Gearbox Fault Prediction of Wind Turbine Based on Improved NEST Model

Di Shuai

Datang Northeast Electric Power Test & Research Institute Co., Ltd, Thermal Automation Department, 130000 Changchun, China

Abstract. This paper studies a fault prediction method for wind turbine gearbox. It uses grey relation analysis to get modeling variables, and makes sample data getting good integrity and redundancy by similarity analysis. Thus it gets the reduced process memory matrix, and trains the improved nonlinear state estimation (NEST) model. When the gearbox fails, the model residual will exceed the threshold value, and the model will give an early warning. Combined with the actual operation data of a wind turbine, the effectiveness and accuracy of the improved model are verified.

1 Introduction

The gearbox as the key component of wind turbines can transfer the moment generated by wind energy to the generator and convert into electrical energy. It can also change the low speed of the main shaft to the high speed required by the generator. The maintenance process of gearbox is complex. The maintenance process needs a lot of equipment, and the weather for maintenance is demanding. Comparing with the electrical system and main control system with the high failure rate, the gearbox has the longest failure time and the highest economic loss. Therefore, this paper studies the fault prediction of gearbox. Literature [1] makes statistics of various faults of wind turbine, and the result shows that gearbox is the component which causes longest downtime and most serious failure. Literature [2-4] summarize the common condition monitoring methods of wind turbine. Literature [5] uses the nonlinear state estimation model to predict gearbox temperature. But it selects observation variables relying on experience, and it lacks theoretical basis. At the same time, the process memory matrix is not optimized for sample, and which affects the timeliness of model.

The bearing of gearbox is the carrier of main shaft, and which directly reflects the meshing state of main shaft and gear. Because of the particularity of bearing work and the complexity of load, bearing becomes a component with high failure rate. This paper uses NEST model to monitor the gearbox bearing temperature. It is aimed at the deficiency of selecting observation vector based on experience, and it uses the grey correlation analysis method to provide theoretical basis. In order to ensure the integrity and small redundancy of data, this paper uses the similarity analysis method to construct the process memory matrix which can realize sample optimization. When the working state of gearbox is abnormal, the residual of the prediction model and mean square root of residual will increase. This paper uses Statistical analysis

for residuals by moving window residuals method and compares with the setting threshold value, which can realize fault early warning. This paper uses SCADA data of a wind turbine to simulate and analysis, the results show that the improved NEST model can effectively predict the bearing temperature and has good timeliness.

2 Data preprocess

2.1 Selection of modeling variables

This paper is based on SCADA data of a wind turbine. The rated power of fan is 2000 kW. The rated frequency is 50Hz. The rated wind speed is 11.5 m/s. The wind turbine SCADA data records average value of each parameter every 10 minutes. This paper selects 5050 sets of effective data to participate in modeling. In order to establish the prediction model of gearbox bearing temperature, this paper selects the variables based on experience which involved in modeling, including ambient temperature, gearbox oil temperature, average wind speed and active power. Next, this paper analyses feasibility of the four parameters involved in modeling by theoretical.

Gearbox oil temperature: the oil and bearing are both in the gearbox, and the gearbox has heater and cooler. When the oil temperature is high, the cooler will work. When the temperature is low, the heater will work. The bearing temperature is affected by the heater and cooler. Ambient temperature: there's a clear temperature difference between day and night in local, and the bearing temperature is greatly affected by ambient temperature.

Active power: the higher the active power, the higher the unit load and the higher the gearbox bearing temperature.

Average wind speed: the higher the wind speed, the faster the gearbox rotates and the higher the bearing temperature.

2.2 Similarity analysis

Constructing observation vector of gearbox bearing temperature X_i:

$$x_{obs}(i) = [x_1 \ x_2 \ x_3 \ x_4]^T \qquad (1)$$

In formula (1), X_1 is ambient temperature; X_2 is the gearbox oil temperature; X_3 is the average wind speed; X_4 is the active power.

Constructing process memory matrix U:

$$U = [X(1) \quad X(2) \quad \dots \quad X(m)]$$

$$= \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(m) \\ x_2(1) & x_2(2) & \cdots & x_2(m) \\ \vdots & \vdots & \vdots & \vdots \\ x_n(1) & x_n(2) & \cdots & x_n(m) \end{bmatrix}$$

m is the number of observation vectors. m=5050. n is the number of modeling variables. n=4.

The SCADA system of wind turbine contains a large number of similar data, which leads to the increase of calculation amount and calculation time. Similarity analysis can effectively solve this problem. Firstly, the data of gearbox in normal working state is selected as observation vector, then it calculates the similarity between any two groups of observation vectors. Finally, the similarity matrix R is obtained.

Figure 1 is the schematic diagram. The closer the similarity value is to 1, the more the two observation vectors contain the same information. This paper eliminates a group of samples to optimize and simplifies the sample, and ensures the integrity of workspace information.
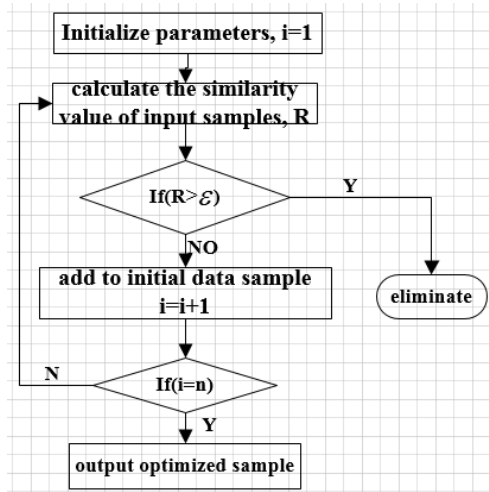

Figure 1. Similarity principle.

The similarity function is:

$$R_{ij} = \exp(-\|x_i - x_j\| / \delta) \qquad (2)$$

In formula (2), $x_i \in R^m$, i=1,2,$\cdots$,n; m is the dimension of data; δ is the normalized parameter; R_{ij} is the similarity value between i group and j group of data.

Calculating similarity normalization parameters δ:

$$\delta = \sum_{i=1}^{m} (\max D_i - \min D_i) \qquad (3)$$

In formula (3), D_i is the data set of parameter i.

The simplification degree of reduced process memory matrix depends on the selection of similarity function threshold. There are 5050 groups of effective historical data in this wind turbine SCADA system. The relationship between number of remaining samples and different threshold is shown in Figure 2.
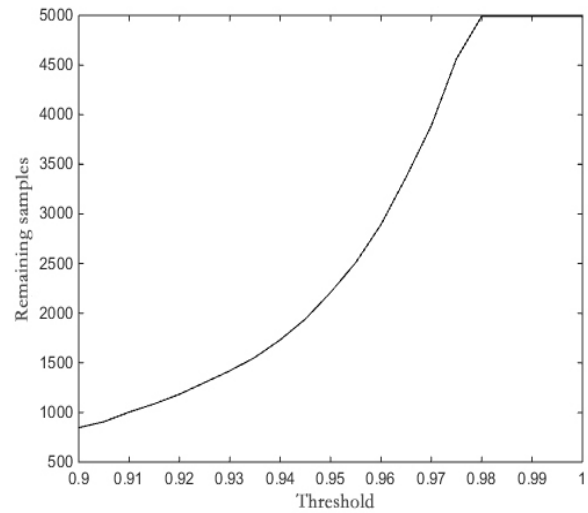

Figure 2. Relationship between number of remaining samples and threshold

Table 1. Similarity analysis results of historical data.

τ	0.91	0.92	0.94	0.96	0.97	0.975	0.99
No. of groups	990	1190	1756	2871	3870	4556	5050

For the requirements of high integrity and low redundancy, when the curve slope is the largest, it means that the increment of redundancy information in this interval is the largest. In order to get best effect, the threshold is 0.97 in this paper. 5050 training sample sets are optimized by similarity function, and 3870 optimized sample sets are obtained.

$$U' = [X(1) \quad X(2) \quad \dots \quad X(s)]$$

$$= \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(s) \\ x_2(1) & x_2(2) & \cdots & x_2(s) \\ \vdots & \vdots & \vdots & \vdots \\ x_n(1) & x_n(2) & \cdots & x_n(s) \end{bmatrix}$$

S is the number of observation vectors in reduced process matrix, s=3870.

2.3 Grey correlation analysis

In order to reduce the information dimension and improve the prediction accuracy, this paper selects the operation parameters which are sensitive to bearing fault, and removes the parameters which are less sensitive. Grey relational analysis (GRA) can achieve this function

better. It can calculate the consistency of changing trend of factors to achieve variable selection.

32 relevant parameters can be reduced to 8 parameters which related to gearbox bearing temperature, including environment temperature, gearbox oil temperature, drive chain swing amplitude, average speed of generator, cabin temperature, average wind speed, active power, average speed of control generator and power generation. Through formula correlation, this paper obtains the Grey correlation degree value of selected variables to the bearing temperature, and the results are shown in Table 2.

Table 2. Grey correlation value of modeling variable.

R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
0.1 10	0.0 70	0.0 78	0.0 62	0.1 03	0.1 04	0.1 20	0.0 53

Among them, R_1 is the grey correlation value of average wind speed; R_2 is the grey correlation value of average speed of control generator; R_3 is grey correlation value of average speed of generator; R_4 is the grey correlation value of power generation; R_5 is the grey correlation value of active power; R_6 is the grey correlation value of ambient temperature; R_7 is the grey correlation value of gearbox oil temperature; R_8 is the grey correlation value of engine room temperature.

According to table 2, it can be seen that the top four items are gearbox oil temperature, average wind speed, ambient temperature and active power. It is completely consistent with the observation vector selected by experience in Chapter 1.1. It provides a theoretical basis for the selection of modeling variables and enhances the credibility of the model.

3 Establishment of NEST model and residual analysis

3.1 Nonlinear state estimation modeling

It takes the new observation vector of gearbox working state X_{obs} as the input of NEST model and outputs the corresponding prediction vector X_{ets}.

Firstly, the NEST model will generate an s-dimension weight vector corresponded input:

$$W = \begin{bmatrix} w_1 & w_2 & \cdots & w_s \end{bmatrix}^T \quad (4)$$

$$X_{est} = RRW = w_1 \cdot X(1) + \cdots + w_s \cdot X(s) \quad (5)$$

In order to find the weight vector, the residual between model input and estimated value ε should be kept to a minimum:

$$\min \varepsilon = \min(|X_{obs} - X_{est}|) \quad (6)$$

In order to facilitate the computing, the sum of squares of the residuals ε should be calculated:

$$S(\varepsilon) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon = \sum_{i=1}^{n} (X_{obs}(i) - \sum_{j=1}^{m} w_j D_{ij})^2 \quad (7)$$

Finding the partial derivative of $S(w)$ to w_k:

$$\frac{\partial S(w)}{\partial w_k} = -2 \sum_{i=1}^{n} (X_{obs}(i) - \sum_{j=1}^{m} w_j D_{ij}) D_{ik} = 0 \quad (8)$$

Simplifying formula (8):

$$\sum_{i=1}^{n} X_{obs}(i) D_{ik} = \sum_{i=1}^{n} \sum_{j=1}^{m} w_j D_{ij} D_{ik} = \sum_{j=1}^{m} (\sum_{i=1}^{n} D_{ij} D_{ik}) w_j,$$
$$k = 1, 2, \cdots, m \quad (9)$$

The matrix form of formula (9):

$$X_{obs} \cdot D^T = D \cdot D^T \cdot W \quad (10)$$

Transforming formula (10):

$$W = (D \otimes D^T)^{-1} \cdot (X_{obs} \otimes D^T) \quad (11)$$

In formula (11), $X \otimes Y = \sqrt{\sum_{i=1}^{n} (x_i - x_j)^2}$ is the similarity calculation based on Euclidean Distance, and the estimated output is:

$$X_{est} = D \cdot W = D \cdot (D^T \cdot D)^{-1} \cdot (D^T \cdot X_{obs}) \quad (12)$$

3.2 Statistical method of residual

Because of the random change of wind speed and the time-varying influence of fan operating conditions, various uncertain factors will affect the residual change. It leads to some isolated abnormal points and false alarm. This paper uses the analysis method of residual mean value of sliding window, which can effectively weaken the influence of isolated points and continuously reflect the trend of residual change. It can avoid false alarm and accurately give early warning signals.

Setting the width of sliding window to N. Processing the mean value of N consecutive residuals:

$$\overline{x_i} = \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i \quad (14)$$

During the normal operation of the equipment, the residual mean value sequence is obtained by sliding window calculation, and the maximum absolute value obtained by statistics is E_x.

Setting the residual threshold E_v of NEST model:

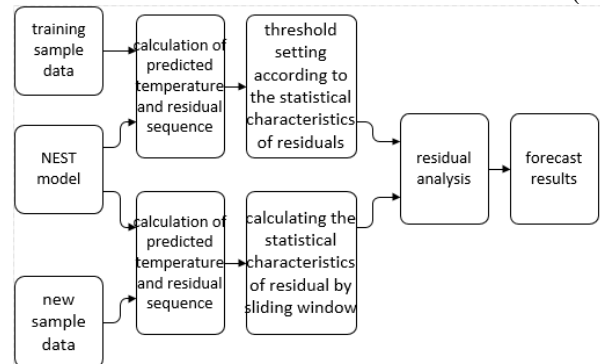$$E_v = \pm k * E_x \quad (15)$$

Figure 3. Prediction process based on NEST.

In formula (15), the setting of K is generally determined according to the experience of staff. The flow chart of gearbox bearing temperature monitoring model based on NEST is shown in Figure 3.

4 Fault prediction model

4.1 Verifying the model

This paper sets the threshold τ to 0.97. The reduced process matrix is constructed by the remaining 3870 groups of data, which trains NEST model 1. NEST model 2 is trained by all 5050 groups of data as a comparison. 800 groups of data were selected as the test set , and it used to test the validity of the model.

The predicted residual of NEST model recording the bearing temperature of gearbox:

$$\varepsilon_{GT} = x_b - \hat{x}_b \qquad (16)$$

In formula (16), x_b is the input temperature variable. $\hat{x}_b$ is the predicted temperature variable. The prediction results of model 1 and model 2 are shown in Figure 4 and figure 5.
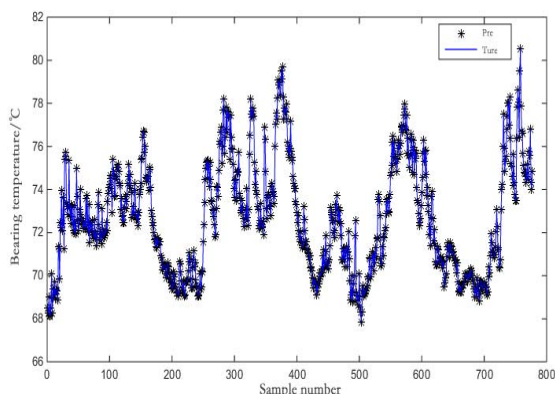

Figure 4. Prediction results of model 1.

The average relative uses MRE to represent. It can be seen from Figure 4 that the MRE of model 1 is 1.42%, and the operation time of the model is 4.9 seconds. It can be seen from Figure 5 that the MRE of model 2 is 1.28%, and the operation time of the model is 9.6 seconds.

The comparison shows that the MRE of the two models is very low. In other words, the prediction accuracy of two models is very high. Two models can track the change trend of gearbox bearing temperature well. The prediction accuracy of two models is the same order of magnitude, and the difference isn't big. But in terms of operation time, the time consumption of model 1 is only half of model 2.
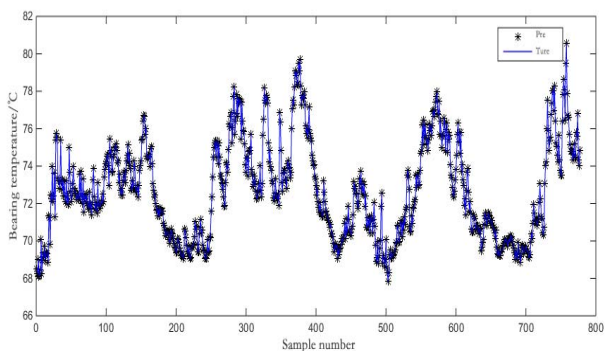

Figure 5. Prediction results of Model 2.

Therefore, the trained NEST model constructed by reduced process matrix and similarity analysis not only has high accuracy, but also shortens the operation time. At the same time, it reduces the operation amount and effectively improves the timeliness of the model.
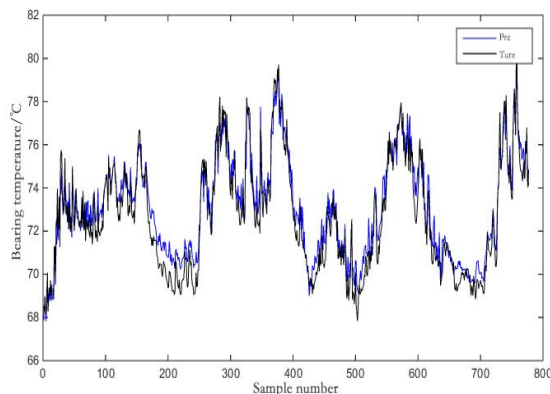

Figure 6. Prediction results of model 3.

Comparing with different models, the least square support vector machine (LSSVM) is used to train the model 3, and the parameters are optimized by the bird swarm optimization (PSO) algorithm. The model 3 is established by six variables. It includes gearbox bearing temperature at the previous time, gearbox bearing temperature at the previous two times, environment temperature, gearbox oil temperature, average wind speed, active power. It takes the same test data for testing, and the predicted results is shown in Figure 6. The MRE of model 3 is 1.85%. Comparing three models, it can be seen the accuracy of NEST model is higher than LSSVM model, and the variables involved in modeling NEST model are less. The NEST model is relatively simple and reliable. This paper verifies the accuracy and superiority of NEST model.

4.2 Residual analysis

According to the temperature residual obtained by inputting the training samples into NEST model, the residual threshold can be determined after statistics. The mean value of residual is 0.02. This paper defines k1 as 1. The threshold Ev is defined as 0.02. This paper carries out the residual analysis by sliding window for model 1, and the result is shown in Figure 7.
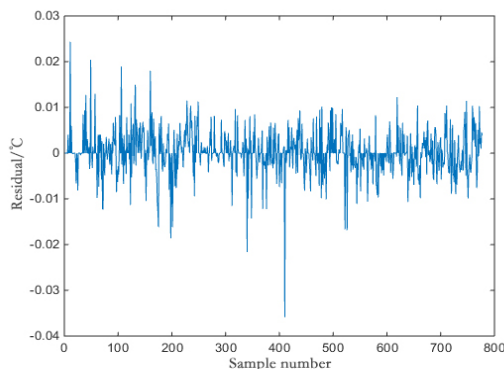
Figure 7. Residual analysis of sliding window.

Table 3. operation log.

No.	date	time	alarm reason
1	2019/7/8	3:05:20	gearbox oil high temperature
2	2019/7/11	19:27:37	gearbox oil high temperature

It can be seen from Figure 7 that the residual value of points 23 and 409 in the figure exceeds the threshold value of 0.02, and a fault warning will be carried out. By querying the SCADA records and operation logs, the unit did have faults and carried out alarm maintenance at that time. At 3:05 on July 8 and 19:27 on July 11, the gearbox oil high temperature fault had occurred, which corresponding to points 20 and 406 in the figure respectively.

4.3 Prediction of bearing temperature trend

In order to further verify the validity of the model, this paper adds temperature offset to gearbox bearing temperature, which is used to simulate the fault of gearbox bearing temperature abnormality. In 800 groups of testing data samples, starting from point 200, it adds 0.01 ℃ cumulative error. The change of residual mean value sequence and standard mean value sequence are shown in Figure 8 and Figure 9. It can be seen from the figure that the residual mean value sequence and the residual standard deviation sequence are gradually increasing, both of which exceed the corresponding thresholds and carry out fault early warning. Therefore, the residual mean value and the standard deviation can quickly reflect the fault status of the gearbox, and the NEST fault prediction model is effective and reliable.
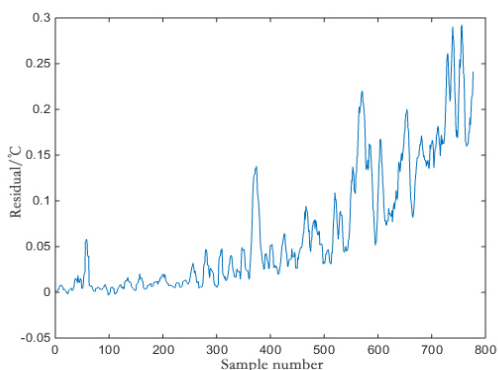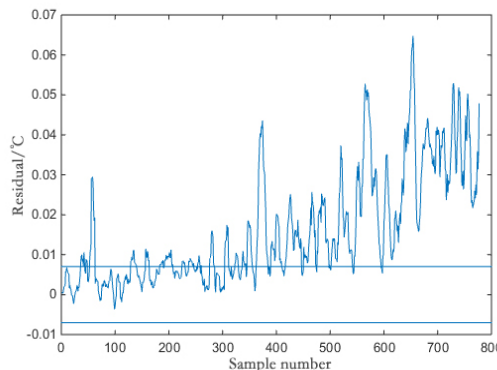

Figure 8. Residual mean analysis.

Figure 9. Residual standard deviation analysis.

5 Conclusion

This paper studies a fault prediction method of gearbox bearing. It uses grey relation analysis to get modeling variables, and makes sample data getting good integrity and redundancy by similarity analysis. Thus it gets the reduced process memory matrix, and trains the improved NEST model which can identify fault symptoms. Comparing with different models, the effectiveness and accuracy of the improved model are verified. The improved NEST model can shorten the operation time and reduce the operation amount, and it has high prediction accuracy. In actual needs of the project, it can effectively avoid equipment damage, and improve the economy and safety of wind turbine.

References

1. Crabtree C J, Feng Y, Tavner P J. Detecting incipient wind turbine gearbox failure: a signal analysis method for on-line condition monitoring. Proceeding of European Wind Energy Conference. C. Poland 2010:11-14.

2. Lu Bin, Li Yaoyu, Wu Xin. A review of recent advance in wind turbine condition monitoring and fault diagnosis. Lincoln: Proceedings of Power Electronics and Machines in Wind Application. C. 2009:1-7.

3. Hameed Z, Hong Y S, Cho Y M, et al. Condition monitoring and fault detection of wind turbines and related algorithm：a review. Renewable and Sustainable Energy Reviews. J. 2009, 13(1):1-39.

4. Amirat Y, Benbouzid M, Al-Ahmar E. A brief status on condition monitoring and fault diagnosis in wind energy conversion systems. Renewable and Sustainable Energy Reviews. J. 2009, 13(9): 2629-2636.

5. Guo Peng, David Infield, Yang Xiyun. Wind turbine gearbox condition monitoring using temperature trend analysis. Journal of Chinese Electrical Engineering Science. J. 2011, 31(32):129-136.

6. Alexios Koltsidopoulos, Papatzimos, Philipp R. Thies, et al. Offshore wind turbine fault alarm prediction. Wind Energy. J. 2019:1779-1788.

7. Alan Turnbull, James Carroll, Alasdair McDonald, et al. Prediction of wind turbine generator failure using two-stage cluster-classification methodology. Wind Energy. J. 2019:1593-1602.

8. Lorenzo Colone, Nikolay Dimitrov, Daniel Straub. Predictive repair scheduling of wind turbine drive-train components based on machine learning. Wind Energy. J. 2019:1230-1242.

9. James Carroll, Sofia Koukoura, Alasdair McDonald, et al. Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques. Wind Energy. J. 2019:360-375.

10. Pere Marti Puig, Alejandro Blanco M, Juan José Cárdenas, et al. Feature Selection Algorithms for Wind Turbine Failure Prediction. Energies. J. 2019:453.

11. Volkan Sevinc, Omer Kucuk, Merih Goltas. A Bayesian network model for prediction and analysis of possible forest fire causes. Forest Ecology and Management. J. 2020:457.

12. James Carroll, Sofia Koukoura, Alasdair Mcdonald, et al. Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques. Wind Energy. J. 2019, 22(3):360-375.

13. Pere Martipuig, Alejandro Blancom, J J Cardenas, et al. Feature selection algorithms for wind turbine failure prediction. Energies. J. 2019, 12(3):453.