fault diagnosis of Fans, *Noise and Vibration Control*, Vol.31, No.2, 2011, pp. 94-98.

[9] GUO Dongjie, WANG Lingmei, GUO Honglong, WU Weihong, HAN Xigui. Fault Diagnosis of Wind Power Generator Based on Improved Wavelet and BP NN, *Proceedings of the CSU-EPSA*, Vol.24, No.2, 2012, pp. 53-58.

[10] ZHANG Xiang-luo. Research on system of online fault feature extraction and fault

diagnosis for wind turbine generator, *South China University of Technology*, Guangzhou, China, 2013.

[11] ZENG Jianwu, LU Dingguo, ZHAO Yue, ZHANG Zhe, QIAO Wei, GONG Xiang. Wind turbine fault detection and isolation using support vector machine and a residual-based method, *American Control Conference (ACC)*, 2013, pp. 3661 -3666.

[12] JIAO Bin, HAO Yun-suo. Support Vector Machine with Composite Kernel in Fault Diagnosis of Wind Power Gearbox, *Journal of Jiangnan University(Natural Science Edition)*, Vol.12, No.5, 2013, pp. 607-612.

[13] AN Xueli, ZHAO Minghao, JIANG Dongxiang, et al. Direct-drive wind turbine fault diagnosis based on support vector machine and multi-source information, *Power System Technology*, Vol.35, No.4, 2011, pp. 117-122.

[14] ZHUANG Zhe-min, YIN Guo-hua, LI Fen-lan, JIANG Zhong-wei. Fault Diagnosis of Wind Power Generation Based on Wavelet Neural Network, *TRANSACTIONS OF CHINA ELECTROTECHNICAL SOCIETY*, Vol.24, No.4, 2009, pp. 224-228.

[15] YANG Ming，SHI Wei-feng. Fault diagnosis of ship generators based on multi-population genetic neural network, *Journal of Shanghai Maritime University*, Vol.34, No.4, 2013, pp. 18-22.

[16] WANG Shan, SU Lu-wei, GU Yu-jiong, YANG Kun. Fault Diagnosis Approach for Incipient Bearing Fault in Wind Turbine under Variable Conditions, *Electric Power Science and Engineering*, Vol.30, No.3, 2014, pp.39-44.

[17] WU Zhao-tong, YANG Shi-xi, WANG Wei-jun, et al. *New Methods for Rotating Machinery Fault Feature Extraction and Pattern Classification*, Sciences Press, 2012.

[18] SHI Xian-jiang, WANG Gui-rong, SI Jun-shan. *Mechanical Fault Diagnosis and Typical Case Analysis*, Chemical Industry Publishing House, 2013.

[19] LIN Jin-bo. Application of Recognition Model for Power Load in Cluster Ensemble and Deep Learning. *South China University of Technology*. Guangzhou, China, 2014.

[20] Eklund P, Villerd J.A. Survey of Hybrid Representations of Concept Lattices in Conceptual Knowledge Processing, *Lecture Notes in Computer Science*, 2010, pp.296-311.

[21] DUAN Bao-bin, HAN Li-xin. Improved Convolutional Neural Networks and Its Application in Stitching of Scrapped Paper. *Computer Engineering and Applications*, Vol.50, No.9, 2014, pp. 176-181.

[22] CAI Ying-feng, WANG Hai, CHEN Long, JIANG Haobin. Robust vehicle recognition algorithm using visual saliency and deep convolutional neural networks, *Journal of Jiangsu University(Natural Science Edition)*, Vol.36, No.3, 2015, pp. 331-336.

[23] QU K.-S., LIANG J.-Y., WANG J.-H., et al. The algebraic properties of concept lattice. *Journal of Systems Science and Information*, Vol.2, No.2, 2004, pp. 271-277.

[24] XU Juan. *Fault Diagnosis System of Machinery Vibration Based on the Neural Network*. Shenyang University of Technology, Liaoning, Shengyang, 2007.