Fault Diagnosis of Pitch System of Wind Turbine based on Improved Stacked Auto-Encoder Network

Sihua Wang [1 2], Tian Wang [1 2], Lijun Zhou [1 2], Tianyu Chen[1 2], Yu Wang [1 2]

[1]College of Automation & Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China
[2]Rail Transit Electrical Automation Engineering Laboratory of Gansu Province, Lanzhou Jiaotong University, Lanzhou 730070, China

Abstract

In order to improve the accuracy of fault diagnosis of the wind turbine's pitch system, an improved stack autoencoder network is proposed. Based on the Supervisory Control And Data Acquisition (SCADA) data of the wind turbine's electric pitch system, the batch normalization (BN) algorithm was introduced for the gradient dispersion problem in the feature extraction of ordinary autoencoder networks when there are many parameters. This article uses the Adam optimizer to iteratively update the neural network weights based on the training data. Then calculate the cross-entropy loss function and train the network with the minimum loss function as the goal. Finally, the Softmax classifier is used, and its output is the diagnosis and probability of each component of the pitch system. The data set in the pitch control system SCADA is selected. This paper selects the verification set in the pitch control system SCADA and substitutes it into the ordinary stack autoencoder and improved stack autoencoder network (SAE) for comparison and verification. The verification results show that the batch-standardized SAE network has a more optimized network model and higher recognition accuracy, and also provides a strategy for fault diagnosis of wind turbines.

Keywords: Insulator; batch standardization; stacking autoencoder; fault diagnosis; wind turbine

Received: 6 April 2020

To cite this article:
Sihua Wang, Tian Wang, Lijun Zhou, Tianyu Chen, Yu Wang, "Fault diagnosis of pitch system of wind turbine based on improved stacked autoencoder network ", in *Electrotehnica, Electronica, Automatica (EEA)*, 2020, vol. 68, no. 3, pp. 47-54, ISSN 1582-5175. https://doi.org/10.46904/eea.20.68.3.1108006

1. Introduction

In the long-term operation of the wind turbine, various component failures will occur frequently. The electric pitch system, as one of the core components of the large-scale variable-speed constant-frequency wind turbine control system, plays an important role in the safe and stable operation of the unit. However, due to the randomness and uncertainty of wind speed and the harsh operating environment, it has become one of the components with high failure rate [1].

Therefore, it is of great significance for the reliability of wind turbines to adopt reasonable and efficient monitoring methods and accurate fault diagnosis methods.

At present, the research on fault diagnosis of wind turbine is mainly based on the operation data of wind turbine data SCADA system [2].

In the era of big data, it is difficult to mine deeper feature information by relying on the traditional mechanical fault diagnosis "feature extraction + pattern recognition" mode, which reflects the characteristics of large amount of data, many forms, and low value density reflecting the status of equipment. Reference [3], based on the wind farm SCADA system data, proposes a method for identifying abnormal state parameters of wind turbines, and uses neural networks to establish state parameter prediction models.

In reference [5], researchers have proposed some advanced signal processing techniques to extract useful fault features for diagnosis. However, these advanced methods often require researchers to have a deep understanding of the induction motor system and its fault signals. In addition, because the required expert knowledge is not easy to obtain, these methods are not universal, which means that these methods are not intelligent enough compared with machine learning.

Reference [6], Guo Huidong and other scholars proposed an on-line identification method for degradation state of wind turbine pitch system according to the requirements of real-time state maintenance of wind turbine.

To some extent, the traditional wind turbine fault diagnosis method realizes fault diagnosis and state recognition, but it cannot automatically mine deeper information from wind turbine big data.

Deep learning can well realize the representation of complex high-dimensional functions, so it has strong representation ability and has incomparable advantages in feature extraction [7].

Deep learning opens up new ideas for intelligent fault diagnosis. In reference [8], Zhao Hongshan scholars designed a fault diagnosis algorithm based on deep autoender (DAE) network and XGBoost by analysing the SCADA data of wind turbine.

Reference [9] Mitra and other scholars, when the wind turbine enters a certain predefined situation, collect the vibration signal through the accelerometer, and then use the standard deviation to identify the defective bearing.

Reference [10], Cong Wei and other scholars proposed a network-based meteorological disaster prevention method based on Synthetic Minority Oversampling Technique (SMOTE) and stacked denoising autoencoder (SDAE), which can accurately and comprehensively establish the association mapping relationship between meteorological information and power grid faults. Whether the specified meteorological conditions will lead to the occurrence of power grid disaster accidents can be accurately predicted.

In reference [11], a novel intelligent fault diagnosis method is proposed for wind turbine rolling bearings based on Mahalanobis Semi-supervised Mapping (MSSM) manifold learning algorithm and Beetle Antennae Search based Support Vector Machine (BAS-SVM).

In [12], Zhang Xining and other scholars introduced a standardization strategy and used standardized SAE networks to diagnose rolling bearing faults. At present, the research on SAE in the field of mechanical fault diagnosis is mostly in the stage of introduction and application of artificial intelligence, and there is still a lack of network performance optimization analysis and model improvement.

In this paper, based on the SCADA data of a wind turbine pitch system of a certain electric field, the batch standardization is referenced to the SAE network. By optimizing the network performance and using the powerful nonlinear feature mining ability of the deep neural network, the SCADA system data and faults are established. Mapping model to realize the state identification and fault diagnosis of the wind turbine pitch system.

2. Deep network model construction

2.1 Auto-ender

Auto-encoded networks are a type of unsupervised learning. A representational learning algorithm proposed by Rumelhart in 1986 [13]. The essence of the self-encoding network is the process of initializing network parameters. The entire process can be divided into two parts, encoding and decoding. The structure is shown in Figure 1.

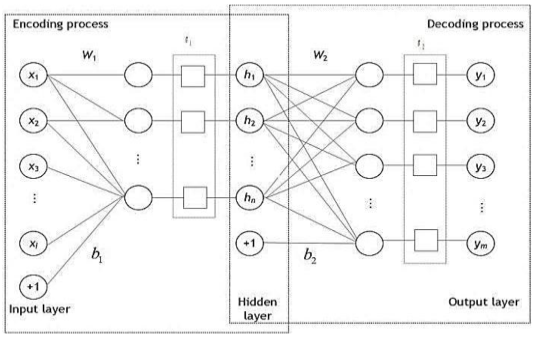

Figure 1. Autoencoder network structure

The network encoding part reconstructs the input value and extracts features; the decoding part compares the reconstructed value with the original input value and calculates the parameters of better data.

The expressions for the encoding and decoding parts are:

$$h = \sigma_1(W_1 x + b_1) \tag{1}$$

$$y = \sigma_2(W_2 h + b_2) \tag{2}$$

In the formula: x, h, y are the input of the autoencoder, the output of the hidden layer, and the output of the output layer; σ_1 and σ_2 represent activation functions for encoding and decoding; W_1 and b_1 represent the weight and offset of the encoder. W_2 and b_2 represent the weight and offset of the decoder.

In this process, in order to make the final reconstruction result can restore the input sample as much as possible. With the input and output errors, such as the minimum mean square error as the goal, the autoender parameter θ are optimized:

$$E = \min_{\theta}(L(x, y)) \tag{3}$$

$$L(x, y) = \frac{1}{m}\sum_{i=1}^{m} \| x - y \|^2 \tag{4}$$

$$\theta = [W_1, W_2, b_1, b_2] \tag{5}$$

In the formula: L is loss function, the mean square error of input *x* and output *y* is generally selected; *E* is the minimum loss error; m is number of training samples; θ is autoencoder parameters. The loss function *L* is minimized by gradient descent method..

2.2 Training of stacked autoencoder networks (SAE)

Hinton proposed SAE network for further improvement of autoencoder [14]. SAE transforms a complex input data into a series of simple high-order features, which is one of the most important research hotspots in the field of deep learning. At present, there have been research results in the field of equipment condition detection, among which literature [15] has verified the effectiveness of SAE deep learning method for generator condition monitoring and fault diagnosis.

In reference [16], based on the data of the state and monitoring system, a depth automatic encoder is used to monitor the parameters of the blades of the wind turbine and diagnose the faults of the blades.

We used an auto-coder to build a deep network model [17] (Figure 2.).

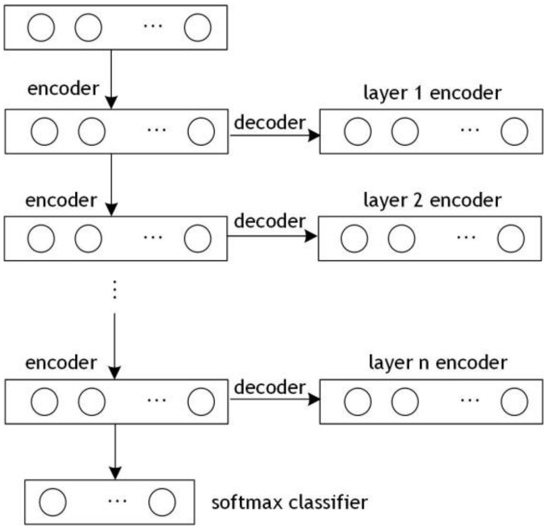

Figure 2. Deep network model

In Figure 2, there are presented two processes:

(1) Unsupervised layer-by-layer training, which trains multiple autoencoders;

(2) Supervised fine-tuning, that is to say, several self-coding parts after training are connected in series one by one, the parameters of the trained coding layer are fine-tuning by using the sample tag, and the classifier softmax is added in the last layer to realize the classification function.

For a training set

$\{(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),...,(x^{(n)},y^{(n)})\}$;

Its input feature is: $x^{(i)} \in R^{n+1}$;

the class label is: $y^{(i)} \in \{0,1,...,k\}$.

Suppose the function *f* calculates the probability value $p(y=j\,|\,x)$ that sample *x* belongs to each class *j*. The softmax function expression is as follows:

$$f_\theta(x^{(i)}) = \begin{bmatrix} p(y=1\,|\,x;\theta) \\ p(y=2\,|\,x;\theta) \\ \cdots \\ p(y=k\,|\,x;\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \cdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (6)$$

In the formula: θ is the vector; $\theta_1,\theta_2,...,\theta_k$ are the parameters of the model, f is the hypothesis function; k is the number of probability values; i is the number of training samples; n is the number of training samples; j is the category.

Selecting different non-linear activation functions for the training SAE will result in different expression and generalization abilities, making them have different training effects.

Currently, the commonly used activation functions are ReLU function, sigmond function, and tanh function. When the input of the ReLU function is a positive number, the problem of gradient explosion and gradient disappearance is avoided, and the calculation speed is faster.

But when the input is negative, ReLU is completely inactive. In this paper, the PReLU function is selected as the activation function [18].

The activation function can adaptively learn and correct the parameters of the linear unit, effectively alleviating the hard saturation problem of the conventional ReLU function on $(-\infty,0)$, and helping to enhance the feature extraction ability and learning ability of the network.

$$ReLU(x) = \begin{cases} x & if x > 0 \\ 0 & if x \le 0 \end{cases} \quad (7)$$

$$PReLU(x_i) = \begin{cases} x_i & if x_i > 0 \\ a_i x_i & if x_i \le 0 \end{cases} \quad (8)$$

In the formula: *a* is changed according to the data, initialized to 0.25; *i* means different channels.

3. Analysis and optimization of network performance

3.1 Adam Optimizer

In this paper, Adam algorithm is selected when optimizing network parameters.

In 2015, Diederik Kingma and Jimmy Ba proposed Adam algorithm in ICLR paper[19]. Adam algorithm is a first-order optimization algorithm, which can replace the traditional stochastic gradient descent (SGD) process. It can train the data to iteratively update the neural network weights. Calculate the first-order moment estimation and second-order moment estimation of the gradient to design independent adaptive learning rates for different parameters. The algorithm formula is as follows:

$$\begin{cases} g = \frac{1}{m}\nabla\theta\sum_i (f(x^i;\theta),y^{(i)}) \\ s = \rho_1 s + (1-\rho_1)g \\ r = \rho_2 r + (1-\rho_2)g^2 \\ \hat{s} = \frac{s}{1-\rho_1^t}, \hat{r} = \frac{r}{1-\rho_2^t} \\ \theta = \theta - \varepsilon\frac{\hat{s}}{\sqrt{\hat{r}}+\delta} \end{cases} \quad (9)$$

In the formula, θ is the initial parameter; m is the number of samples; $x^{(i)}$ is the small batch of training samples; $y^{(i)}$ is the objective function with parameter θ ; g is the gradient of the objective function derivation of θ ; s,r are the first moment estimation and second moment estimation of the gradient; ρ_1,ρ_2 are the decay rate of the moment estimation, usually

0.9 and 0.999; $\hat{s}, \hat{r}$ are the correction deviations; δ is a numerically stable constant, which is generally10-8 ; a is the learning rate, generally 0.001.

3.2 Batch standardization SAE

After the data passes through the layer-by-layer network, the output data distribution will change, which brings difficulties to the next layer of network learning and slows the learning speed of the entire network. Although the PReLu activation function selected in this paper can alleviate the gradient saturation, if the input values of neurons in one layer of SAE are mostly concentrated in the negative range, the expression ability of the back network will still decline.

Therefore, a batch normalization (BN) algorithm strategy is introduced, so that the value of the activation function falls in the region where the nonlinear function is sensitive to the input and solves the problem of the loss of learning ability of the neuron in the saturation stage [20].

The principle of BN algorithm is shown in Figure 3.

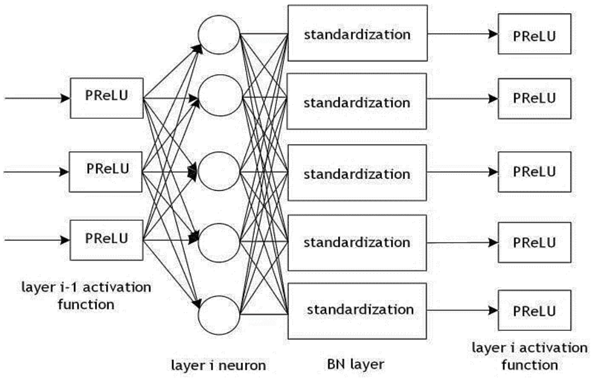

Figure 3. Join the network structure of the BN layer

BN layer is added between all layers of neural network and activation function. wo network parameters γ and β are learned in each network layer. As long as these two network parameters are obtained, the network input X of the upper layer can be normalized to get y as the input of the next network layer. The algorithm formula is as follows:

$$\begin{cases} \mu_b = \dfrac{1}{m}\sum_{i-1}^{m} x_i \\ \sigma_b^2 = \dfrac{1}{m}\sum_{i-1}^{m} (x_i - \mu_b)^2 \\ \hat{x} = \dfrac{x_i - \mu_b}{\sqrt{\sigma_b^2 + \varepsilon}} \\ y_i = \gamma \hat{x}_i + \beta \end{cases} \tag{10}$$

In the formula, $b = \{x_1, \ldots x_m\}$; m is the number of training samples; x_i, y_i are the activation values of the neurons before and after normalization; μ_b, σ_b^2 are the mean and variance of the smallest batch; ε is a small stable constant, γ and β are optimizable parameters of the BN layer.

4. Fault diagnosis and establishment process of SAE wind turbine pitch system based on standardization

The fault diagnosis of wind turbine pitch system based on SAE is mainly divided into two steps:

- first step is based on the operation data and deep learning theory of wind turbine SCADA system, the mapping relationship between the characteristic parameters of wind turbine pitch system and the fault type of pitch system is determined.
- the second step is based on the established mapping relationship, the pitch fault diagnosis is carried out for the given test set, and the mapping relationship is further established by adjusting the diagnosis results.

Fault diagnosis process shown in Figure 4.

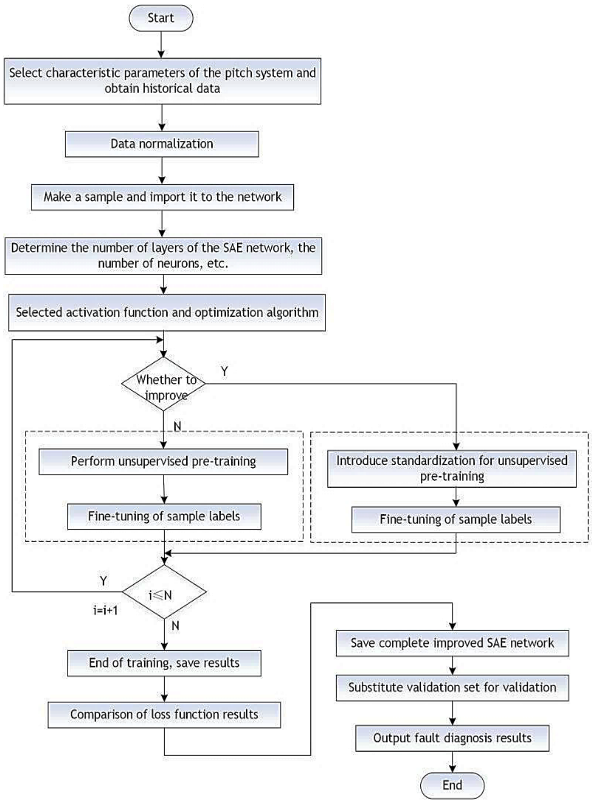

Figure 4. Fault diagnosis flowchart of wind turbine pitch system

5. Example analysis

5.1 Data acquisition and variable selection of wind turbine pitch system

To realize the safe and stable operation of the wind turbine's pitch system, it is necessary to monitor the operating parameters of each component of the pitch system.

The main parameters of the online monitoring include the operating parameters of the pitch motor, blade pitch torque, battery cabinet voltage, blade pitch angle, and control cabinet temperature. In order to quantitatively analyse the accuracy and versatility of the proposed algorithm in fault diagnosis of the wind turbine pitch system, this article selects the SCADA data of a certain electric field Huarui 3 MW wind turbine pitch system under normal conditions from 1 October to 28 December 2019.

In order to ensure the validity of the data in the research process, when selecting the research data, the relevant data of the wind turbine unit in the process of zero power is eliminated, and the data of the process of negative power growth and the data of the process of power reduction from positive value to zero or negative value are eliminated [21].

After abnormal data processing, the number of samples is 40000 × 10, and the total time is one month.

According to the calculation results of the weights of the primary parameters of the pitch system, the parameters with a larger weight for the pitch system fault are selected as the characteristic parameters of the pitch system fault diagnosis.

The characteristic parameters given in Table 1.

Table 1. Related parameters of pitch system

Parameter number	Characteristic parameter
B_1	1 # blade B encoder
B_2	2 # blade B encoder
B_3	2 # blade B encoder
F_1	1 # blade pitching torque
F_2	2 # blade pitching torque
F_3	3 # blade pitching torque
T_1	1 # Blade motor temperature
T_2	2 # Blade motor temperature
T_3	3 # Blade motor temperature
U_1	1 # blade battery voltage
U_2	2 # blade battery voltage
U_3	3# blade battery voltage
V	impeller speed

The collected data samples are used for general SAE network and batch standardized SAE network model training.

The data samples are divided into training set, verification set, and test set according to 9:2:1.

The final network model is determined by training the SAE network with different parameters.

The number of network layers is 10, and the entire model is fully connected.

The encoder has 5 layers and the decoder has 5 layers.

The number of neuron inputs and outputs in each layer are (13, 64), (64, 128), (128, 256), (256, 256), (256, 128), (128, 256), (256, 256), (256, 128), (128, 64), (64, 14).

This paper uses Google's TensorFlow to build a deep learning framework in the Anaconda environment.

We used the Pycharm encoder to write the code. After selecting the super-parameter, the training data set is brought into the established model for training, and the weight updating module is used to automatically update the weight.

5.2 Verification results and analysis

5.2.1 Pitch system fault diagnosis result output

The output of the verification is determined based on the output probability of the last Softmax classifier.

Output probability refers to the probability of failure of each component of the output of the stacked self-coding network.

The maximum probability is determined as the final output result. The probability of occurrence is for a single component. The data of a group of faulty samples in the 18th unit of Huarui Phase III is selected for verification.

The verification results are shown in Table 2.

Table 2. Simulation parameter setting

Fault type	Occurrence probability	Diagnostic result	Actual results
The angle difference between the three main encoders is more than 2°	1.55 %	pitch universal drive fault	reactor malfunction
The angle difference between the three B encoders is more than 2°	0.67 %		
Pitch position control error is greater than 0.1°	0.58 %		
pitch universal drive fault	81.38 %		
Overvoltage fault of DC bus of pitch converter	0.02 %		
Undervoltage fault of DC bus of pitch converter	0.49 %		
Pitch inverter over temperature fault	0.43 %		
Pitch main state voltage fault	0.09 %		
Pitch battery cabinet fault	0.43 %		
Pitch motor fault	13.55 %		
Pitch position fault	0.81 %		

Anomalies in the pitch system are usually reflected by abnormal changes in variables at a certain location.

With the further development of the anomaly, the operating status of the pitch system may affect the operation of the entire unit.

Taking this fault event as an example, the probability of a fault occurring in the diagnosis is shown in Table 2, and the highest probability is judged as a component failure.

Therefore, the fault diagnosis result is a pitch drive failure. Pitch fault handling process and failure analysis: The failure probability of each component in the table is the failure probability of other components due to the failure of the pitch drive.

According to the pitch drive fault indication, the cause of the fault is IGBT overcurrent or motor overcurrent, and the motor failure probability in Table 2 is 13.55 %. The probabilities of other variables have been fluctuating around zero values, which indicates that pitch drive has caused the pitch motor to overcurrent.

In practice, faults of the pitch motor, battery cabinet, emergency module, angle encoder, and wiring are eliminated by replacing parts that may be faulty. The fault point is locked to the reactor, and the blade 1 reactor is replaced. After the maintenance is released, when the blade is at the preparation angle, the motor does not overheat, and the fan starts to run. Consistent with network diagnosis results, indicating that the fault diagnosis results are accurate. If the diagnosis results are not consistent with the actual situation, the network model needs to be revised and trained to obtain a more optimized and accurate model.

5.2.2 Comparison between the number of iterations and different SAE network training

A BN layer is added between the full connection layer and the activation function of the neural network to build a batch standardized SAE network.

Compare the network loss values of normal SAE and standardized SAE, and the comparison results are shown in Figure 5.

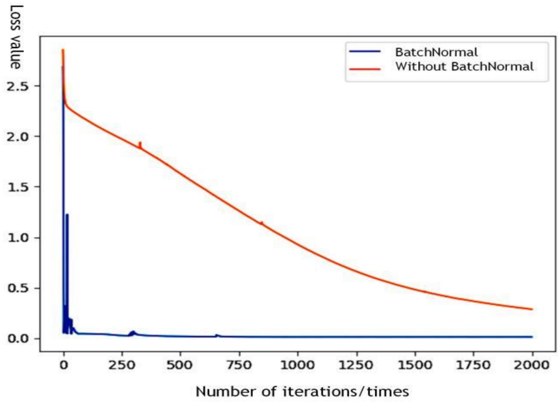

Figure 5. Comparison chart of SAE network loss value training

As can be seen from Figure 5, the loss value of the ordinary SAE network gradually decreases with the increase of the number of iterations during training; but the loss of the standardized SAE network decreases significantly faster with the same number of iterations. The loss value of the standardized SAE network drops rapidly and reaches a stable value at the beginning of training. When the number of iterations is more than 150, the loss value keeps stable and converges to more cells, which greatly improves the fault identification ability of the autoender network. Therefore, the standardized SAE network improves the learning performance and has better training accuracy. In this paper, the number of iterations is 150, the training accuracy meets the expected requirements, and better network parameters are found.

5.2.3 Sample number and different SAE network analysis

To further illustrate the advantages of SAE networks based on batch standardization over ordinary SAE networks in feature extraction and pattern recognition of high-dimensional large-capacity data, about 3614 groups of data were collected, and the number of samples in each fault state was 278, among which 1807 groups were training samples and the remaining 1807 groups were test samples.

Comparing the training samples and test samples on the general SAE network and the batch standardized SAE network (Setting up a common SAE and a batch standardized SAE have the same learning algorithm and incentive function). The correct recognition rate of training set and test set under different sample numbers is shown in Figure 6 and Figure 7.

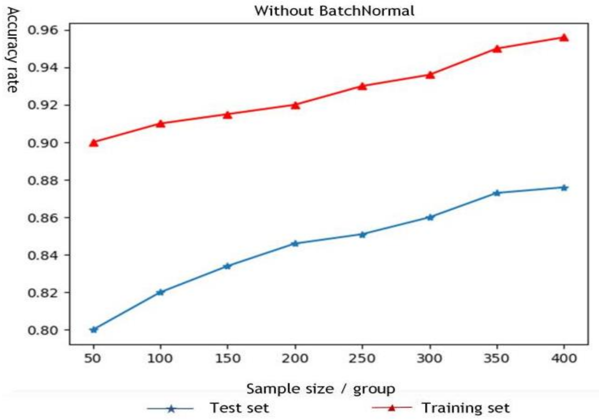

Figure 6. Judgment results of ordinary SAE networks under different iterations

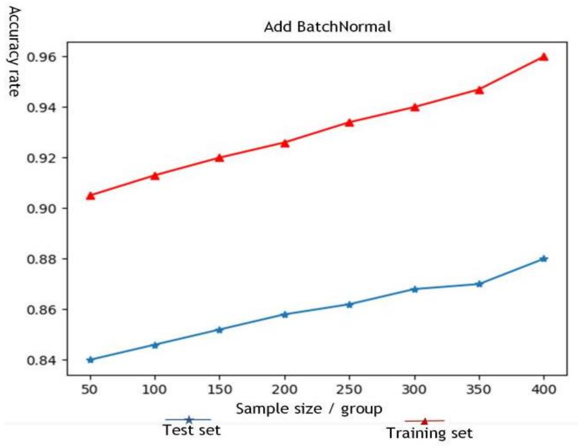

Figure 7. Judgment results of standardized SAE networks under different iterations

It can be seen from Figures 6 and 7 that when the number of samples is small, the recognition effect of the two SAE networks is relatively low, because the deep learning method itself requires a large amount of data.

As the number of samples increases, the correct rate of fault judgment is constantly increasing in the training samples and test samples of ordinary SAE networks and standardized SAE networks.

When the number of samples continues to increase, the test accuracy of ordinary SAE networks will be closer to that of standardized SAE networks; the training accuracy of ordinary SAE networks is also closer to the accuracy of standardized SAE network tests. This shows that the training process of the standardized SAE network's fault diagnosis method is extremely important.

But with the same sample size, the correct rate of batch standardized SAE network is significantly higher than that of ordinary SAE network when tested through the test set.

Therefore, for multi-dimensional sample data, compared with ordinary SAE, batch standardized SAE has higher recognition accuracy and better extraction ability. The more training samples and input types, the more features the corresponding mapping relationship can obtain, the more reliable the method will be, and the network model will be more optimized.

5.3 Method comparison

In order to further illustrate the advantages of this method compared with the traditional method, the results of BP neural network, SVM classifier and SAE network are compared with those of the standardized SAE (BN-SAE) network (Setting up BP network, SAE network and bn-sae network have the same learning algorithm and incentive function).

The number of iterations of SAE network and BN-SAE network is 150, and that of BP network is 220. The core parameters of SVM are selected as (C,σ)=(110,11) . Change the size of the number of training samples, the results are shown in Table 3.

Table 3. The accuracy of different networks under different sample numbers

Number of samples	Correct recognition rate			
	BP	SVM	SAE	BN-SAE
50	0.364	0.625	0.902	0.906
100	0.439	0.746	0.914	0.917
150	0.538	0.823	0.918	0.922
200	0.530	0.834	0.920	0.931
250	0.519	0.820	0.928	0.938

Compared with traditional methods, the accuracy of common SAE network and BN-SAE network has been greatly improved.

The BP network is easy to produce the problem of non-absorption and over fitting when dealing with large-scale samples. The SVM classification is better than BP, but the recognition rate is still low.

Compared with SAE, bn-sae network has higher recognition accuracy in each sample number. Therefore, the BN-SAE network has more advantages.

6. Conclusion

In this paper, the data sample of Wind Turbine Pitch System SCADA is taken as the input feature. In order to solve the problem of gradient dispersion in the process of feature extraction from traditional self-coding networks, an improved SAE network is proposed. The algorithm proposed in this paper is verified by experiments. The conclusions are as follows:

The network model based on batch standardization is verified. The results show that the model can effectively diagnose the pitch system and meet the actual needs. However, the fault sample is a short-term fault, and the accuracy of the gradual fault recognition still needs to be collected for further training and verification.

The performance of general SAE network and batch standardized SAE network is better than that of traditional methods. In the process of sample training, the increase of the number of iterations and samples has an impact on both SAE networks, the batch standardized SAE network has better feature extraction ability and higher recognition accuracy.

Using deep learning to diagnose wind turbine faults is one of the important research directions in this field in the future. In this paper, the SAE network is improved by optimizing the network performance to realize the fault diagnosis of wind turbine pitch system, which has important research value and good application prospect.

7. References

[1] YAN, Y. L. “Investigation to approaches of online condition assessment and short-term reliability prediction of wind turbines”, Chongqing, China: *Chongqing University*, 2015.

[2] Qiu, Y. N., Feng, Y. H., Sun, J. “Applying thermo-physics for wind turbine drivetrain fault diagnosis using SCADA data”, *IET Renewable Power Generation*, 2016, vol. 10, no. 5, pp. 661–668.

[3] SUN, P., LI, J., KOU, X. S. “Wind turbine status parameter anomaly detection based on prediction models and fuzzy theory”, *Automation of Electric Power Systems*, 2017, vol. 37, no. 08, pp. 90–98.

[4] Martin, B., David, B., Marie, C., Jean, F. “On-line monitoring of mechanical faults in variable-speed induction motor drives using the Wigner distribution”, IEEE Transactions on Industrial Electronics, 2008, vol. 55, pp. 522-533.

[5] GUO, H. D. “On-line Identification for Wind Turbine Pitch System Degradation Based on Data Mining Technology”, *Proceedings of the CSEE*, 2016, vol. 36, no. 09, pp. 2389–2397.

[6] Zhang, Q., Yang, L. T,. Chen, Z, “Deep Computation Model for Unsupervised Feature Learning on Big Data”, *IEEE Transactions on Services Computing,* 2016, vol. 9, pp. 161-171, ISSN:1939-1374.

[7] ZHAO, H. SH., YAN, X. H., WANG, G. L. “Fault diagnosis of wind turbine generator based on deep autoencoder network and XGBoost”, *Automation of Electric Power Systems,* 2019, vol. 43, no. 1, pp. 81-90.

[8] MITRA, S., KOLEY, C. “An automated SCADA based system for identification of induction motor bearing fault used in process control operation”, in: 2nd International Conference on Control, Dubai, United Arab Emirates, Instrumentation Energy & Communication (CIEC), IEEE 2016：294-298.

[9] Cong, W., Hu, L. L., Sun, S. J. “Meteorological Disaster Prevention Method for Power Grid Based on Improved Stacked Denoising Auto-encoder Network”, *Automation of Electric Power Systems,* 2019, vol. 43, no. 02, pp. 42-50.

[10] Du, B. L., He. Y. G., He, Y. Z., Zhang, C. L., “Progress and trends in fault diagnosis for renewable and sustainable energy system based on infrared thermography: A review”, *Infrared Physics & Technology*, 2020, vol. 109.

[11] Zhang, X. N., Xiang, Z., Xia, X. R., Li, L. F., “Optimization of Stacking Auto-Encoder with Applications in Bearing Fault Diagnosis”, *Journal of Xi'an Jiaotong University,* 2018, vol. 52, no. 10, pp. 49-56+87.

[12] Rumelhart, D. E., Hinton, G. E., Williams, R. J., “Learning representation by back-propagating errors”, *Nature,* 1988, vol. 323, no. 6088, pp. 399-421.

[13] Hinton, G. E., Salakhutdinov, R. R., “Reducing the dimensionality of data with neural networks”, *Science,* 2006, vol. 313, no. 5786, pp. 504-507.

[14] Zhao, H. S., Liu, H. H., Liu, H. Y., Lin, Y. K., “Condition Monitoring and Fault Diagnosis of Wind Turbine Generator Based on Stacked Autoencoder Network”, *Automation of Electric Power Systems*, 2018, vol. 42, no. 11, pp. 102-108.

[15] Wang, L., Zhang, Z. J., Xu, J.,“Wind turbine blade breakage monitoring with deep autoencoders”, *IEEE Transactions on Smart Grid,* 2018, vol. 9, no. 4, pp. 2824-2833.

[16] Zhao, Z. Y., *Python Machine Learning Algorithm,* 1st ed., vol. 1, Beijing, China: Electronic Industry Press, 2017.

[17] Yang, Y., Du, F., Deep Learning in Action[M]. Beijing, Tsinghua University Press, 2018.

[18] Kingma, D. P., Ba, J. "Adam: a method for stochastic optimization", in: Proceedings of the 3rd International Conference on Learning Representations, Computer Science, 2015, pp. 1-15.

[19] Ioffe, S., Szegedy, C., "Batch normalization: accelerating deep network training by reducing internal covariate shift", 32nd International Conference on Machine Learning, Cambridge, MA, USA, IMLS, 2015, pp. 448-456.

[20] Guo, T. T.,"On-line abnormal state identification of pitch system based on transitional mode for Wind Turbine", Beijing, China: *Beijing Jiaotong University*, 2019.

Acknowledgements

This work was financially supported by the Natural Science Foundation of China, No.51867013 and 51767014.

Biography

Sihua Wang was born in Jiangsu Province (China), on May 21,1968.
He graduated the East China Jiaotong University, Electrical engineering and automation in Jiangxi(China), in 1995.
He is a Professor at Lanzhou Jiaotong University ,in Lanzhou(China).
His research interests concern: Reliability of Substation Integrated Automation and Direction of HV Insulation.
Correspondence address: ws_h@163.com

Tian Wang was born in Gansu Province (China), on Sept 16,1991.
She is studying for a master's degree in the Lanzhou Jiaotong University, Electrical Engineering in Lanzhou(China).
Her research interest concern: mainly based on deep learning of wind turbine fault diagnosis.
Correspondence address: 1047937197@qq.com

Lijun Zhou is studying for a master's degree in the Lanzhou Jiaotong University, Electrical Engineering in Lanzhou(China).
Her research interest is mainly in flexible DC grid protection and control.
Correspondence address: 101264731@qq.com

Tianyu Chen is studying for a master's degree in the Lanzhou Jiaotong University, Electrical Engineering in Lanzhou(China).
His research interest concern: mainly wireless monitoring of catenary voltage of electrified railway.
Correspondence address: 787693391@qq.com

Yu Wang is studying for a master's degree in the Lanzhou Jiaotong University, Electrical Engineering in Lanzhou(China).
His research interest is mainly in the protection and control of flexible DC power grid.
Correspondence address: 1192958264@qq.com