Fault Diagnosis of Wind Turbine Generator System Based on PMI-LSSVM

Wei Zhang1,*, Zhizhi Zhang1, Qi Yao2, Xiao Zhang3, Di Liu1, Zesan Liu1, Hongmin Meng1 and Weiming Qin1

1 State Grid Information and Telecommunication Group Co., Ltd, Beijing 102211, China
2 Energy and Electricity Research Center, Jinan University, Zhuhai 519070, Guangdong, China
3 Communication University of China, Nanjing, Nanjing 211172, Jiangsu, China

*Corresponding author email: ncepuzw@163.com

Abstract. Considering the multi-source wind power information such as wind speed, rotation speed, spindle horizontal and vertical vibration, a fault diagnosis method of wind turbine generator system based on partial mutual information (PMI) and least squares support vector machine (LSSVM) was proposed. A large amount of data containing fault status, such as blade fault, converter fault, generator fault, pitch bearing fault and yaw system fault, was analyzed. The PMI method was used to screen the characteristic parameters of the operation state of the wind turbine to identify the fault of the unit. The characteristic parameters of the wind turbine in various states were trained by LSSVM method to establish the mapping relationship between the parameter vectors of different characteristics and the fault types, so as to achieve the purpose of fault diagnosis. Besides, the different fault history data of wind turbine was used to test the fault model performance. The results compared with artificial neural network (ANN) method showed that the proposed method had good fault recognition ability and fast operation speed, which was suitable for fault diagnosis of multibrid technology wind turbine generator system, and can meet the requirements of online fault diagnosis.

Keywords: Fault diagnosis; Wind turbine generator system; Partial mutual information; Least squares support vector machine.

1. Introduction

The condition monitoring and fault diagnosis of wind turbines are the key to ensure the long-term stable operation and safe generation of wind turbines. In order to reduce the fault rate of wind turbine and the operation and maintenance cost of wind turbine, it is necessary to carry out the condition monitoring and fault diagnosis research of wind turbine. Timely mastering the operation status of wind turbine and early detecting the potential fault symptoms can lower the fault rate, reduce the operation and maintenance cost, and ensure the safe and efficient operation of large wind turbine.

With the rapid development of China’s wind power industry, it is faced with the situation of frequent faults of wind turbines. However, in the case of wind turbine fault, the existing solutions still remain after the occurrence of failures, relying on artificial experience to diagnose and analyze the operation status information of equipment failures. Nevertheless, for such a complex system as wind turbine, the causes of some faults are strongly coupled. How to quickly screen the causes of the system fault, comprehensive fault diagnosis and location is particularly important.
There are many fault diagnosis methods for wind turbine, including traditional diagnosis method, mathematical diagnosis method and intelligent fault diagnosis method. First, the most of the traditional diagnosis methods are based on the data analysis of condition monitoring technology, such as the comparison of state similarity, but the ability of traditional methods to identify fault information isn't comprehensive, and they are usually combined with other methods to diagnose faults. Second, the mathematical diagnosis methods mainly include the pattern recognition, the time series model based on probability and statistics, diagnosis based on distance criterion, fuzzy diagnosis, grey system diagnosis, fault tree analysis, wavelet analysis, chaos analysis and fractal geometry. But the above existing mathematical diagnosis methods have some limitations, such as fault tree analysis. In practical application of fault tree analysis, because it is difficult to determine the classification of some unknown states, the fault diagnosis results can not be accurately obtained simply by using fault tree analysis [1]. Third, the intelligent diagnosis methods mainly include fuzzy logic, expert system, neural network, genetic algorithm, support vector machine and so on. Among them, expert system and artificial neural network (ANN) are widely used in fault diagnosis of rotating machinery [2,3], but there are still the following main defects in practical application. For example, expert system is difficult to establish knowledge base and verify its completeness, its fault tolerance is poor, and there is no effective methods to identify error information yet [4]. Besides, the maintenance of knowledge base of large-scale expert system is very difficult, and there are problems of combination explosion and slow reasoning speed in the complex fault diagnosis task. Correspondingly, artificial neural network has the characteristics of strong fault tolerance and learning ability [5]. It can combine the running state of the diagnosis object to construct a pattern recognition classifier for state detection, and analyze the observed feature data information for fault diagnosis [6]. However, the multi-layer forward neural network needs back propagation neural network algorithm to calculate the back error propagation. Thus, the training time of back propagation neural network model is long and it is easy to produce local optimum, and the accuracy of classification is affected [7]. In comparison, the support vector machine (SVM) method is a small sample learning method, which realizes efficient transductive reasoning from training samples to prediction samples, and simplifies the common problems of classification and regression [8]. However, in terms of a large number of multi-class data training, there are also problems such as long training time and difficulty in solving. To solve these problems, the least squares support vector machine (LSSVM) method optimizes the model parameters, which not only has good fitting performance and it improves the generalization ability of the model [9]. In this paper, based on a large number of historical operation data of wind turbines, the partial mutual information (PMI) method was used to select the characteristic state variables that affect the state of wind turbines. Then, different types of fault data were taken as training samples and trained in the LSSVM model to get the mapping relationship between different state vectors and fault types, so as to establish the state fault diagnosis model of wind turbine. The performance of the diagnosis model was proved effective through the comparison with ANN model and the verification of historical fault state data.

2. Method Introduction

In this part, PMI and LSSVM methods were introduced respectively. The selection of characteristic state variables that affect the state of wind turbine and the algorithm steps of fault data type modeling were introduced in detail.

2.1. Fault Feature Screening Based on PMI

Based on the entropy of information theory, the mutual information method is used to reflect the linear and nonlinear correlation between variables. It can quantitatively reflect the correlation degree between variables, and has good performance in describing linear and nonlinear variables. On the basis of mutual information, PMI uses conditional expectation to eliminate the correlation between variables, and then calculates the mutual information, which eliminates the relationship between variables and effectively improves the accuracy of variable selection. The entropy has been used as a measure of information [10]. According to Reference [11], the entropy, joint entropy, conditional entropy can be calculated as Formula (1) - (3):
The information shared by $X$ and $Y$ is the partial mutual information $I(X,Y)$ [11], which can be calculated as Formula (4).

$$I(X,Y) = H(X) - H(X/Y)$$  \hspace{1cm} (4)$$

The partial mutual information can be calculated as Formula (5) - (7):

$$I(X,Y) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{f(x_i, y_i)}{f(x_i) f(y_i)}$$  \hspace{1cm} (5)$$

$$f(x) = \frac{1}{n \sqrt{2 \pi h^d}} \exp \left( -\frac{1}{2h^2} \sum_{i=1}^{n} \|x - x_i\|^2 \right)$$  \hspace{1cm} (6)$$

$$h = \left( \frac{1}{d+2} \right)^{\frac{1}{d+4}} n^{-\frac{1}{d+4}}$$  \hspace{1cm} (7)$$

The relationship between entropy, joint entropy, condition entropy and mutual information is shown in Figure 1.

**Figure 1.** Relationship between entropy, joint entropy, condition entropy and mutual information.

For multi-input systems, the information of partial inputs $X$ contained in output $Y$ and their inputs $Z$ should be eliminated using the conditional expectation, and the correlation degree among variables should be measured by PMI [11]. After processing, the output and input are expressed as $u$ and $v$ respectively. The relevant formula is shown in Formula (8) - (10):

$$m_y(x) = E[y|X = x] = \left[ \sum_{i=1}^{n} y_i f(x) \right] / \left[ \sum_{i=1}^{n} f(x) \right]$$  \hspace{1cm} (8)$$

$$u = Y - m_y(X)$$  \hspace{1cm} (9)$$

$$v = Z - m_z(X)$$  \hspace{1cm} (10)$$
The partial mutual information between $Y$ and $Z$ is expressed as Formula (11):

$$\text{PMI}(Z, Y) = I(v, u)$$  \hspace{2cm} (11)$$

The procedure of variable selection method based on PMI are the same as those in Reference [11]. With the selection of independent variables, the value of Akaike information criterion (AIC) constantly decreases. When the AIC reaches the minimum value, the optimal set of independent variables is completely extracted [11].

2.2. LSSVM Data-driven Fault Diagnosis Model

Based on the theory of structural risk minimization, SVM maps the nonlinear samples in low dimensional space into linear samples in high dimensional space through kernel function, so as to reduce the complexity of solution. The basic principle is shown in Figure 2.

![Figure 2. Basic schematic of support vector machine.](image)

LSSVM is an improvement of SVM. LSSVM is a method of applying kernel to Ridge region. It uses least square error to fit all samples. The fitting is in the kernel transformed high-dimensional space. The inequality constraints are replaced with equation constraints, and the sum of squared error loss function is taken as the experience loss of training set. Thus, solving quadratic programming problem is transformed into solving linear equations, which improves the speed and convergence accuracy of solving the problem.

$$\{ (x_i, f_i) \}_{i=1}^{n}$$ is a given n-dimensional data set, where the input $x_i$ is the independent vector including the control parameters of operation state, the output $f_i$ is the operation cost of unit, the input sample $x_i \in \mathbb{R}^p$, the output sample $f_i \in \mathbb{R}$. The LSSVM regression model can be expressed with the following Formula (12):

$$f_i = w^T \phi(x_i) + b$$  \hspace{2cm} (12)$$

where $w$ is the weight vector; $\phi(\cdot)$ is a nonlinear mapping; $b$ is the bias.

As in reference 11, the LSSVM is finally obtained as Formula (13):

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i) + b$$  \hspace{2cm} (13)$$

where $\alpha = [\alpha_1, \ldots, \alpha_n]^T$ is the Lagrange multiplier vector; $f = [f_1, \ldots, f_n]^T$. Besides, the widely-used radial basis function (RBF) is adopted as the kernel function of LSSVM, which can be expressed as Formula (14):

$$K(x, x_i) = \exp \left( -\frac{\|x - x_i\|^2}{\sigma^2} \right)$$  \hspace{2cm} (14)$$

where $\sigma$ is the kernel parameter.

The LSSVM method can find the best function matching of data by minimizing the sum of squares of errors, and can be used to obtain the unknown data easily and make the sum of squares of the errors between the obtained data and the actual data minimum.
3. Simulation Verification
Wind turbine is a complex electromechanical integrated system. The main components of wind turbine fault include blade, gearbox, generator, pitch system, frequency converter and so on. Among them, the gearbox fault, the electrical system fault and the generator fault are the three main faults. There are many mapping relationships between past symptoms and faults. According to the fault information in the actual operation data of MySE 3000 multibrid technology wind turbine generator system, the following case study is introduced through feature variable selection, fault diagnosis model establishment and model simulation verification.

3.1. Data Acquisition
The research object of simulation data is MySE multibrid technology wind turbine generator system, and its model is MySE3000. The historical operation data of the multibrid technology wind turbine generator system from December 2018 to June 2019 was selected, and 10 minutes was taken as the sampling time to sample the operation data. According to the actual operation fault records from December 2018 to June 2019, the operation data within the fault time was retrieved. The following Table 1 was shown part of these fault record.

Table 1. The fault type and the fault duration of wind turbine in historical operation data.

| Fault Type number | Fault details | Duration               |
|-------------------|---------------|------------------------|
| 1                 | Filter element 1 of gearbox main filter pump fault | 2019/6/11 10:06 - 2019/6/11 11:03 |
| 2                 | Filter element 2 of gearbox main filter pump fault | 2019/6/10 23:17 - 2019/6/10 23:43 |
| 3                 | Safety relay triggered | 2019/6/8 10:02 - 2019/6/8 12:29 |
| 4                 | Pitch blade out of synchronization | 2019/5/26 10:22 - 2019/5/26 10:35 |
| 5                 | Protection trip signal of gearbox fan motor in engine | 2019/5/9 1:44 - 2019/5/9 10:54 |
| 6                 | Oil leakage of gearbox | 2019/4/12 14:18 - 2019/4/13 9:06 |
| 7                 | Starting time of converter exceeded limit | 2019/2/20 13:52 - 2019/2/20 14:11 |
| 8                 | Gearbox accumulator pressure below limit | 2019/2/18 22:04 - 2019/2/19 7:27 |
| 9                 | Generator speed 1 below set value & blade below set value | 2019/1/29 20:21 - 2019/1/29 20:48 |
| 10                | Oil pressure at the gearbox port below limit | 2019/1/22 10:59 - 2019/1/22 11:18 |

Taking 10 minutes as sampling time, the statistical fault record data was sampled. 480 groups of characteristic variable training data and 150 groups of characteristic variable test data were obtained from the historical data.

3.2. PMI Variable Selection
Firstly, according to the procedure of variable selection method based on PMI in 2.1, PMI method was used to screen the characteristic variables of wind turbine. As shown in the Figure 3, a group of characteristic variables with minimum AIC value was obtained as the input variables of fault diagnosis model.
It was shown in Figure 3 that when the number of input variables was 25, the AIC value initially reached the minimum. In addition to the relationship between AIC value and the number of variables, according to the PMI algorithm, the input variables of the fault diagnosis model obtained in the screening process include: the wind speed, the impeller speed, the generator speed, the converter generator speed, the converter torque, the engine room vibration x, the engine room vibration y, the gearbox oil pump high speed, the gearbox oil pump low speed, the axle box temperature, the gearbox main bearing temperature, the gearbox front bearing temperature, the gearbox rear bearing temperature, the wind direction, the pitch angle setting value, the motor temperature, the generator winding temperature, the generator front bearing temperature, the generator rear bearing temperature, the gearbox oil level, the energy storage tank pressure, the yaw pressure, the blade angle 1, the blade angle 2, the blade angle 3.

3.3. PMI-LSSVM Model Training

The wind turbine state vector composed of above 25 variables by PMI method was used as the model input, and the fault type number in Table 1 was used as the model output. Thus, based on PMI feature screening, PMI-LSSVM model was built to judge the types of different fault states. Based on rough and fine cross validation optimization, the optimal parameters pair \((\gamma, \sigma)\) of LSSVM model was determined as \((7408.2538, 2186.4722)\). The training result including the actual fault type and diagnosis fault type was shown in Figure 4. Besides, the red dotted line in the enlarged area represented the diagnostic error result.

In Figure 4, it can be seen that the training curve of diagnosis fault type was in good agreement with the curve of actual fault type. Although there were some small diagnostic error between the fault type 1 and the fault type 2, the overall effect of fault diagnosis was effective. The above two confused fault type corresponded to the fault of the filter element 1 of gearbox main filter pump and the fault of the filter element 2 of gearbox main filter pump. The gearbox fault is a very common problem in wind
power companies. The main reason of the gearbox main filter pump fault is that, the filter element is blocked by too many impurities in the lubricating oil. In general, each filter element may be blocked to varying degrees, so it was easy to make above diagnostic errors.

Further, the model training can be evaluated by error statistical analysis. The mean square error (MSE) and mean relative error (MRE) of model training were 0.1708 and 0.0292, respectively. These index values confirmed the good training results of the PMI-LSSVM model.

3.4. PMI-LSSVM Model Testing

150 groups of fault data were used as test data to verify the built model. The test result was shown in the Figure 5.

![Figure 5. The test results of PMI-LSSVM model.](image)

It can be seen from the Figure 5 that the fault diagnosis model of wind turbine generator system have good fault recognition ability and robustness. However, there were still a small part of diagnostic error between the fault of the filter element 1 of gearbox main filter pump and the fault of the filter element 2 of gearbox main filter pump. But on the whole, PMI-LSSVM model can diagnose all kinds of faults well, and the error between diagnosis fault type and actual fault type was kept in a small error range.

3.5. Model Comparison and Validation

In order to compare the effect of fault diagnosis, ANN algorithm was used to diagnose the test case, and 25 variables screened by PMI method was also used as the model input, and Figure 6 was obtained.

![Figure 6. The test results of PMI-ANN model.](image)

By comparing with the diagnosis results of PMI-LSSVM, the test results of PMI-ANN model also had the diagnostic errors of fault type 1 and fault type 2, and even had a higher diagnosis error rate. The comparison results of PMI-LSSVM and PMI-ANN in diagnosis error rate and diagnosis time were shown in Table 2.
Table 2. The comparison results of PMI-LSSVM and PMI-ANN in diagnosis error rate and diagnosis time.

| Model index | PMI-LSSVM | PMI-ANN |
|-------------|-----------|---------|
| MSE         | 0.2582    | 0.4472  |
| MRE         | 0.0667    | 0.2000  |
| Diagnosis time (s) | 0.0668 | 0.6963 |

Comparing the two models, the following results could be obtained: Firstly, the test errors of the two models both were smaller, but the PMI-LSSVM method had higher test accuracy than the PMI-ANN model; the PMI-LSSVM method had shorter operation time, lower model complexity, and better generalization ability. In addition, compared with the PMI-ANN model, the PMI-LSSVM model didn’t have the disadvantage of unstable performance caused by the randomness of the weights during neural network training. Therefore, in terms of comprehensive indicators, the PMI-LSSVM model had more advantages. Future research work should focus on how to reduce the diagnostic errors between the failure of gearbox main filter pump filter element 1 and gearbox main filter pump filter element 2.

4. Conclusion

(1) The PMI method eliminated the relationship between variables, effectively improved the accuracy and screening efficiency of feature screening, and effectively provided support for big data learning and modeling technology.

(2) Considering the multi-source wind power information, the PMI-LSSVM model can quickly and effectively identify the fault types. Compared with PMI-ANN model, PMI-LSSVM model had advantages in model diagnosis accuracy and operation time, and the diagnose results could provide effective guidance for unit maintenance.

(3) Wind turbine is a complex electromechanical integrated system. The PMI-LSSVM fault monitoring model driven by historical wind turbine state data can effectively diagnose and classify the faults on-line, which is conducive to lowering the unit failure rate, reducing the maintenance time and improving the economic benefits of the wind farm.

Acknowledgments

This project is supported by the National Key Research and Development Program of China (Grant number: 2020YFB2104500). The authors would like to thank the company and colleagues for the support. In addition, the authors expect positive criticism and suggestions.

References

[1] B Qin, G D Sun, L Y Zhang, et al. Fault features extraction and identification based rolling bearing fault diagnosis. Journal of Physics: Conference Series, 2017, 842: 12-55.
[2] Tairan Han. Fault diagnosis strategy of coal mine high voltage cable based on neural network expert system. Coal technology, 2019 (11).
[3] Juzheng Liang. Research on automatic diagnosis of mechanical fault. Electronic test, 2018, 0(3): 97-98.
[4] Jun Zeng, Yanfeng Chen, Ping Yang, et al. Review of fault diagnosis methods of large-Scale wind turbines, Power System Technology, 2018, 42(3).
[5] Sheng Yu, Hongbo Zou, Fan Yu, et al. Summary of power grid fault diagnosis based on combination of multiple intelligent methods. Journal of Electric Power, 2019, 148(1): 30-36.
[6] Longlong Chen, Bo Wang, Ling Yuan. A neural network fault diagnosis method for power transformer. Journal of Nanjing University of Information Engineering: Natural Science Edition, 2018, 54(2): 199-202.
[7] Pu Yuan, Jianlin Mao, Fenghong Xiang, et al. Improved BP neural network fault diagnosis based on genetic optimization. Journal of power system and automation, 2017, 29(1): 118-122.
[8] Zengsheng Kuang, Zhe Zhang, Xiuli Shao, et al. Fault diagnosis method based on support vector machine and cosine similarity. Data mining, 2020, 10(2): 136-142.
[9] Guo Wei, Lei Li. Research on fault diagnosis of reciprocating compressor based on least squares support vector machine. Microcomputer information, 2018, 0(21): 95-96.

[10] Jizhen Liu, Tianmu Qin, Tingting Yang, et al. Variable selection method based on partial mutual information and its application in SCR system modeling of thermal power plant. Chinese Journal of electrical engineering, 2016, 036(009): 2438-2443.

[11] Wei Zhang, Jizhen Liu, Mingming Gao, et al. A novel operation cost optimization system for mix-burning coal slime circulating fluidized bed boiler unit. Applied Thermal Engineering, 2019, 148: 620-631.

[12] Bernard W. Silverman, Density estimation for statistics and data analysis, Chapman & Hall, London, 1986.