Renewable Energy 127 (2018) 825–834

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Anomaly detection and fault analysis of wind turbine components based on deep learning network

Hongshan Zhao [a, *], Huihai Liu [b], Wenjing Hu [a], Xihui Yan [a]

[a] Department of Electrical and Electronic Engineering, North China Electric Power University, 071003, Baoding, China
[b] Bohai New Area Power Supply Branch, Cangzhou City, State Grid Hebei Electric Power Co., Ltd., Cangzhou 061113, China

ARTICLE INFO

Article history:
Received 21 October 2017
Received in revised form
3 May 2018
Accepted 4 May 2018
Available online 5 May 2018

Keywords:
Wind turbine
SCADA data
Anomaly detection
Deep learning networks
Extreme value theory

ABSTRACT

Continuous monitoring of wind turbine health using early fault detection methods can improve turbine reliability and reduce maintenance costs before they reach a catastrophic stage. To achieve anomaly detection and fault analysis of wind turbine components, this paper proposes a deep learning method based on a deep auto-encoder (DAE) network using operational supervisory control and data acquisition (SCADA) data of wind turbines. First, a component DAE network model using multiple restricted Boltzmann machines (RBM) was constructed. Previously collected normal SCADA data from wind turbines were used to train this multilayer network model layer-wise to extract the relationships between SCADA variables. Then, a reconstruction error (R_e) was calculated by using the DAE network input and its output reconstruction value, which was defined as the condition detection index to reflect the component health condition. Due to the acute changes and disturbances of wind speed in actual operation, the calculated detection index always has an extreme distribution that can cause false alarms. Therefore, an adaptive threshold determined by the extreme value theory was proposed and used as the rule of anomaly judgement. The method can not only implement early warning of fault components but also deduce the physical location of a faulted component by DAE residuals. Finally, the effectiveness of the proposed method was verified by some reported failure cases of wind turbine components.

1. Introduction

The harsh and variable operating environment of wind turbines is the main reason for failure of their components, such as the gearbox, main bearing, generator, inverter and controller [1,2]. Frequent component malfunctions lead to low availability and expensive repair costs of wind turbines. To analyse and evaluate the health status of wind turbines in a timely manner, many types of condition monitoring or SCADA sensors are allocated in different functional units. Hence, it is beneficial to achieve condition assessment and abnormal detection of wind turbine components by utilizing the data collected through these sensors. Generally fault detection and isolation (FDI) approaches can be divided into model-based methods and data-driven methods. The model-based methods, such as the observer-based methods [3–6], use explicit system dynamic models and control theories to generate residuals for FDI. On the other hand, the data-driven model methods use data mining techniques, such as artificial intelligence algorithms [7–9], to capture discrepancies between observed data and that predicted by a model.

The acquisition data from wind turbine sensors usually contains condition monitoring data and operation data (namely, SCADA). The turbine condition monitoring data, such as vibration and oil monitoring signals, are mainly used for equipment fault diagnosis and health analysis [10–14]. In the literature [11,12], Zhang and Huang et al. proposed using the time-wavelet energy spectrum to extract the time domain and frequency domain features of the bearing vibration signal for use in bearing fault diagnosis; Ma and Zhao *et al* [13,14] combined variational modal decomposition and the Teager energy operator to demodulate the recorded signals of the bearing in failure and extracted the features of fault successfully. Random changes of wind speed also lead to some restrictions for the fault diagnosis algorithm by the use of vibration signals. Based on this, Ren [15] used the angular domain cascade maximum correlation kurtosis (CMCKD) technique for bearing fault diagnosis by converting the time-domain vibration signal to an angular-domain signal. In addition, oil detection techniques, such as

* Corresponding author.
E-mail address: zhaohshcn@ncepu.edu.cn (H. Zhao).

https://doi.org/10.1016/j.renene.2018.05.024
0960-1481/

particle counting, image recognition, etc., can effectively be used to detect the fault types of the gearbox [16,17]. These methods are mostly used in off-line analysis. Since the application of the above methods is expensive and complicated, some researchers are beginning to study online algorithms [18,19].

Though wind turbine SCADA data, such as voltage, current, power, temperature and pressure, cannot contain detailed transient information such as vibration signals, there is a large volume of this data which is easily obtained in wind farm, and it contains multiple types of parameters and timeliness characteristics. Therefore, SCADA data can be explored to help with turbine component fault diagnostics and O&M improvements, and most importantly, the use of these data for condition detection is cost effective. Yang [20] developed a method to process the raw SCADA data for reducing the statistical error, and then extracted the correlation from these data to achieve the condition analysis of a wind turbine driven-chain system under varying operating conditions. Based on the cointegration method, Phong [21] extracted the cointegration residual as an index of condition monitoring for wind turbine fault detection. Considering that SCADA data contains both trend and abrupt information, the clustering and principal components analyses are used for fault detection, and gearbox failure is identified based on rotor speeds and powers [22]. Sun [23] utilized a genetic algorithm combined with partial least squares regression (GAPLS) for input parameter selection to reduce the redundant parameters and achieve anomaly detection of a wind turbine generator based on the normal behaviour models of SCADA data. The temperature of wind turbine components, such as bearings and the gearbox, can also reflect their current status. In Ref. [24], three temperature signals, a main bearing temperature, a lubrication oil temperature of the gearbox, and a winding temperature of the generator, were used for fault detection of wind turbine components via a back-propagation neural network (BPNN) approach. Much effort has been made in developing wind turbine condition monitoring systems, however due to high cost and available achievement limitations, focused study of the cost-effective wind turbine condition detection technique using existing SCADA data is required for the smart operation and maintenance of large-scale wind farms today.

Combined with the SCADA characteristics of wind turbines, this paper presents a deep auto-encoder (DAE) method for anomaly detection and fault analysis of wind turbine components. Compared with the traditional learning models that exists some shortcomings such as insufficient learning ability and low accuracy [25,26], the DAE network model is more accurate in modelling wind turbine component dynamic behaviour by working on a closer level of mimicking the working process of a natural brain.

For DAE method, we first built a wind turbine component DAE model by taking advantage of SCADA data associated with the corresponding component. Then, the DAE network was trained on the SCADA data samples of wind turbines in normal operation, and the internal intrinsic relationships between input and output is mined to form parameters of the component DAE model. Next, we defined the index of component health condition by the reconstruction error of the input and output of the DAE network. To monitor the trend of the index effectively, a new method to determine the dynamic adaptive threshold was proposed on the basis of the detection index calculated by the wind turbine SCADA data. Finally, the condition index and the adaptive threshold and the residual of the input-output were used to achieve the abnormal detection and fault location analysis of wind turbine components.

2. Deep learning network

A deep learning network is a deep neural network with multiple hidden layers between the input and output layers, which can model complex non-linear relationships among multiple types of variables. The parameters of this network are initialized by unsupervised learning for input data layer by layer, and then supervised learning is used for fine-tuning. In this network framework, deep learning models lead to more complex features at higher output layers, and the learned complex features will be invariant with the change of input data [27–29]. In this paper, the deep auto-encoder deep learning method is used to extract the relationships and features from the SCADA data of wind turbine components. For a DAE network, its training process including two phases, pre-training and fine-training, that are applied to obtain parameters of the DAE network model.

2.1. Pre-training of the DAE

The DAE is a deep learning network composed of multiple Restricted Boltzmann Machine (RBM) stacks. In the DAE, the output of each RBM is considered as a new input of a higher level RBM to achieve the transmission of learning results layer by layer. This procedure is repeated many times to initialize the parameters of each hidden layer.

The mechanism of the DAE network includes two processes, encoding and decoding. In the encoding process, the input X is first transformed to produce a set of features for further layer-wise transformations, and the more complex features are obtained in higher layers. Finally, the code Y is acquired through the encoding process. Similarly, the code Y is iteratively transformed back to the original input via RBMs and the reconstruction of X, $\widehat{X}$, is generated in the decoding process. Fig. 1 shows the mechanisms of encoding and decoding.

The RBM is a two-layer-network of a random Markov type with N visible units $v_i = \{0,1\}^N$ and M hidden units $h_j = \{0,1\}^M$, as shown in Fig. 2. The energy model is introduced to the describe energy of the joint configuration units in the RBM, by

$$E(v,h;\theta) = -\sum_{i=1}^{N} a_i v_i - \sum_{j=1}^{M} h_j b_j - \sum_{i=1}^{N}\sum_{j=1}^{M} w_{ij} v_i h_j \tag{1}$$

Where $\theta = \{w_{ij}, b_i, a_j\}$, w_{ij} is the weight between visible unit i and hidden unit j and b_i and a_j are the bias of visible and hidden unit, respectively. They are the model parameters.

In the RBM, the joint distribution between units based on the energy model is described as

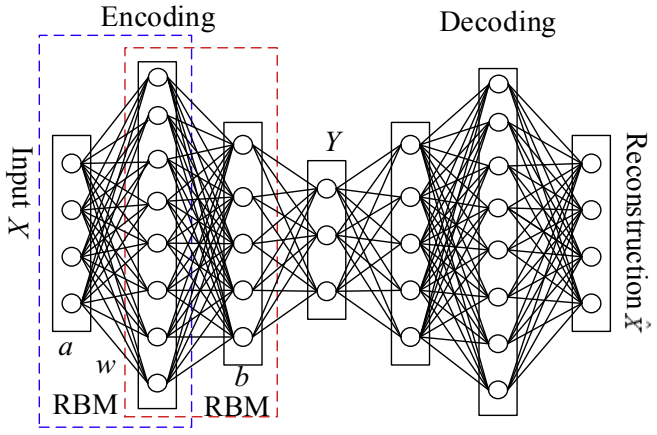

Fig. 1. Structure of DAE network.

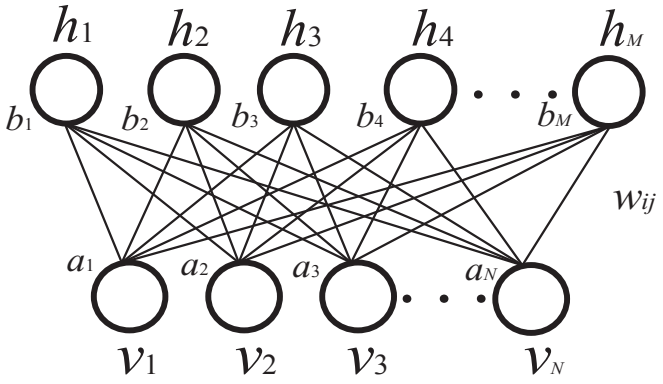

Fig. 2. Structure of RBM.

$$P(v,h;\theta)=\frac{1}{Z(\theta)}\exp(-E(v,h;\theta)) \tag{2}$$

$$Z(\theta)=\sum_{v}\sum_{h}E(v,h;\theta) \tag{3}$$

where $Z(\theta)$ is the normalizing constant. The network gives the probability values of each input vector via the energy function, and the probability can be raised by changing parameter θ to adjust the energy value in Eq. (1).

The conditional distributions of hidden units h and input vector v in the RBM is given by

$$P(h_j=1|v)=f\left(\sum_{i=1}^{N}W_{ij}v_i+b_j\right) \tag{4}$$

$$P(v_i=1|h)=f\left(\sum_{j=1}^{M}W_{ij}h_j+a_i\right) \tag{5}$$

$$f(x)=\frac{1-e^{-2x}}{1+e^{-2x}} \tag{6}$$

where $f(x)$ is an activation function, taken as *tanh*. The essence of the activation function is to retain the characteristics of the activated neuron and map it out. In this paper, *tanh* is substituted for *sigmoid* as the activation function. When the input is between [-1, 1], the *sigmoid* value changes sensitively. Once the input is close to or out of the interval, the *sigmoid* value will lose its sensitivity, and the accuracy of this network will be reduced when the *sigmoid* value is in the saturated condition. In addition, *tanh* has a faster output convergence than function *sigmoid*. *Tanh*'s output and input can maintain a non-linear monotonic rise and fall relationship to meet the BP network gradient solution [30].

In each RBM network, the activation information of the hidden layer units is represented as the extracted features of the input data. In other words, the learning goal of the RBM is to obtain the parameter θ to recover the original input data perfectly. Therefore, the likelihood function of the visible layer v is constructed as in (7), to obtain the parameter θ.

$$L(\theta;v)=\prod_{v}P_\theta(v)=\prod_{v}\frac{\sum_h e^{-E_\theta(v,h)}}{\sum_{v,h}e^{-E_\theta(v,h)}} \tag{7}$$

Then, in Eq. (7), the logarithm is taken on both sides, and the logarithm function derivative with respect to $\theta=\{w,b,a\}$ is presented as follows,

$$\begin{aligned}\frac{\partial \ln(L(\theta;v))}{\partial\theta}&=\frac{\partial\sum\ln(P_\theta(v))}{\partial\theta}\\&=\sum\left\{E_{P_\theta(h|v)}\left[-\frac{\partial E_\theta(v,h)}{\partial\theta}\right]-E_{P_\theta(v,h)}\left[-\frac{\partial E_\theta(v,h)}{\partial\theta}\right]\right\}\end{aligned} \tag{8}$$

Next, the contrastive divergence (CD) algorithm is applied to estimate the gradient. We use one-step CD learning to update the parameters θ [31]. To reduce the information loss of input X and obtain accurate parameters for the DAE, the loss function can be defined as Eq. (9),

$$J_{AE}(\theta)=\frac{1}{N}\sum_{x\in X}R_e\left(X,f\left(\widehat{X}\right)\right) \tag{9}$$

The parameter θ is obtained by using the gradient descent method to minimize the loss function. Thus, the update rule is described as,

$$w^{k+1}=w^k+\varepsilon\frac{\partial J_{AE}(\theta)}{\partial w} \tag{10}$$

$$a^{k+1}=a^k+\varepsilon\frac{\partial J_{AE}(\theta)}{\partial a} \tag{11}$$

$$b^{k+1}=b^k+\varepsilon\frac{\partial J_{AE}(\theta)}{\partial b} \tag{12}$$

where ε is the learning rate. So, a reasonable parameter θ can finally be obtained by the RBM's hierarchical training process.

After training each RBM, the learnt information from the SCADA data of wind turbine components lies in the hidden layer, which can be used as the input of a higher layer to produce essential information and obtain its hidden layer parameter θ until the entire DAE network training is completed.

During the training, the SCADA data, as unlabelled data, for long-term normal operating conditions of the wind turbine is selected as the training data. The variable set (vector) of each component i is defined as X_i,

$$X_i=[x_{i1},x_{i2},\cdots,x_{im}] \tag{13}$$

where i is the name of wind turbine component, x_{ij} is the jth parameter in the SCADA variable set of the component i. For example, the generator SCADA variable set includes voltage, current, power, winding temperatures, and generator bearing temperatures, etc. to be able to reflect the condition of generator.

To reduce calculation error caused by the numerical differences of the different types of parameters of wind turbine components and to keep the original data structure invariant, the SCADA data is preprocessed into the interval [0, 1] by normalization.

2.2. Fine-tuning of DAE parameters

After the pre-training of the DAE hidden layers with the RBMs, the weights and biases of each hidden layer of multi-layer RBMs are updated, and the framework of the DAE model is constructed. The pre-training of the DAE network is an unsupervised learning of the SCADA data. Thus, the results of the learning can be used as a priori values for supervised learning of the DAE network.

Taking advantage of the SCADA labelled data in long-term normal operation SCADA data, the back-propagation (BP) algorithm is used for a supervised learning to improve the representation of data features and optimize the parameters of hidden layers in the fine-tuning. Because the fine-tuning training only needs local search based on the parameters obtained by pre-training, the convergence time of the optimization is significantly shortened in this process. Finally, the parameters obtained by this training process are better than those trained by the BP algorithm alone [32,33]. After fine-tuning, the globally optimized parameters of the DAE network are obtained.

For DAE networks of wind turbine components, the mapping of input to output is one-to-one, and each item has the same physical meaning. $\widehat{X}$ shown in Fig. 1 is the reconstruction of input X, which corresponds to SCADA variables in X. Therefore, the condition of wind turbine components can be evaluated by analysing the relationship between X and $\widehat{X}$.

3. Condition evaluation and anomaly detection of wind turbine components

When the wind turbines are operating normally, the SCADA variables generally should satisfy their relevance in the physical sense. Once an abnormality occurs, the internal relevance between these variables will be destroyed. Therefore, the SCADA data X in the abnormal condition will deviate from the reconstruction $\widehat{X}$ estimated by DAE network.

The residual (r) and reconstruction error (R_e) are respectively defined by (14) and (15),

$$r = [r_1, \cdots, r_s] = \widehat{X} - X \tag{14}$$

$$R_e = \widehat{X} - X_2 \tag{15}$$

where the residual r denotes the difference between $\widehat{X}$ and X, r_i is the residual of i-th variable x_i in vector r; the reconstruction error is a Euclidean norm that describes the ordinary distance from X to $\widehat{X}$.

The residual can reflect each SCADA variable change and trend in the operation of the wind turbine, while the reconstruction error describes the overall behaviour of monitored component; hence, R_e is defined as the overall condition evaluating index of the wind turbine component.

To monitor the trend of R_e and detect its anomaly variation, an adaptive threshold is designed as an alarm decision criterion. When the R_e crosses the threshold and keeps above its threshold, this means that an incipient fault may occur. Therefore, the rule for determining the abnormal condition of wind turbine components is defined as follows,

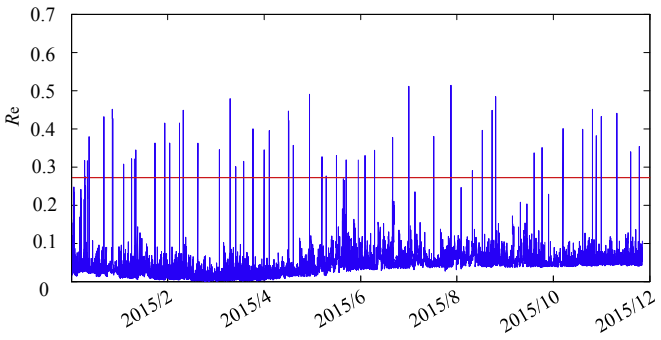

Fig. 3. R_e of wind turbine generator under normal condition.

$$D = \begin{cases} 1(alarm), & R_e \geq R_{th} \\ 0(normal), & R_e < R_{th} \end{cases} \tag{16}$$

For example, after the establishment of a wind turbine generator's DAE network, the R_e is constantly computed as shown in Fig. 3. We can see that R_e is non-stationary even when the wind turbine runs normally. Due to the fluctuations in wind speed, especially the gust, many extreme exist in R_e. If a constant threshold is applied, false alarms may be caused. Therefore, the non-stationarity of R_e should be taken into consideration in determining a proper threshold. In this paper, the extreme value theory is applied to design the adaptive threshold to determine whether the turbine component condition is abnormal.

4. Adaptive threshold based on extreme value theory

Considering the non-stationarity and multiple extreme value points of condition evaluating index R_e, we apply the extreme value theory to analyse the distribution of extreme values in R_e and to design a threshold for abnormal detection and condition evaluation of wind turbine components.

Let $X_1, X_2, \cdots X_n$ be sample vectors of independent and identically distributed variables with the distribution F. X_i usually represents the values in a unit time, one minute is chosen to be the unit time in this paper. For these sample vectors, we find that the most appropriate statistic to study the tails is,

$$M_n = \max(X_1, X_2, \cdots X_n) \tag{17}$$

where M_n is the maximum value of R_e in each minute. For a set of M_n, the probability distribution function can be derived by,

$$\begin{aligned} &P_r(M_n \leq z) = P_r\{X_1 \leq z, X_2 \leq z, \cdots, X_n \leq z\} \\ &= P_r\{X_1 \leq z\} \times P_r\{X_2 \leq z\} \times \cdots \times P_r\{X_n \leq z\} \\ &= \{F(z)\}^n \end{aligned} \tag{18}$$

As the distribution F is unknown, approximate families of models for F^n which can be estimated only using the extreme data are considered. However, $F^n \to 0$ *as* $n \to \infty$. To avoid the degradation of M_n to a point, we assume that there exist μ and σ such that

$$P_r\left\{\frac{(M_n - \mu)}{\sigma}\right\} \to G(z) \quad as \quad n \to \infty \tag{19}$$

where μ is a location parameter, σ is the scale parameter, and $G(z)$ is the generalized extreme value distribution function [34] and can be described as follows,

$$G(z) = \exp\left\{-\left[1 + \left(\frac{z - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \tag{20}$$

defined on $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, where the parameters satisfy $-\infty < \mu < \infty$, $\sigma > 0$, $-\infty < \xi < \infty$ respectively, and ξ is the shape parameter.

The DAE networks of different components have learned their respective intrinsic correlation from SCADA data in a variety of wind conditions under healthy conditions of the wind turbine. Therefore, when collected SCADA data are used to calculate R_e, though R_e of each component is non-stationary, the mean of R_e will not shift greatly. Therefore, their means can be approximated as constants, but their variances are non-constant. Hence, the location parameters and scale parameters are expressed as follows,

$$\mu(t) = \beta_0 \tag{21}$$

$$\sigma(t) = \exp(\beta_0 + \beta_t g(t)) \tag{22}$$

where β_0 and β_t are constant coefficients and $g(t)$ is a function describing the variable operating condition, i.e., affected by the change of the SCADA data. The parameters μ,σ and ξ can be ensured by the maximum likelihood estimation approach. Having obtained these parameters for the appropriate extreme value distribution, the cumulative distribution function (CDF) is used to establish the confidence limits. If the CDF is expressed in terms of a generalized extreme value distribution, then the final determination of warning thresholds can be calculated by,

$$R_{th} = \mu(t) - \frac{\sigma(t)}{\xi}\left[1 - \left\{-\ln(1-p)\right\}^{-\xi}\right] \tag{23}$$

where p is a confidence limit.

5. Case studies

To verify the proposed model framework validation, the generator and gearbox training datasets are established by collecting nearly 4-years of normal SCADA operation data of 1.5 MW wind turbines from an actual wind farm. The two datasets are applied to develop the component DAE model, which allows the DAE network to fully extract the features in normal operating conditions including different wind speeds, climates and seasons, etc., to meet the needs of the wind turbine component's condition detection. In the following three cases analyses, the DAE models for wind turbine components are all chosen to have seven hidden layers.

5.1. Detection analysis of wind turbine gearbox condition (Case A)

The gearbox DAE network is constructed with wind turbine gearbox variables and other variables closely related to the gearbox condition. These SCADA variables as inputs to the gearbox DAE model are shown in Table 1, and the vector of these variables can be expressed as,

$$X_B = [v_0, T_i, T_o, T_r, T_c, F_i, F_w, T_w, \Omega, P] \tag{24}$$

We applied the gearbox dataset to train the gearbox DAE model and neural network (NN), then calculate R_e by using the recorded failure SCADA data, as shown in Fig. 4.

We see that the gearbox R_e in Fig. 4(a) fluctuates within the adaptive threshold range before time t_0, but after t_0, R_e approaches its upper limit and then quickly crosses its upper limit. Approximately 5 h later, R_e increases continuously until the downtime. According to the proposed DAE model analysis, this means that at time t_0, the gearbox detected an abnormality, and there may be an impending fault. Compared with the actual fault record, the DAE method can detect the fault more than 14 h ahead of actual downtime, which can provide sufficient time to take some actions for wind turbine gearbox maintenance. From Fig. 4, it can be seen that the R_e calculated by the DAE network is smaller than that of the NN, which shows that DAE network can extract more sufficient dynamic features from normal SCADA data, in other words, the DAE network model is more accurate in describing wind turbine component dynamic behaviour. Moreover, the DAE condition index can be earlier to detect anomaly condition than that of NN. The proposed DAE network can detect incipient fault earlier nearly 10 h than the NN.

Next, in Fig. 5, we further analyse the residuals of each of the variables in (24) calculated by the gearbox DAE network. Seen from Fig. 5, the residual of the lubrication oil filter inlet (LOFI) press shows an upward trend as early as t_1 on August 14th, which indicates that the gearbox may already have some hidden problems. Although the growth of the gearbox inlet oil (GIO) and gearbox output shaft (GOS) temperature residuals is not as fast as that of the LOFI press, they still slowly rise. However, other parameter residual changes, such as gearbox inlet oil (GIO) temperature and gearbox input shaft (GIS) temperature, are far below the residual change of the LOFI press. Thus, these parameters stay in their normal operation range. Therefore, according to the analysis of the DAE method, the reason for the increased gearbox LOFI press and GIO/GOS temperature may be that the gearbox oil filter inlet is clogged or the oil contains impurities. An actual inspection verified that the shutdown of gearbox was caused by the clogged gearbox oil filter inlet.

Table 1
Description of wind turbine gearbox operation variables.

No.	Parameter	Notation
1	wind speed	v_o
2	gearbox input shaft temperature	T_i
3	gearbox output shaft temperature	T_o
4	gearbox inlet oil temperature	T_r
5	gearbox oil sump temperature	T_c
6	gearbox lubrication oil filter inlet press	F_i
7	gearbox water cool water press	F_w
8	gearbox cooling water temperature	T_w
9	Generator speed	Ω
10	power	P

5.2. Anomaly analysis of generator rear bearing (Case B)

The wind turbine generator DAE model was established based on the SCADA data, such as voltage, current and bearing temperature, etc., and the model input vector was expressed as

$$X_G = [v_0, P, U_1, U_2, U_3, I_1, I_2, I_3, T_t, T_s, T_{ai}, T_{ba}, T_{bb}, T_{u1}, T_{v1}, T_{w1}] \tag{25}$$

and the detailed variables of input vector are described in Table 2.

This is a case of a wind turbine (301#) generator rear bearing failure caused by lack of lubricating oil in July 2015. The recorded SCADA data was analysed by the proposed DAE network model and the NN, and the curve of generator R_e and the key input variable residuals are shown in Fig. 6 and Fig. 7.

Seen from the variation tendency of the condition index R_e in Fig. 6(a), the generator DAE method finds that R_e exceed its threshold at time t_0 on July 14, and the over-limited state was kept on until the downtime on July 18. This means that the DAE model can detect the abnormality three days earlier than the actual wind turbine downtime. Compared with Fig. 6(b), we can find that the proposed DAE network can detect incipient fault earlier nearly 8 h than the NN.

In Fig. 7, we found that the maximum variation residual curve is the temperature of the generator rear bearing, while other parameters residuals, such as the winding temperature, cooling air temperature and voltage, almost always fluctuate near zero. This means the main reason of the closing down was due to the high temperature of the bearing. Therefore, we are sure that it may be due to a lack of lubricating grease in the generator rear causing the bearing to wear and its temperature to rise. The analysis result is exactly same as the actual situation.

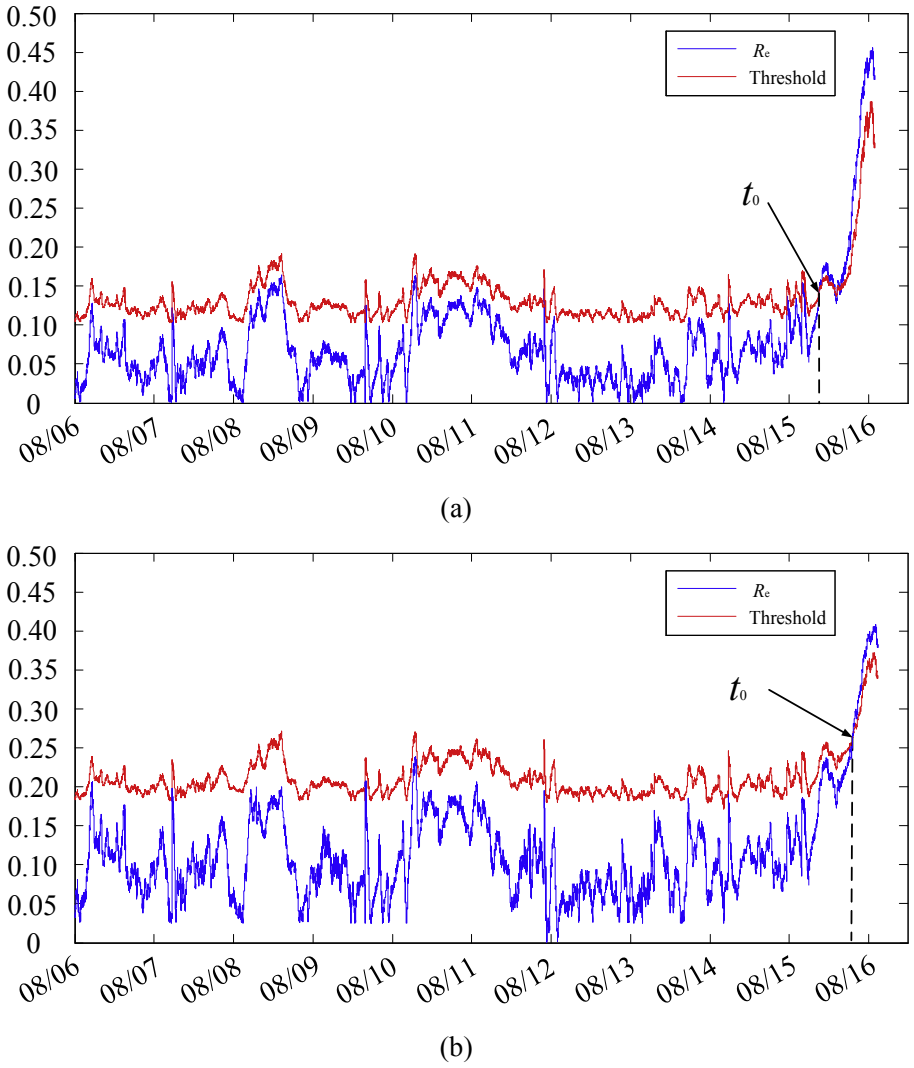

Fig. 4. Control Chart of Wind Turbine Gearbox Re (a) by DAE network (b) by NN.

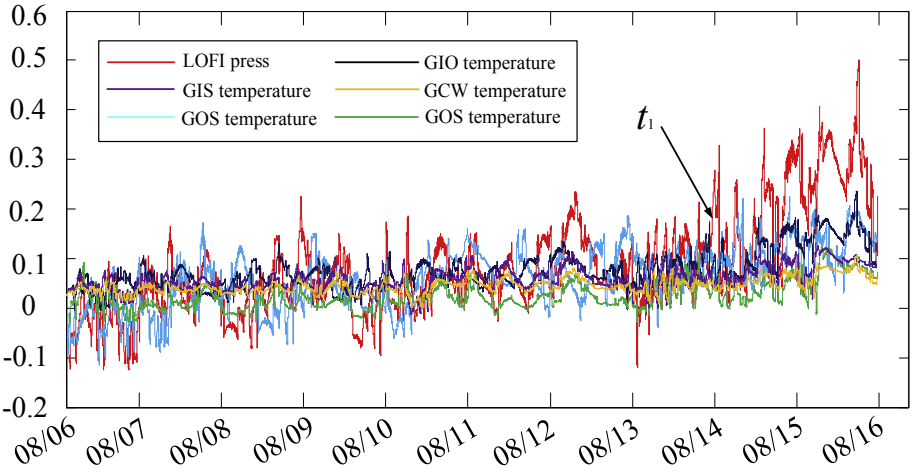

Fig. 5. The residual of gearbox SCADA variables.

Table 2
Description of wind turbine generator variables.

No.	Parameter	Notation
1	wind speed	v_0
2	grid side power	P
3	grid U1 voltage	U_1
4	grid U2 voltage	U_2
5	grid U3 voltage	U_3
6	grid I1 current	I_1
7	grid I2 current	I_2
8	grid I3 current	I_3
9	generator slip temperature	T_s
10	generator cooling air temperature	T_{ai}
11	generator front bearing temperature	T_a
12	generator rear bearing temperature	T_b
13	U1 winding temperature	T_{u1}
14	U2 winding temperature	T_{v1}
15	U3 winding temperature	T_{w1}

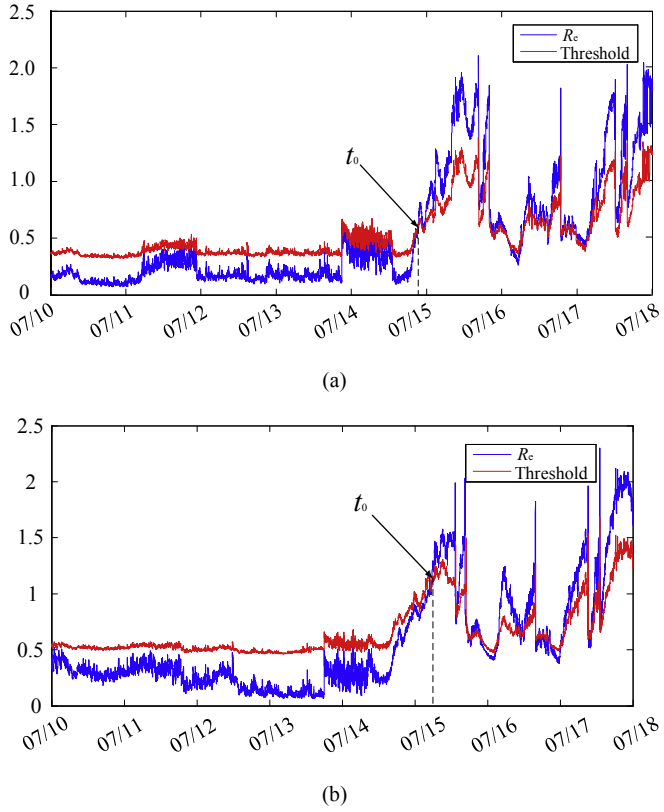

Fig. 6. The control chart of 301# wind turbine generator R_e (a) by DAE network, b) by NN.

5.3. An inverter failure analysis via wind turbine generator DAE model (Case C)

A record of actual fault alarms and treating processes that occurred in a wind turbine electrical system is shown in Table 3. There were multiple times alarms of generator 3-phase current asymmetry during the period of June 21 to June 24, 2016. The steps the operators took were as follows: After the warning on July 21, they prejudged the B-phase CT and PT faults of both the generator side and line side, and the CTs and PTs were replaced during 10:30–16:25. There were still 3-phase asymmetric fault alarms on July 22 after replacement, and technicians entered the wind turbine nacelle to inspect and found no abnormality; then, they reset the alarm signal. The IGBT fault on the side of the inverter B-phase was not found until July 24 after the three-phase asymmetrical fault warning occurred again. After the repair, the three-phase asymmetrical fault never occurred, the wind turbine operated normally.

Analysis by the generator DAE model and NN with input vector the same as in Table 2, it can be seen from the R_e index in Fig. 8. The alarm signal of DAE network was transmitted on t_0, which is 4 h earlier than the actual downtime on July 21. Compared with the detection result based on NN in Fig. 8(b), the proposed DAE network can detect incipient fault earlier about 2 h than the traditional learning network.

Through analysis of DAE model residuals in Fig. 9, asymmetry characteristics are seen in both the 3-phase voltage and current on the generator side, namely, the voltage and current of phase B deviated seriously, while the residuals of the voltage and current of phase A and C is small. From this, it can be concluded that the possible failure may be: (i) a B-phase winding fault of the generator, (ii) a B-phase PT and CT fault of the generator, or (iii) a B-phase bridge fault of the inverter connected to generator.

In Fig. 10, we analysed the changes of the grid-side active power and the wind speed, the wind turbine's output power to power grid was normal, which satisfied the relationship between the wind power and wind speed. Therefore, the fault of B-phase generator winding cannot be the culprit.

Upon deep analysis of the residual waveform of the asymmetric 3-phase voltage and current in Fig. 9, it is seen that only the residual of B-phase voltage and current is relatively large, which is the major cause of the index R_e alarm, but it cannot be concluded that there was a fault in PT and CT of phase B. Although the residual change of the B-phase voltage and current was huge, the probability of PT and CT simultaneous failure is very low, and the residual change trend of the B-phase voltage and current was very consistent with that of phase A and phase C. Thus, the PT and CT outputs of the B-phase can reflect the change of its primary voltage and current. Therefore, it can be speculated that there might be a failure on the inverter connected to generator. If there was a failure on the B-phase power electronic components of the inverter, the fault behaviour is completely consistent with each variable residual change of the generator DAE model, and the change in the inverter power output was very small when a single bridge failed on any one phase in inverter [35]. This conclusion is verified to be consistent with the actual inspection on the inverter B-phase fault.

5.4. Computation time comparison between DAE and NN

After analysing performance of DAE fault detection and comparison results between DAE method and standard NN, we further evaluate computation time of two methods. Since their input data matrix size is selectable, the input data matrices of two methods in three cases is respectively selected for data volume of 1 min, 5 min and 10 min. In other words, they calculated one time every 1 min, 5 min and 10 min. The calculation time of two methods is shown in Table 4.

Shown as Table 4, the computation time of DAE is longer than NN under the condition of the same input data volume. Although the calculation time of DAE method is longer than that of NN method, the time is millisecond and can meet practical application in wind turbine fault detection. Once determining proper size of input data matrix, the computation time of DAE does not affect actual fault detection accuracy of turbine components.

Seen from the above three cases, the advantages of DAE method mainly show in two aspects: (1) The learning ability of DAE method

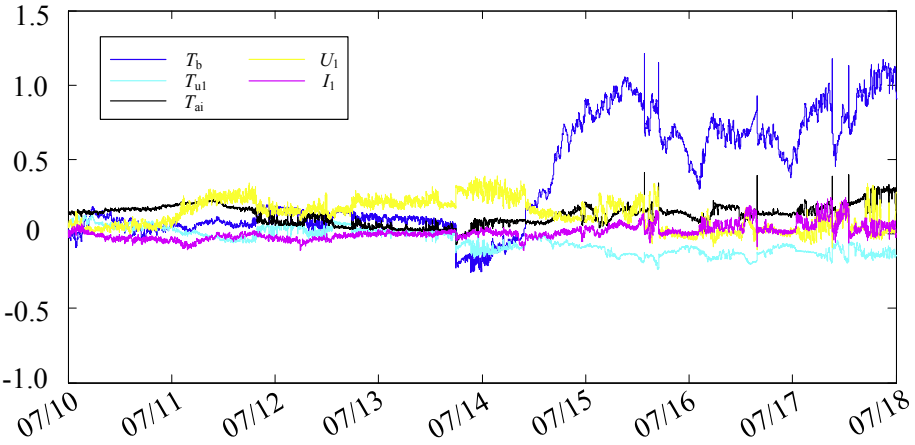

Fig. 7. The residual trend of generator SCADA variables.

Table 3
Fault alarms and processing records for the generator-inverter of the wind turbine.

Data/Time	Wind speed (m/s)	Fault warning	Fault judgement	Repairing	Handling time
2016/6/21 10:15	3.62	Generator 3-phase current asymmetry	Generator side and line side CTs failure	Replace PT and CT of B phase	2016/6/21 10:30 −16:25
2016/6/22 17:20	5.3	Repeatedly occurred generator 3-phase currents asymmetry	Nothing abnormal detected	Reset of warning signal	2016/6/23 8:43 −12:58
2016/6/24 10:05	8.25	Generator 3-phase currents asymmetry	One-side IGBT fault in inverter	Replace the B phase IGBT	2016/6/24 12:50 −15:53

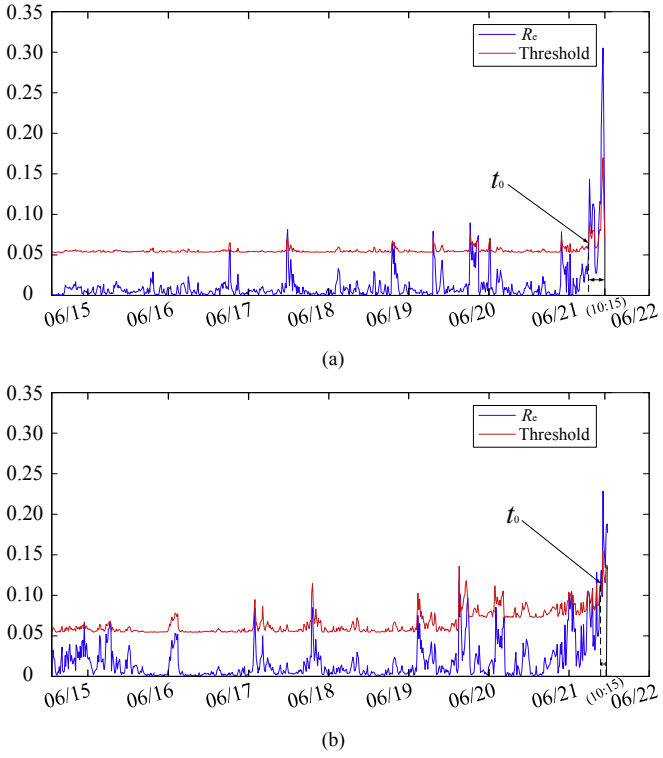

Fig. 8. The control chart of wind turbine generator R_e (a) by DAE network (b) by NN.

is much better than the traditional NN method; (2) the condition index of DAE method has very good sensitivity and accuracy in reflecting condition changes of wind turbine components.

6. Conclusion

This paper presents a DAE deep learning network for component early anomaly detection and fault detection using available SCADA data from wind turbines. The DAE network model is trained by normal operation SCADA data, the reconstruction error and adaptive threshold, and residuals were used for anomaly detection and fault location. The actual case study analysis results confirm the following conclusions:

(1) The validity of the developed DAE network model for early anomaly detection of wind turbine components is verified according that the analysis results based on the DAE model of the generator and the gearbox, which is consistent with the actual recorded fault cases of the wind turbine.
(2) The influence of sharp wind speed fluctuations and external disturbances on the anomaly detection of wind turbine components is reduced by using the adaptive threshold based on the extreme value theory to monitor R_e. The proposed adaptive threshold is capable of avoiding false alarms and giving valid warnings at an early stage, as shown by three fault case studies.
(3) In addition, after the DAE model gives an early warning, the possible fault location of the component can be further determined by analysing the change trends of the SCADA variable residuals of the wind turbine components.

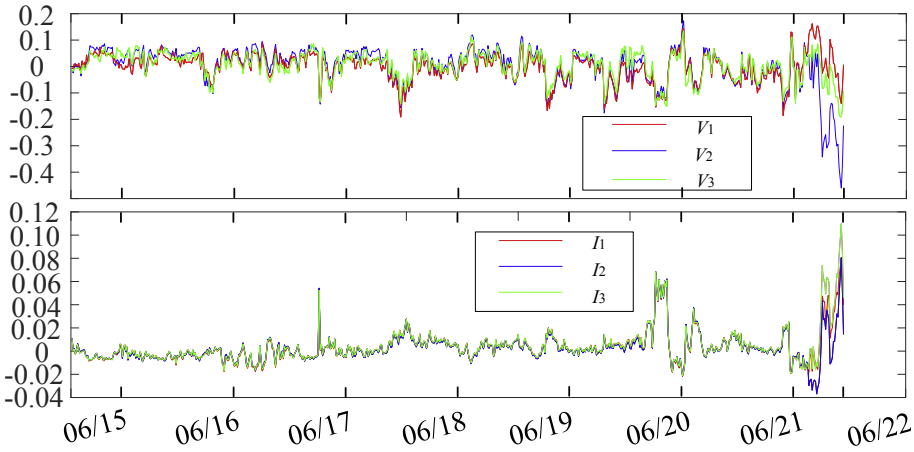

Fig. 9. The residual trend of generator three-phase voltages and three-phase currents.

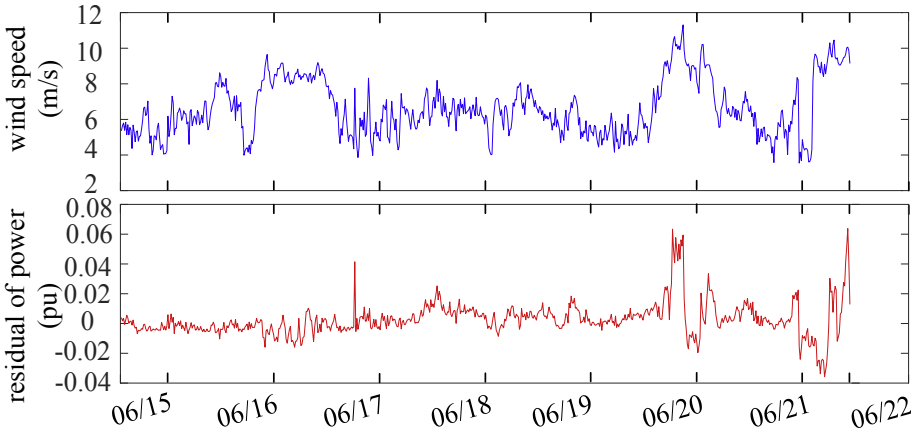

Fig. 10. Wind speed and the residual trend of generator power.

Table 4
Computation time of DAE and NN methods.

Methods	Case A (Seconds)			Case B (Seconds)			Case C (Seconds)		
	1min	5min	10min	1min	5min	10min	1min	5min	10min
DAE	0.0469	0.2138	0.4122	0.0491	0.2006	0.4929	0.0439	0.2183	0.4429
NN	0.0070	0.0314	0.0670	0.0067	0.0341	0.0685	0.0069	0.0334	0.0685

Notes: The 1min, 5min and 10min respectively represents the input data volume of the corresponding time length (SCADA data is sampled every second), the Case A, Case B, and Case C correspond to the three fault cases in Section 5.

Acknowledgements

We thank the funding of National Key Technology Research and Development Program [No. 2015BAA06B03].

References

[1] A. Kusiak, A. Verma, A data-mining approach to monitoring wind turbines, IEEE Trans. Sustain. Energy 3 (1) (2012) 150–157.
[2] H. Zhao, L. Li, Fault diagnosis of wind turbine bearing based on variational mode decomposition and teager energy operator, IET Renew. Power Gener. 11 (4) (2017) 453–460.
[3] I. Hwang, S.W. Kim, Y. Kim, C.E. Seah, A survey of fault detection, isolation, and reconfiguration methods, IEEE Trans. Control Syst. Technol. 18 (3) (2010) 636–653.
[4] P.F. Odgaard, J.A. Stoustrup, Benchmark evaluation of fault tolerant wind turbine control concepts, IEEE Trans. Control Syst. Technol. 23 (3) (2015) 1221–1228.
[5] S. Dey, Pisu, P, B. Ayalew, A comparative study of three fault diagnosis schemes for wind turbines, IEEE Trans. Control Syst. Technol. 23 (5) (2015) 1853–1868.
[6] P. Casau, P. Rosa, S.M. Tabatabaeipour, C. Silverstre, Fault detection and isolation and fault tolerant control of wind turbines using set-valued observers, in: 8th IFAC SAFEPROCESS, August 29-31, 2012. Mexico City, Mexico.
[7] I.V.D. Bessa, R.M. Palhares, M.F.S. Angelo, J.E.C. Filho, Data-driven fault detection and isolation scheme for a wind turbine benchmark, Renew. Energy 87 (2016) 634–645.
[8] N. Laouti, N.S. Othman, S. Othman, Support vector machines for fault detection in wind turbines, in: 18th IFAC World Congress, Milano (Italy) August 28-September 2, 2011.
[9] J.F. Dong, M. Verhaegen, Data driven Fault detection and isolation of a wind turbine benchmark, in: 18th IFAC World Congress, Milano (Italy) August 28-September 2, 2011.
[10] F.P.G. Márquez, A.M. Tobias, J.M.P. Pérez, et al., Condition monitoring of wind turbines: techniques and methods, Renew. Energy 46 (5) (2012) 169–178.
[11] J. Zhang, Extraction of rolling bearing fault feature based on time-wavelet energy spectrum, J. Mech. Eng. 47 (17) (2011) 44.
[12] Q. Huang, D. Jiang, L. Hong, et al., Application of wavelet neural networks on vibration Fault diagnosis for wind turbine Gearbox[C], in: International Symposium on Neural Networks: Advances in Neural Networks, Springer-

Verlag, 2008, pp. 313–320.
[13] Z. Ma, Y. Li, Z. Liu, et al., Rolling bearings' fault feature extraction based on variational mode decomposition and Teager energy operator, J. Vib. Shock 35 (13) (2016) 134–139.
[14] H. Zhao, L. Li, Fault diagnosis of wind turbine bearing based on variational mode decomposition and teager energy operator, IET Renew. Power Gener. 11 (4) (2017) 453–460.
[15] X. Ren, Y. Zhang, Y. Xing, et al., Rolling bearing early fault diagnosis based on angular domain cascade maximum correlation kurtosis deconvolution, Chin. J. Sci. Instrum. 36 (9) (2015) 2104–2111.
[16] Zhu Junda, He David, Jae M. Yoon, Online particle contaminated lubrication oil condition monitoring and remaining useful life prediction for wind turbines, Wind Energy 18 (6) (2015) 1131–1145.
[17] Shawn Sheng, Investigation of Oil Conditioning, Real-Time Monitoring and Oil Sample Analysis for Wind Turbine Gearboxes (Presentation). No. NREL/PR-5000–50301, National Renewable Energy Laboratory (NREL), Golden, CO, 2011.
[18] W. Qiao, D. Lu, A survey on wind turbine condition monitoring and fault Diagnosis—Part II: signals and signal processing methods, IEEE Trans. Ind. Electron. 62 (10) (2015) 6546–6557.
[19] J. Zhu, J.M. Yoon, D. He, et al., Lubrication oil condition monitoring and remaining useful life prediction with particle filtering, Int. J. Prognostics Health Manag. 4 (Sp2) (2013) 1–15.
[20] W. Yang, R. Court, J. Jiang, Wind turbine condition monitoring by the approach of SCADA data analysis, Renew. Energy 53 (9) (2013) 365–376.
[21] Phong B. Dao, et al., Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data, Renew. Energy 116 (2018) 107–122.
[22] K. Kim, G. Parthasarathy, O. Uluyol, et al., Use of SCADA data for failure detection in wind turbines, in: Proceedings of the ASME 2011 5th International Conference on Energy Sustainability, Washington DC, USA, 2011, pp. 2071–2079.
[23] P. Sun, J. Li, Y. Yan, et al., Wind turbine anomaly detection using normal behavior models based on SCADA data[C], in: International Conference on High Voltage Engineering and Application, IEEE, 2014, pp. 1–4.
[24] A. Zaher, S.D.J. Mcarthur, D.G. Infield, et al., Online wind turbine fault detection through automated SCADA data analysis, Wind Energy 12 (6) (2009) 574–593.
[25] Xuejun LI, Ping LI, Lingli JIANG, Class mean kernel principal component analysis and its application in fault diagnosis, J. Mech. Eng. 50 (3) (2014) 123–129.
[26] R. Huang, L. Xi, X. Li, et al., Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods, Mech. Syst. Signal Process. 21 (1) (2007) 193–207.
[27] Lei Yaguo, Jia Feng, Zhou Xin, et al., A deep learning-based method for machinery health monitoring with big data 51 (21) (2015) 49–56.
[28] Y. Chen, X. Zhao, X. Jia, Spectral–spatial classification of hyperspectral data based on deep belief network, IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. 8 (6) (2015) 2381–2392.
[29] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[30] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks[C], in: International Conference on Neural Information Processing Systems, Curran Associates Inc, 2012, pp. 1097–1105.
[31] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (1) (2006) 504–507.
[32] Chen Yushi, Lin Zhouhan, Zhao Xing, et al., Deep learning-based classification of hyperspectral data, Observ. Remote Sens. 7 (6) (2014) 2094–2107.
[33] T. Nakama, Theoretical analysis of batch and on-line training for gradient descent learning in neural networks, Neurocomputing 73 (1) (2009) 151–159.
[34] Daniela Toshkova, et al., Applying Extreme Value Theory for alarm and warning levels setting under variable operating conditions, 2016.
[35] Zhao Hongshan, Cheng Liangliang, Open switch Fault diagnostic method for back to back converters of doubly fed wind power generation system, IEEE Trans. Power Electron. PP (99) (2017) 1, 1.