Article

Condition Monitoring and Fault Diagnosis of Wind Turbines Gearbox Bearing Temperature Based on Kolmogorov-Smirnov Test and Convolutional Neural Network Model

Peng Guo, Jian Fu* and XiYun Yang

School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China; huadiangp@163.com (P.G.); yangxiyun916@ncepu.edu.cn (X.Y.)

* Correspondence: huadianfj@ncepu.edu.cn; Tel.: +86-188-1092-2716

Received: 2 August 2018; Accepted: 22 August 2018; Published: 27 August 2018

Abstract: Wind turbine condition-monitoring and fault diagnosis have important practical value for wind farms to reduce maintenance cost and improve operating level. Due to the special distribution law of the operating parameters of similar turbines, this paper compares the instantaneous operation parameters of four 1.5 MW turbines with strong correlation of a wind farm. The temperature-power distribution of the gearbox bearings is analyzed to find out the main trend of the turbines and the deviations of individual turbine parameters. At the same time, for the huge amount of data caused by the increase of turbines number and monitoring parameters, this paper uses the huge neural network and multi-hidden layer of a convolutional neural network to model historical data. Finally, the rapid warning and judgment of gearbox bearing over-temperature faults proves that the monitoring method is of great significance for large-scale wind farms.

Keywords: wind turbines; condition monitoring; Kolmogorov-Smirnov test; convolutional neural network

1. Introduction

As a distributed energy system, the wind farm's harsh working environment, special geographical location and randomness and volatility brought by generators are the most unique places compared with traditional power generation. The Literature [1] suggested that many wind farms around the world, such as Canada, have been in service for decades. The gearbox is one of the important components of the wind turbine, and its manufacturing technology is relatively mature and has high reliability. Although the failure rate of the gearbox is low, the maintenance process is complicated compared with the electronic control system and hydraulic system with the highest frequency of failure. Especially for offshore wind turbines, the maintenance process requires special equipment such as ships and cranes, and suitable weather. As a result, downtime and maintenance costs caused by gearbox failures are highest among all types of failures. Therefore, it is especially important to choose a good condition-monitoring method.

For the traditional fault diagnosis methods of gearbox bearing, the Literature [2] introduced the improved range overlap method of the fuzzy expert system to improve the accuracy of the classifier. Literature [3] used methods such as trend, clustering, damage modeling and expert system evaluation to analyze fault anomalies. This type of diagnostic method is based on knowledge diagnosis and is more suitable for solving non-numeric problems in the nonlinear domain. Literature [4] proposed qualitative fault tree analysis. The vibration acceleration and current signal analysis used in [5–7] determine the fault by diagnosing the abnormal signal. The hybrid fault decoupling method based on bounded component analysis proposed in [8] is quite effective and representative for solving the

complexity problem. Literatures [9,10] used the PCA (Principal Component Analysis) and MPCA (Multi-way Principal Component Analysis) models to model normal units to evaluate faulty units. The Literature [11] compared the average temperature measurement of a 2 MW unit with several other units, which is convenient for finding short-term anomalies and dangerous peaks, and provides a good idea for lateral comparison between units. The fault diagnosis method proposed by the above literature from different angles has achieved good results in the past practical applications. Based on this, the paper finds the unique distribution of gearbox bearing temperature-power scatter plots, which provides more valuable comparison parameters for comparison between units. At the same time, with the deep application of deep learning ideas in various fields, fault diagnosis and condition-monitoring methods have also ushered in new changes. The BP (Back Propagation) neural network model used in [12] has been not entirely applied to the current training requirements for large data volumes. Literatures [13,14] demonstrated the application of artificial neural networks in process surface roughness and mining. Literature [15] combined convolutional neural networks with other artificial intelligence algorithms and achieved good effect. Literature [16] demonstrated the accuracy and efficiency of deep neural networks in load forecasting. Therefore, the great potential of deep learning ideas in wind turbine monitoring and fault diagnosis is obvious.

On the one hand, the immediate comparison of adjacent units can reflect the deviation of the parameters of a single unit from the overall trend of the wind farm, and observe the parameter changes simply and efficiently. On the other hand, for the processing demand of large data volume, the convolutional neural network has the characteristics of “end-to-end”. That is, the whole process of feature extraction, feature dimension reduction and classification, and huge neurons can be completed through a neural network. The network and multiple hidden layers can quickly and efficiently train data. It can not only check the working condition of the unit during most normal working hours, but also avoid the early warning caused by the accidental parameter change and play the role of condition-monitoring. After the first section of the article introduces the method, horizontal contrast and longitudinal analysis are performed in the second and third sections respectively. Finally, the rapid warning and judgment of gearbox bearing over-temperature faults proves that the monitoring method is of great significance for large-scale wind farms.

2. Condition-Monitoring Method

The condition-monitoring method proposed in this paper is divided into two parts. The structural block diagram is shown in Figure 1.

(1) Horizontal comparison: Before the comparative analysis, this paper selects appropriate wind power plant data from the wind farm data collected by the SCADA system and selects a turbine for comparison with other turbines to analyze the differences. After the data is preprocessed by the grid method, the temperature-power scatter plot of the gear box is plotted. The average value of the scattered points is calculated, and the distribution of the data is tested by KS test. For different parameters of different units in different work environments, the scatter points conform to different data distributions (normal distribution, gamma distribution, wilbur distribution).

(2) Longitudinal analysis: The values of active power, wind speed, ambient temperature, cabin temperature, and gearbox oil temperature at different times are selected as inputs, and the gearbox bearing temperature is used as the output. The model is used to train the known normal data as a sample, and the actual data is selected for verification. After drawing the gearbox temperature-power statistics scatter distribution, this part analyzes the effect of modeling test by comparing the measured data. In the case of condition-monitoring, the unit operating conditions are analyzed by combining the predicted results with the horizontal comparison of other units.

Figure 1. Block diagram of the overall structure of the paper.

3. Horizontal Contrast Analysis

3.1. Grid Data Preprocessing

At first, after using the Pearson correlation coefficient formula to calculate the fan correlation, the four units E17, C12, E16 and E18 with the strongest correlation are selected for lateral comparison. Among them, x_i , y_i , and $\bar{x}$, $\bar{y}$ are the sample value and mean value of the reference unit and the control unit respectively. Figure 2 shows the distribution map of the wind farm turbines.

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\delta_x} \right) \left(\frac{y_i - \bar{y}}{\delta_y} \right) \quad (1)$$

Select the point where the power is greater than 0 from the raw data of the four units, and set the active power interval to $[P_{min}, P_{max}]$. In this paper, for data on shutdown periods, when drawing the gearbox bearing temperature–power scatter plot, the power points less than 0 have been filtered out. In the drawn gearbox temperature–power scatter plot, the power band is divided into $PNum$ intervals by the power band at intervals of 40 kW. The reason for selecting the split power band here is that the change is continuously and easily observed in (0, 1600) kW, while the temperature band is concentrated only in a densely small range. Moreover, the variance of the power data on the same temperature band is too large, which is not convenient to the continuity of data observation. The rated power $PRated$ is 1600 kW.

$$PNum = \frac{PRated + 100}{40} = 40 \quad (2)$$

After counting the number of scattered dots in $XBoxNumber$, the data is deposited and the power band is aggregated with each layer, and the control chart is used to observe the range between the different units. The control chart contains the control limit UCL , the average difference $MRange$, the total mean M , the control center line CL , and the control lower limit LCL .

Figure 2. Wind farm unit distribution satellite map.

In this paper, the unit capacity is chosen to be 4, and the maximum difference value in each subgroup is *Range*. x_j is the total number of power strip scatter points for each of the four units ($j = 1, 2, 3, 4$).

Average difference *MRange*.

$$MRange = \frac{Max(x_j) - Min(x_j)}{40} \quad (3)$$

$$Range = Max(x_j) - Min(x_j) \quad (4)$$

Control limit *UCL*.

$$UCL = M + 0.729MRange \quad (5)$$

Control lower limit *LCL*.

$$LCL = M - 0.729MRange \quad (6)$$

For the mean $\bar{x}_n$ and variance σ_n of the scatter points for each row of the power band: $\bar{x}_n$ is the statistical mean of the power scatter points for the n row; k is the total number of data for each segment (100), and $x_{i,n}$ is the number of scattered points.

The mean and variance of the scatter points of each power band are $\bar{x}_n$ and σ_n respectively. n is the number of segments corresponding to the y-axis (40), and $x_{i,n}$ is the number of scatter points corresponding to the grid power ($i = 1, 2, 3 \dots 40; n = 1, 2, 3, 4$).

$$\bar{x}_n = \frac{1}{k} \sum_{i=1}^k x_{i,n} \quad (7)$$

$$\sigma_n = \sqrt{\frac{1}{k} \sum_{i=1}^k (x_{i,n} - \bar{x}_n)^2} \quad (8)$$

3.2. Kolmogorov-Smirnov Check Scatter Distribution

The Kolmogorov-Smirnov test is a non-parametric test in which successive one-dimensional probability distributions are equal and it is used to compare sample and reference probability

distributions. The Kolmogorov-Smirnov statistic quantifies the distance between the empirical distribution function of a sample and the cumulative distribution function of a reference distribution or the empirical distribution function of two samples.

This paper proposes hypotheses for the distribution of scatter, and uses $F_n(x)$ to indicate the sample size as a cumulative distribution function of the observations of n samples, and $F_n(x) = i/n$ ($i = 1, 2, \dots, n$). $F_n(x)$ represents the cumulative probability distribution function of the theoretical distribution. The cumulative distribution function $F_n(x)$ and the theoretical distribution function $F(x)$ are compared to test the goodness of fit test. The test statistic D_n is the maximum deviation between $F_n(x)$ and $F(x)$: where n is the number of samples and a is the level of significance. $D(n, a)$ is a standard value corresponding to a certain sample and significance level. In this paper, a is 0.05 and n is 40. Table 1 is the partial value of $D(n, a)$.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{[-\infty, x]}(X_i) \quad (9)$$

$I_{[-\infty, x]}(X_i)$ is the index function and if $X_i < x$, it equals 1 otherwise it is equal to 0.

Table 1. Partial value of $D(n, a)$.

Level of Significance(a)				
<i>n</i>	0.40	0.20	0.10	0.05
5	0.369	0.447	0.509	0.562
10	0.268	0.322	0.368	0.409
20	0.192	0.232	0.264	0.294

Given the cumulative distribution function of the Kolmogorov-Smirnov statistic is

$$D_n = \max |F(x) - F_n(x)| \quad (10)$$

If the sample comes from the distribution $F(x)$, D_n will converge to 0. The critical value of a single-sample K-S test is determined based on a given significance level of 0.05 and the number of sample data of 40. If $D_{max} < D_a$, then at the significance level of a , the accepted sample conforms to the distribution.

After statistical analysis of the scattered data used in this paper, it is concluded that the model generally conforms to the continuous probability function of the gamma distribution. Parameters K in the gamma distribution is called shape parameters and θ is called scale parameter.

The variance and mean are shown in the following formulas.

$$E(x) = K\theta \quad (11)$$

$$Var(x) = K\theta^2 \quad (12)$$

The feature function is

$$\ell(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha} \quad (13)$$

4. Longitudinal Analysis: CNN Model for Gearbox Bearing Temperature Forecast

The above-mentioned horizontal comparison process combines the operating conditions of multiple units, which can effectively avoid the abnormal alarm caused by the active power change of the wind turbine caused by accidental reasons. For example, in the weather with high wind speed, the power generation capacity of the unit is generally higher than before. By comparing the operating parameters between the units, the cause of the alarm can be identified. However, the horizontal

comparison can only roughly judge whether the current state is abnormal through Real-time data, and the fault warning needs to be implemented in combination with an effective modeling method.

This chapter introduces the idea of deep learning into the neural network by modeling the single unit and using historical data as a training sample. Through the convolution operation to extract different levels of features from shallow to deep, the neural network training process is used to make the whole network automatically adjust the parameters of the convolution kernel, so that the most suitable classification features are generated unsupervised, and the above process is assisted in state judgment. While training the historical normal operation data and the reliable data of the relevant units, this model can predict the scatter distribution of the unit components in a short period of time, which can play a good role.

4.1. Convolutional Neural Network Structure

To establish the gearbox temperature CNN model, it is necessary to determine the modeling variables in the training sample, which are closely related to the gearbox bearing temperature. Through the analysis of 40 parameters of 1-min class recorded by the SCADA system, 5 variables of active power, wind speed, ambient temperature, cabin temperature and gearbox oil temperature closely related to bearing temperature are selected as inputs, and the gearbox bearing temperature is used as the output.

For the input of the convolutional neural network in this paper, the method of selecting variables is based on the Expert Evaluation method and Partial Least Squares (PLS) method. The wealth of experience and breadth of knowledge of experts provides a reference value for the initial selection of variables. The PLS method is an applied statistical method that integrates the basic functions of multivariate statistical regression, canonical correlation analysis and Principal Component Analysis (PCA). This method can select the combination of independent variables with the largest correlation with the dependent variable, and solve the multiple correlation of variables in the regression analysis. The extracted principal components are well interpreted and can eliminate correlation between multiple variables. As shown below:

(1) Active power (P): It is closely related to the gearbox temperature. When the output power of the unit is high, the load on the gearbox is large, resulting in high gearbox temperature.

(2) Wind speed (u): Wind is the source of energy for wind turbines. For the variable speed wind turbines studied in this paper, the speed of the transmission system is proportional to the wind speed for the purpose of achieving the best tip speed ratio for maximum wind energy tracking. The higher the wind speed, the higher the gearbox speed, which causes the gearbox temperature to rise.

(3) Ambient temperature (T): Since the ambient temperature of the unit varies greatly in the short-term (day and night) and long-term (week, month) time scales, the ambient temperature must be considered as one of the factors. Especially in the spring of March and April, due to the wind and cold, the ambient temperature difference can reach 30 °C. At different times, even if the power and wind speed of the unit are the same, the temperature of the gearbox will vary greatly due to the difference in ambient temperature.

(4) Cabin temperature (T1): Due to uneven distribution of temperature field inside the wind turbine cabin, it will cause the main heat source inside the engine room (such as generator, gear box, etc.) to stop due to over temperature alarm.

(5) Gearbox oil temperature (T2): The oil temperature of the gearbox is too high, which will cause the fan to stop, and has a serious impact on the power generation.

At the same time, because the gearbox bearing temperature parameter changes have a large inertia, the gearbox temperature at the previous moment has a direct impact on the current temperature. Thus, six adjacent historical moments are selected as a set of samples. The training samples included a total of 14,000 samples from May 1 to June 30, each of which is a 6×6 vector. The vector contains 6 sets of data of 6 variables adjacent to each other at the historical moment. Since the convolutional neural network is a multi-level neural network, it includes a filtering level and a classification level. The filtering

stage is used to extract the characteristics of the input signal, and the classification level classifies the learned features, and the two-level network parameters are jointly trained. The filtering stage consists of three basic units: a convolutional layer, a pooled layer and an active layer. The classification level is generally composed of a fully connected layer.

(1) Convolution layer

The convolutional layer uses a convolutional kernel to perform convolution on the local area of the input signal (or feature) and produces corresponding features. The most important feature is weight sharing, in which the same convolution kernel traverses input once in fixed steps. Weight sharing reduces the network parameters of the convolutional layer, and avoids overfitting due to excessive parameters, and reduces system memory requirements. In practice, correlation operations are mostly used instead of convolution operations to avoid flipping the convolution kernel when back propagation happens. The concrete convolutional layer operation is shown in Equation (14).

$$y^{l(i,j)} = K_i^l * x^{l(x^j)} = \sum_{j'=0}^{w-1} K_i^{l(j')} x^{l(j+j')} \quad (14)$$

$K_i^{l(j')}$ —the j' weight of the i convolution kernel of layer l

$x^{l(x^j)}$ —the j th convolved local area in layer l

w —the width of the convolution kernel

When one-dimensional convolutional layer operations are performed, each convolution kernel traverses the convolutional layer once and performs convolution operation at the same time. Take the first convolution kernel as an example. In the convolution operation, the nucleus multiplies the coefficients corresponding to the neurons in the volume area, then moves the convolution kernel in steps of 1 and repeats the previous operation until the volume. The nucleus traverses all areas of the input signal.

(2) Activation layer

After the active layer is convolved, the activation function will nonlinearly transform the logits value of each convolution output. The purpose of the activation function is to map the linearly inseparable multidimensional features to another space. Since the ReLU (Rectified Linear Unit) function always has a derivative value of 1 when the input value is greater than 0, the gradient diffusion phenomenon is easy to overcome. Therefore, ReLU is used as the activation function of the convolutional neural network. The ReLU function is as follows.

$$a^{l(i,j)} = f(y^{l(i,j)}) = \max\{0, y^{l(i,j)}\} \quad (15)$$

$a^{l(i,j)}$ —the activation value of the convolutional output $y^{l(i,j)}$

(3) Pooling layer

The pooling operation downsamples the width and depth of the original feature to the output by adjusting the size and step size. The downsampling operation used in this paper is a maximum pooling. It takes the maximum value in the sensing domain as an output, and the advantage of that is it can obtain position-independent features that are critical for periodic time-domain signals.

$$p^{l(i,j)} = \max_{(j-1)W+1 \leq t \leq jW} \{a^{l(i,t)}\} \quad (16)$$

$a^{l(i,t)}$ —the activation value of the t neuron in the i layer of the l layer

$p^{l(i,j)}$ —the width of the pooled area

(4) Fully connected layer

The fully connected layer classifies the features extracted by the filter stage. The specific approach is to first spread the output of the last pooling layer into a one-dimensional feature vector as the input of the full-connected layer; then to fully connect the input and output, in which ReLU is used as the activation function of hidden layer. The last output layer's activation function is Softmax.

$$z^{l+1(j)} = \sum_{i=1}^n W_{ij}^l a^{l(i)} + b_j^l \quad (17)$$

W_{ij}^l —the weight between the i neuron in the l layer and the j neuron in the $l + 1$ layer

b_j^l —the bias value of the j neuron in the $(l + 1)$ layer for all neurons in the l layer

(5) Objective function

The output of the segment input signal in the neural network should be consistent with its target value. The function to evaluate this consistency is called the objective function of the neural network. The commonly used objective function has a squared error function and a cross entropy loss function. The actual output of the convolutional neural network has a Softmax value of q , and its target distribution p is a one-hot type vector. That is, when the target category is j , $p^j = 1$, otherwise 0.

$$L = \frac{1}{m} \sum_{k=1}^m \frac{1}{2} \sum_j (p_k^j - q_k^j)^2 \quad (18)$$

$$L = -\frac{1}{m} \sum_{k=1}^m \sum_j p_k^j \log q_k^j \quad (19)$$

Compared with the squared error function, the cross entropy function measures the consistency of the two probability distributions. The cross entropy function is often regarded as the negative log likelihood of the Softmax distribution in machine learning, so the cross entropy function is used as the objective function in this paper. The network consists of two convolutional layers, two pooling layers, a fully connected hidden layer, and a Softmax layer.

4.2. Convolutional Neural Network Error Back Propagation

Error backpropagation is a key step in weight optimization for neural networks. The main method is to solve the derivative function of the objective function with respect to the last layer of neurons. Through the chain rule, the derivative value of the objective function with respect to the ownership value is calculated layer by layer from the back to the front.

(1) Full connection layer reverse derivation

First, calculate the derivative of the objective function L about the last level of the logits value $z^{l+1(j)}$.

$$\frac{\partial L}{\partial z^{l+1(j)}} = \sum_{k=1}^m p_k^j q_k^j - p_k^j \quad (20)$$

Then calculate the derivative W_{ij}^l of the objective function L of the fully connected layer and the derivative of the offset b_j^l .

$$\frac{\partial L}{\partial W_{ij}^l} = \frac{\partial L}{\partial z^{l+1(j)}} \cdot \frac{\partial z^{l+1(j)}}{\partial W_{ij}^l} = \frac{\partial L}{\partial z^{l+1(j)}} \cdot a^{l(i)} \quad (21)$$

Finally, the objective function L is calculated. The activation value $a^{l(i)}$ of the fully connected hidden layer whose activation function is ReLU and the derivative of the logits value $z^{l(j)}$.

$$\frac{\partial L}{\partial a^{l(i)}} = \sum_j \frac{\partial L}{\partial z^{l+1(j)}} \cdot \frac{\partial z^{l+1(j)}}{\partial a^{l(i)}} = \sum_j \frac{\partial L}{\partial z^{l+1(j)}} \cdot W_{ij}^l \quad (22)$$

$$\frac{\partial L}{\partial z^{l(i)}} = \frac{\partial L}{\partial a^{l(i)}} \cdot \frac{\partial a^{l(i)}}{\partial z^{l(i)}} \quad (23)$$

(2) Pooled layer reverse derivation

After the value is calculated by the above formula, the objective function L is continuously solved for the derivative of the weight W_{ij}^{l-1} of the fully connected hidden layer and the bias term b_j^{l-1} . When the objective function L is solved for the logits value of the fully connected layer and the derivative of the weight, then the derivative of each parameter of the pooling layer is calculated. Since the pooling layer has no weight, it is only necessary to calculate the derivative of the input neurons of the pooling layer.

The specific practice for pooling the maximum value is: record the maximum position of the pooled area during forward propagation. When $t = t_m$, $\max_{(j-1)w+1 \leq t \leq jw} \{a^{l(i,t)}\} = a^{l(i,t_m)}$. In the case of backpropagation, the derivative value is passed to the t_m neuron, and the other neurons do not participate in the transmission, which the derivative is zero.

$$\frac{\partial L}{\partial a^{l(i,t)}} = \frac{\partial L}{\partial p^{l(i,j)}} \cdot \frac{\partial p^{l(i,j)}}{\partial a^{l(i,t)}} = \frac{\partial L}{\partial p^{l(i,j)}} (t = t_m) \quad (24)$$

(3) Convolutional layer reverse derivation

First, calculate the derivative of L 's logits value for each convolutional layer, since the convolutional layer uses the ReLU activation function.

$$\frac{\partial L}{\partial y^{l(i,j)}} = \frac{\partial L}{\partial a^{l(i,j)}} \cdot \frac{\partial a^{l(i,j)}}{\partial y^{l(i,j)}} = \frac{\partial L}{\partial a^{l(i,j)}} (y^{l(i,j)} > 0) \quad (25)$$

Next, calculate the derivative of L about the input value $x^{l(j)}$ of the convolutional layer.

$$\frac{\partial L}{\partial x^{l(j)}} = \sum_i \frac{\partial L}{\partial y^{l(i,j)}} \cdot \frac{\partial y^{l(i,j)}}{\partial x^{l(j)}} = \sum_i \frac{\partial L}{\partial y^{l(i,j)}} \cdot \sum_{j'=0}^{W-1} K_i^l(j') \quad (26)$$

Finally, calculate the derivative of L on the convolution kernel $K_i^l(j')$.

$$\frac{\partial L}{\partial K_i^l(j')} = \frac{\partial L}{\partial y^{l(i,j)}} \cdot \frac{\partial y^{l(i,j)}}{\partial K_i^l(j')} = \frac{\partial L}{\partial y^{l(i,j)}} \cdot \sum_j x^{l(j)} \quad (27)$$

(4) Adam optimization algorithm

After calculating the derivative of each weight of the objective function by using the error back propagation algorithm, the next step is to use the optimization algorithm to update the weight. Solve the optimal weight, so that the value of the objective function is minimized. This process is described by the following formula:

$$\theta^* = \arg \min_{\theta} L(f(x^i; \theta)) \quad (28)$$

$L(), f()$ —objective function value and output value

θ —all parameters of the convolutional neural network

θ^* —the optimal parameter of the convolutional neural network

x^i —input of the convolutional neural network

For shallow neural networks, the use of stochastic gradient-decreasing (SGD), which is widely used in BP neural networks, can converge to the global. However, for the deep convolutional neural network proposed in this chapter, due to the large number of parameters and hyperparameters, if the superparameter selection is not good, the use of SGD training tends to fall into the local best. Therefore, the Adam algorithm is used in this paper. Adam is a learning rate adaptive optimization algorithm that dynamically adjusts the learning rate of each parameter by using the first moment estimation and the second moment estimation of the gradient. Adam's advantage lies in the correction of the first moment and non-central second moment of the initialization from the origin after offset correction, so that each iteration learning rate has a certain range. Adam is generally robust to the selection of multiple parameters and therefore helps in parameter adjustment of the neural network.

5. Cases Analysis

5.1. Gearbox Shaft Temperature Normal Operation Analysis

In order to verify the effectiveness of the condition-monitoring method, this paper takes the gearbox of four 1.5 MW units of a wind farm as the research object, and selects the contents recorded by the SCADA system at one-minute level in June and July 2013, which includes time and active power, wind speed, environment and cabin temperature, gearbox bearing temperature, generator speed, transmission chain vibration acceleration and other parameters. At the same time, the SCADA system also saves the operating status information of the unit, such as unit start-up, shutdown, generator over-temperature, and pitch system faults.

According to Pearson's correlation coefficient method, the correlation coefficient values of E16 fan output and C12, E17, E18 output are 0.9483, 0.9526, and 0.9646 respectively, which indicates that the correlation between E16 and other three fans is higher than the standard value. It indicates that the temperature of the gearbox of each unit is normal. Figure 3 is a schematic diagram of the temperature-power dispersion of the gearbox. After the pretreatment, the grid method is used to count the distribution of scattered points.

Figure 3. Gearbox bearing temperature–power scatter plot.

The control chart of Figure 4 shows the distribution of the mean values of the four selected units. Due to the strong correlation, the internal residuals of the unit sample are evenly distributed, and the residual distribution is far less than the control lower limit, indicating that the unit output meets the historical characteristics and E16 output is normal.

Figure 4. Horizontal comparison control chart of four turbines.

After statistical pretreatment of scatters by the grid method, the KS test sample is selected. Figure 5 shows that the gearbox temperature data conforms to the gamma distribution. The distribution parameters are normal unit parameters $K = 2.9943$, $\theta = 1.0428$.

Figure 5. Schematic diagram of scatter gamma distribution of normal turbine.

The KS test is performed on the C12, E17, and E18 units, and the distribution parameters of the control units are calculated. The values are fluctuating at (3.0000 ± 0.2000) and the values are fluctuating at (1.0000 ± 0.1000) . Figure 6 is a lateral contrast gamma distribution graph. Compared with the other three experimental units, the trend and parameters are approximately the same within the acceptance range. Therefore, it is concluded that the measured data between the units with strong correlation are highly comparable. When monitoring the operation status of the unit at a later stage,

the comparative analysis of online data of similar units can be selected to see whether the operation status of the unit is normal.

Figure 6. Schematic diagram of gamma distribution of control turbines.

At the same time, historical data is substituted into the CNN model for modeling prediction. The training samples included a total of 10,000 samples from June 1 to July 10, each of which is a 6×6 vector. And the data is converted into the corresponding format for normalization. The verification sample is 2400 data from July 11 to July 20. The learning rate is 0.01 and the accuracy is 0.001. Each time 50 samples are randomly selected, and the total is iterated 10,000 times. The above sample processing method uses the idea that the CNN model divides the image into multi-dimensional arrays and random sample selection for image processing. After training and verifying that the correct model parameters are obtained, the data for the next 10 days is predicted. According to the predicted results, the gearbox temperature–power scatter plot is drawn, and the number of power strips per segment is counted. As shown in Figure 7, the prediction effect is basically consistent with the actual sample.

Figure 7. Comparison of convolutional neural network model prediction data and measured data.

5.2. Gearbox Shaft Temperature Overtemperature Fault Diagnosis

It is known that starting from August 5 of that year, the gearbox temperature of E16 is higher than that of P13.8 (80 °C). Since the wind turbine bearing temperature is over temperature, the wind turbine will limit the power further to lower the temperature before the rated power. Therefore, compared with the normal unit, the power point of the over-temperature unit will be more concentrated in a certain area of the high temperature section. And the power change is relatively small. Relatively speaking, under the same amount of observation data, the data distribution of the low temperature section is reduced due to the increase in the distribution of the high temperature section. This can be clearly displayed from the statistic number of the scatter and the range of the control chart. Figure 8 is a faulty gearbox bearing and test site.

Figure 8. Fault gearbox bearing and test picture.

Figure 9 is a control diagram for lateral comparison of other units. Obviously, the measured shaft temperature of the normal E16 unit is approximately the same as that of other units. The control chart for August shows that the difference between the E16 unit and the other three units is much higher than the average difference. The high temperature section has a large difference, and the residual distribution fluctuates sharply.

Figure 9. Horizontal comparison control chart of four turbines.

Continue to do KS test after data preprocessing at the beginning of August. This part observes Figure 10 to get the group of data $K = 5.9824$, $\theta = 0.7256$, which is much larger than $K = 3.0770$, $\theta = 1.0947$ for three normal units in the same period. Combined with the above model parameters obtained from the June and July data training, the input variable data collected from August 1st to 3rd is substituted

into the model to predict the gearbox temperature for the next 10 days. The predicted results obtained by the data are shown in Figure 11. The figure shows the measured and predicted data of the abnormal unit, as well as the data of the normal unit, which reflects the accuracy of the prediction. Compared with the normal unit, it is found that the residuals of the abnormal unit and the normal unit are large and the fluctuation is strong. Moreover, since the power is limited before reaching the rated power, it is mainly concentrated in the 20–30th points, thereby judging the overheating fault of the gearbox bearing. Obviously it shows the role of early warning of failure. The actual records of the wind turbine fault log in Table 2 prove the validity and reliability of the condition-monitoring method in this paper.

Figure 10. Schematic diagram of gamma distribution of E16 and control turbine.

Figure 11. Comparison of convolutional neural network model prediction data and measured data.

For Table 2, fault logs 78 and 97 represent the temperature measurement points recorded on low-speed end and high-speed end of the gearbox bearing respectively. The braking step depends on the level at which the fault occurred. Due to the over-temperature fault of the bearing, the unit is restarted, so the brake level is 2. P13.8 represents the threshold of the low-speed end temperature measurement point of the gearbox bearing. P13.9 represents the threshold of the high-speed end temperature measurement point of the gearbox bearing. Considering that the friction at the high-speed

end end is large, the threshold temperature is higher than that of the low-speed end. When the bearing temperature is 15 °C lower than the threshold temperature, the turbine restarts.

Table 2. Overtemperature fault log of gearbox bearing.

Number	Description	Break Step	Fault i = Identification	Reset Temperature
78	Gearbox bearing over temperature	2	Higher than P13.8 (80 °C)	Below P13.8–15 °C
97	Gearbox bearing over temperature	2	Higher than P13.9 (95 °C)	Below P13.9–15 °C

6. Conclusions

Based on the gearbox bearing temperature, this paper uses the horizontal contrast and longitudinal analysis to achieve wind turbine condition-monitoring. In this paper, KS method is used to test the temperature–power scatter distribution of the gearbox of similar units, and the convolutional neural Network is used to construct the prediction model to analyze the over-temperature of the gearbox bearing. The following is a summary and outlook on the work of the thesis:

(1) The turbine can be monitored in real time, and the abnormal condition of the wind turbine can be found in advance.

(2) For the accidental failure of a single turbine, the fault is found by comparing the operating parameters of the relevant turbine.

(3) In the seasons with different wind speeds, the fluctuations in the parameters of the turbine can also be well reflected.

(4) Due to the efficient and accurate processing of large amounts of data by convolutional neural networks, the application prospects of this method is good.

(5) This method can detect the hidden troubles early, and make precaution, which can improve the reliability of the operation of the wind turbine and reduce the maintenance cost.

The combination of lateral contrast and longitudinal modeling can complement each other. For example, horizontal comparison can be analyzed by real-time data in a short time. But in the case of short-term failure warning, it is necessary to combine historical data modeling prediction. At the same time, in the event of an emergency or due to the general fluctuations of the wind farm unit parameters caused by weather changes, the comparison between the units can eliminate the fault alarm. In summary, the method proposed in this paper is of great significance for the monitoring of the operation condition of large wind farm turbines.

Author Contributions: The study idea, plan and design were conceived by P.G.; Calculations and analysis were carried out by J.F. and X.Y.

Funding: Project Supported by National Natural Science Foundation of China (51677067).

Acknowledgments: The authors would like to thank the cooperation company whose wind turbine data were used for this study. We are also grateful to the anonymous reviewers for their insightful comments, which helped improve the overall quality of this publication.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural network
KS	Kolmogorov-Smirnov
Adam	Adaptive moments
SGD	Stochastic gradient-decreasing
BP	Back propagation
SCADA	Supervisory control and data acquisition

References

1. Cambron, P.; Masson, C.; Tahan, A.; Pelletier, F. Control chart monitoring of wind turbine generators using the statistical inertia of a wind farm average. *Renew. Energy* **2016**, *116*, 88–98. [[CrossRef](#)]
2. Berredjem, T.; Benidir, M. Bearing faults diagnosis using fuzzy expert system relying on an improved range overlaps and similarity method. *Expert Syst. Appl.* **2018**, *108*, 134–142. [[CrossRef](#)]
3. Tautz-Weinert, J.; Watson, S.J. Using SCADA data for wind turbine condition-monitoring—A review. *IET Renew. Power Gener.* **2017**, *11*, 382–394. [[CrossRef](#)]
4. Márquez, F.P.G.; Tobias, A.M.; Pérez, J.M.P.; Papaelias, M.P. Condition-monitoring of wind turbines: Techniques and methods. *Renew. Energy* **2012**, *46*, 169–178. [[CrossRef](#)]
5. Hossain, M.L.; Abusiada, A.; Muyeen, S.M. Methods for advanced wind turbine condition-monitoring and early diagnosis: A literature review. *Energies* **2018**, *11*, 1309. [[CrossRef](#)]
6. Saucedo-Dorantes, J.J.; Delgado-Prieto, M.; Ortega-Redondo, A.J.; Osornio-Rios, A.R.; Romero-Troncoso, R. Multiple-fault detection methodology based on vibration and current analysis applied to bearings in induction motors and gearboxes on the kinematic chain. *Shock Vib.* **2016**, *2016*, 5467643. [[CrossRef](#)]
7. Khan, A.S.; Kim, J.M. Automated Bearing Fault Diagnosis Using 2D Analysis of Vibration Acceleration Signals under Variable Speed Conditions. *Shock Vib.* **2016**, *2016*, 8729572. [[CrossRef](#)]
8. Li, Z.; Jiang, Y.; Hu, C.; Peng, Z. Recent progress on decoupling diagnosis of hybrid failures in gear transmission systems using vibration sensor signal: A review. *Measurement* **2016**, *90*, 4–19. [[CrossRef](#)]
9. Pozo, F.; Vidal, Y. Wind turbine fault detection through principal component analysis and statistical hypothesis testing. *Energies* **2016**, *9*, 3. [[CrossRef](#)]
10. Pozo, F.; Vidal, Y.; Salgado, Ó. Wind Turbine condition-monitoring Strategy through Multiway PCA and Multivariate Inference. *Energies* **2018**, *11*, 749. [[CrossRef](#)]
11. Astolfi, D.; Castellani, F.; Terzi, L. Fault prevention and diagnosis through SCADA temperature data analysis of an onshore wind farm. *Diagnostyka* **2014**, *15*, 71–78.
12. Ren, Y.; Qu, F.; Liu, J.; Feng, J.; Li, X. A universal modeling approach for wind turbine condition-monitoring based on SCADA data. In Proceedings of the 2017 6th Data Driven Control and Learning Systems (DDCLS), Chongqing, China, 26–27 May 2017.
13. Ganovska, B.; Molitoris, M.; Hosovsky, A.; Pitel, J.; Krolczyk, J.B.; Ruggiero, A.; Krolczyk, G.M.; Hloch, S. Design of the model for the on-line control of the AWJ technology based on neural networks. *Indian J. Eng. Mater. Sci.* **2016**, *23*, 279–287.
14. Tadeusiewicz, R. Neural networks in mining sciences—General overview and some representative examples. *Arch. Min. Sci.* **2015**, *60*, 971–984. [[CrossRef](#)]
15. Li, Y.; Huang, Y.; Zhang, M. Short-term load forecasting for electric vehicle charging station based on niche immunity lion algorithm and convolutional neural network. *Energies* **2018**, *11*, 1253. [[CrossRef](#)]
16. Ryu, S.; Noh, J.; Kim, H. Deep neural network based demand side short term load forecasting. *Energies* **2016**, *10*, 3. [[CrossRef](#)]

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).