Anomaly Identification of Wind Turbine Gearboxes Based on Similarity Theory

Xiao Yang

School of Energy Power and Mechanical Engineering
North China Electric Power University
Baoding, China
e-mail: 350964520@qq.com

Xinli Li

School of Control and Computer Engineering
North China Electric Power University
Baoding, China
e-mail: 601279228@qq.com

Minglei Hou

Baoding well foundry machinery Co., Ltd.
Baoding, China
e-mail: houminglei8888@126.com

Xiaoliang Fan

School of Energy Power and Mechanical Engineering
North China Electric Power University
Baoding, China
e-mail: fxl188@263.net

Abstract—Wind turbine condition monitoring and fault warning has important practical value to reduce maintenance costs, improve operational efficiency and reliability. In this paper, the characteristic parameters of SCADA system monitoring, based on the principle of similarity modeling techniques, through reasonable storage matrix construction process, the establishment of the gear box "health model" to cover the gearbox to work space. When the gearbox operation occurs, characterized in health parameters deviate from the model, when the offset distance exceeds the threshold value, the system gives the warning. Finally, the health model was validated, experiments show that the method used in this paper can be found early signs of abnormal gearbox and give warning. The results show that the health model by the similar principle established can identify abnormal state timely and accurately, and warning given before a failure occurs, which makes it easier to advance plans to organize maintenance equipment and personnel, to provide a reference for on-site maintenance.

Keywords- gearbox; similarity theory; Anomaly recognition; health model; failure warning

I. INTRODUCTION

The traditional wind turbines maintenances are essentially reactive maintenance，which can repair the fault [1], but cannot play predictable, prevention. With the growth of China's wind power installed capacity and the wind turbine running time, the reaction maintenance is no longer able to meet the needs of maintenance now, preventive maintenance based on the state is becoming more important [2]. Failure prediction technology is based on the analysis process to wind turbine condition monitoring data to determine the health of the system before the failure to identify potential trouble sign monitoring system sends alarms to prevent catastrophic failure or an accident, and timely repairs or replaces the unhealthy system.

Currently prediction method is widely used by the data processing algorithms for trend forecasting, which includes forecasting model based on time series analysis, forecasting model based on gray theory, artificial neural network forecasting model [3,4] and so on. Domestic and foreign scholars on the wind turbine condition monitoring, condition assessment and fault diagnosis has been extensively studied [5-7]. [8] The gearbox temperature trends were analyzed by statistical gearbox operating state with the model prediction residuals, early detection of potential failure of the gearbox. [9] given the options Relief method based on the characteristic parameters of the wind turbine pitch system characteristic parameters mining, as well as on the pitch system status more characteristic parameters from the abnormal recognition, a single parameter absolute threshold assessment compared to more timely and accurate identification its abnormal state.

High costs of wind turbine gearbox, and the event of failure, recovery time is longer, therefore, the gearbox early fault detection process is particularly important. This paper uses a similar principle to the proper classification and historical data on the fan to the gearbox, for example, to establish a healthy model gearbox, by comparing real-time data and health models, and tap abnormal gearbox and give fault warning.

II. WIND TURBINE GEARBOX MONITORING PARAMETERS

In this paper, we take Xinjiang Hami wind 1.5MW wind turbine gearbox for example, which is one of the main gearbox transmission components, its main function is to force the impeller generated by wind generator and passed to the appropriate speed .

SCADA system monitoring parameters of the unit up to a few dozen, we only analyzed for wind turbine gearboxes, selected nine parameters related to the gearbox run from SCADA system variables collected, including power generating units, wind speed, rotor speed , ambient temperature, engine temperature, gear box front bearing temperature, bearing temperature after the gearbox, gear box oil, gearbox oil temperature. The contents of each data

SCADA system includes a time stamp, active power, reactive power, power generating units, wind speed, rotor speed, ambient temperature, engine temperature, and so dozens of parameters. Meanwhile SCADA system also records the operational status of the unit, such as unit start, stop, gearbox oil pressure, oil temperature and low gearbox. Each log includes time, state ID, status descriptions.

III. THE PRINCIPLE OF SIMILARITY MODELING

A. Similarity theory

Similarity theory studies are more cutting-edge hotspot objects in data mining. Between patterns and pattern similarity measure between the sample and the sample, or a pattern of a sample and the sample is called the similarity, which is often used as a measure of the difference between the target scale[10]. To classify the samples during data analysis or data mining, we often need the degree of difference between samples, to judge samples.

Vector set X = {x1, x2, ..., xn} and the vector Y = {x1, x2, ..., xn}, constructor similarity function sim (X, Y), the similarity function sim (X, Y) $<\varepsilon$ when (ε is the set of similarity threshold), said vector X, Y to ε is bounded in similar circumstances, denoted $X \approx Y$. At the same similarity function to meet at three conditions:

- positive definiteness, sim (x, y) ≥ 0, if and only if X = Y when, sim (x, y) = 0;
- symmetry, sim (x, y) = sim (y, x);
- satisfy the triangle inequality, sim (a, c) $\leq$ sim (a, b) + sim (b, c).

Euclidean distance is commonly used in a similar principle operator, is a measure of the absolute distance between the midpoint of the multidimensional space, reflecting the degree of difference between the two phasor, the greater the distance the greater the difference between samples, is a the simplest and most common distance metric. Euclidean distance formula:

$$\text{dist}\ (X,Y) = \sqrt{\sum_{i}^{n}\left(x_i - y_i\right)^2} \qquad (1)$$

B. Nonlinear state estimation modeling principle

Nonlinear state estimate technology(NSET) method was developed by Singer and put forward a non-parametric modeling approach, the current generation of equipment failure warning, monitoring equipment, electronic products and other aspects of life prediction is widely used in a modeling method. NSET model extraction system through normal point, and then use these points constructed normal operation state space model [11].

There are n interrelated variables of a device, at time i observation vector written as:

$$X(i) = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$$

In the period of time the device to work properly, assume that there are m observation vector represent different working conditions, constitute the process of memory matrix D:

$$D = \begin{bmatrix} X(1) & X(2) & \cdots & X(M) \end{bmatrix}$$
$$= \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

Wherein each column vector of the matrix represent a particular state of normal operation, each row vector matrix representative of a time series of observed parameters. Through the rational selection process memory matrix m historical observation vector subspace spanned (i.e. process memory matrix D) to cover the dynamic range of the observation system uptime parameter.

Xobs vector system for the observation of a moment, NEST model by comparing the current observed state of the memory matrix in the process of the m conditions were compared to generate a m-dimensional weight vector W.

$$X_{est} = D \bullet W = \omega(1) \bullet X(1) + \omega(2) \bullet X(2) + \cdots + \omega(m) \bullet X(m) \qquad (2)$$

So Xest vector model for the prediction of outputs. It is a linear combination of m observation vectors.

The weights are state estimation similarity measure and process the memory matrix between the input and output vectors by NEST models residual minimize strike, as follows:

$$\min\varepsilon = \min\ (|X_{obs} - X_{est}|) = \min\ (|X_{obs} - D \bullet W|) \qquad (3)$$

Derivative of the formula number, you can get right to the residual value of the smallest cases:

$$W = (D^T \otimes D)^{-1} \bullet (D^T \otimes X_{obs}) \qquad (4)$$

Where $\otimes$ nonlinear operator based on similar principle is used in place of the dot product, to characterize the degree of similarity calculated by weighting the data between states to solve the matrix data associated caused irreversible.

When the memory matrix in the process covered space observation vector Xobs when the device is working properly located, and Euclidean distance matrix D is small, the corresponding NEST forecast vector Xest with high accuracy. When the device is abnormal, the observation vector will deviate from the normal work space, resulting in increased residuals.

Constitute the entire model is to study the characteristics of the equipment operation, using a combination of these modes will be able to produce an estimate of the model.

Fault warning flowchart NEST model show in Fig. 1

C. Nonlinear State Estimation Modeling Implementation

The above describes specific methods based on the basic principles of the theory of similarity modeling analysis are combined with specific examples of modeling.

- Data normalization process

Due to the observed variable gearbox model different dimension, and the absolute value of the difference between the different variable data is large, in order to eliminate the difference in value, to ensure the model can correctly assess the Euclidean distance between different

observation vector, we need for the selected nine measurements characteristic variables were normalized so that the actual measurement values are mapped to [01] range.
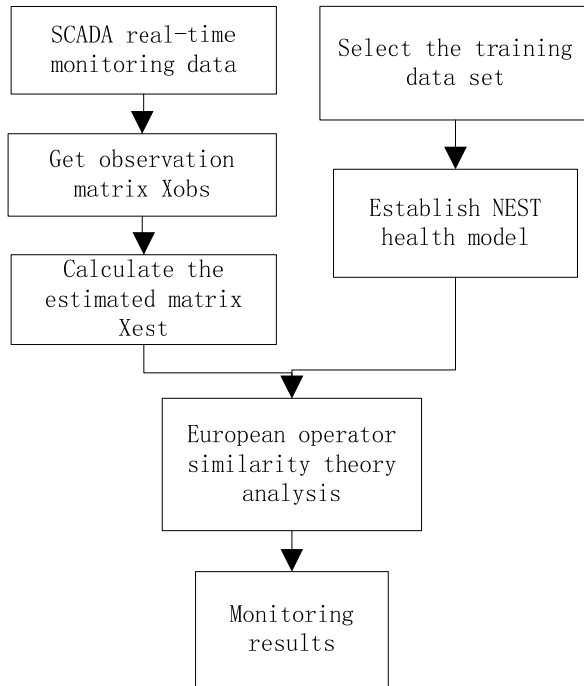
```
┌─────────────────────┐      ┌─────────────────────┐
│  SCADA real-time    │      │ Select the training │
│  monitoring data    │      │     data set        │
└─────────┬───────────┘      └──────────┬──────────┘
          │                             │
          ▼                             │
┌─────────────────────┐                 │
│  Get observation    │                 ▼
│   matrix Xobs       │      ┌─────────────────────┐
└─────────┬───────────┘      │  Establish NEST     │
          │                  │   health model      │
          ▼                  └──────────┬──────────┘
┌─────────────────────┐                 │
│  Calculate the      │                 │
│  estimated matrix   │                 │
│      Xest           │                 │
└─────────┬───────────┘                 │
          │                             │
          ▼                             ▼
       ┌───────────────────────────────────┐
       │      European operator            │
       │      similarity theory            │
       │          analysis                 │
       └─────────────────┬─────────────────┘
                         │
                         ▼
               ┌───────────────────┐
               │    Monitoring     │
               │     results       │
               └───────────────────┘
```

Figure 1. Warning flowchart model

● Selection of training data K

Establish NEST model is based SCADA massive historical data as the basis, sample training data K determines the accuracy of the model, so you first need to train K data modeling used for screening. The training data used to generate the device model should meet the following requirements [12]:

a) covers the dynamic process of the system under normal operating conditions characteristic monitored parameters, such as a wind turbine gearbox, the unit must contain start and stop the whole process in line with relevant monitoring parameters state;

b) data expressed in each wind turbine gearboxes - a normal state, the training data K in the state can not contain fault state.

● The construction process of the memory matrix D

Process memory matrix D contains m observation vector (i.e. the m conditions), to try to override the normal workspace wind turbine gearboxes. No fault data for Unit 2 of this article, Xinjiang Hami wind farm 2014/1/1 00:00 to 2014/4/1 00:00 SCADA system to collect gear box training, establish health gearbox model. To establish a healthy model for wind turbine gearboxes, we use Max Min method selected automatically, the data where the value of each observation mode process variables is selected into the memory array. According to this method of wind turbine gearboxes health model is comprised of 300 data samples, on behalf of the fans in the first quarter of 2014, gearbox normal operation. The observation vector for each point was normalized to the distribution in the phase space, consisting of a discrete hyper sphere. Use of

equipment healthy period of time to run at all state health facilities period were classified to form a normal range of values for each parameter under different conditions shown in Fig. 3, showing the upper and lower limits of a condition of the individual parameters.

Figure 2. Some working parameters change range

As shown above, each data set represents a normal condition, and gives the upper and lower limits of the specific conditions of each variable.

IV. GEARBOX ANOMALY IDENTIFICATION VERIFICATION

In order to verify the validity of the gearbox health model established above, is applying the data unit is not running during the same quarter in the test training. Discover The unit 2015 first quarter operating data records found that the crew at 15:12 on February 23, 2015 in time SCADA system displays "alarm _T117_ gearbox temperature is above limits" error message. Before selecting a fault occurs 15 hours of operating data as verification data model. Verification data into the model by comparing it can be seen from Fig. 4 wind turbine gearboxes monitoring data starts to deviate from the model time at 10:00, 12:27 time in its deviation from the Euclidean distance exceeds a set threshold, then the system gives warning, (Euclidean distance contribution rate that is greater when monitoring data Parameter deviate health model) while viewing the cause of the alarm as shown in Fig. 5, the model gives the contribution rate alarm parameters, you can determine the gearbox oil is continuing to rise, departing from the gearbox health model, given warning before a failure occurs. So we use the gearbox health model established principle of similarity can be given before a failure occurs in anticipation, timely and accurately identify the abnormal state of the gearbox, troubleshooting and engineering applications than the absolute threshold based on a single parameter assessment methods system has a high sensitivity.
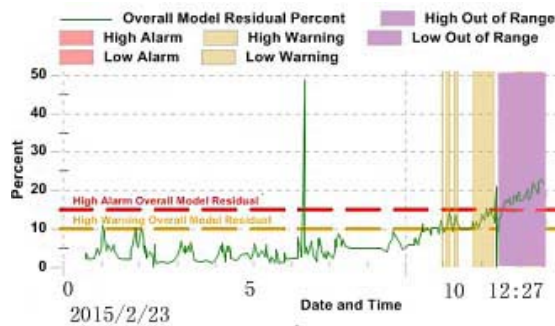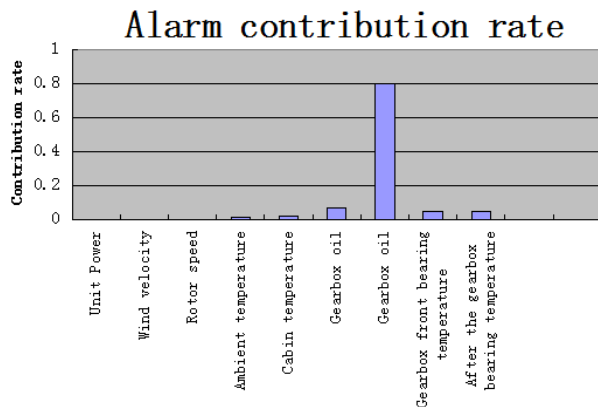
Figure 3.　Model deviation warning

Figure 4.　Model parameters deviate contribution rate

V. CONCLUSION

In this paper, a similar principle, a reasonable choice associated with monitoring gear operating parameters established gearbox own history based on device operating data "health model", change the system equipment health model will be randomly group operating conditions change accordingly , closer to the actual operation of the equipment. By extracting the scene of trouble-free operation of the model data to establish the validity of the above verification results show that health model by a similar principle established timely and accurately identify abnormal state, warning given before a failure occurs. By early detection of faults, avoid premature failure caused more serious damage fault, easy to advance plans to organize maintenance equipment and maintenance personnel, to provide a reference for on-site maintenance.

REFERENCES

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [1]. Where appropriate, include the name(s) of editors of referenced books. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first . . ."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] ZhANG Xiaotian. Research on Fault Prediction of Wind Turbine Main Components Based on Regression Analysis. North China Electric Power University, 2013.

[2] GU Yujiong, SU Luwe, ZhONG Yang. Online Fault Early Warning Method for Wind Turbine Gearbox Based on Operational Condition Division[J]. Electric Power Science and Engineering,2014, 30(8): 1-5.

[3] WANG Xue. The test of intelligent information processing[M]. Tsinghua University press,2008.33-56.

[4] PENG Yu, LIU Datong, PENG Xiyuan. A review: Prognostics and health management[J]. JOURNAL OF

[5] Arabian-Hoseynabadi H, Oraee H, Tavner P J. Failure modes and effects analysis (FMEA) for wind turbines[J]. International Journal of Electrical Power & Energy Systems, 2010, 32(7): 817-824.

[6] LI Hui, HU Yaogang, TANG Xianhu. A Matter-element Assessment Method of a Grid-connected Wind Turbine Driven Generator System Under On-line Operating Conditions[J]. AUTOMATION OF ELECTRIC POWER SYSTEMS, 2010 (33): 103-109.

[7] WANG Ruichuang, LIN Fuhong. Research on on-line monitoring and fault diagnosis of wind turbine[J]. EAST CHINA ELECTRIC POWER, 2009, 37(1): 190-193.

[8] GUO Peng, YANG Xiyun. Wind Turbine Gearbox Condition Monitoring Using Temperature Trend Analysis[J]. Proceedings of the CSEE, 2011, 31(32): 129-136.

[9] YANG Chao, LI Xuewei. Conditions Characteristic Parameters Mining and Outlier Identification for Electric Pitch System of Wind Turbine[J]. Proceedings of the CSEE, 2014, 34(12): 1922-1930.

[10] Gross K C, Singer R M, Wegerich S W, et al. Application of a model-based fault detection system to nuclear plant signals[C]//Proc. 9th Int. Conf. Intelligent Systems Applications to Power Systems. 1997: 6-10.

[11] Guo P, Infield D. Wind Turbine Tower Vibration Modeling and Monitoring by the Nonlinear State Estimation Technique (NSET)[J]. Energies, 2012, 5(12): 5279-5293.

[12] CHANG Shuping, GUO Jianglong, LV Yukun. The Application of NSET Modeling Method in Failure Prognostic System[J]. COMPUTER ENGINEERING & SOFTWARE, 2012, 32(7): 57-60.