Research Article

Abnormal Detection of Wind Turbine Based on SCADA Data Mining

Liang Tao , Qian Siqi, Yingjuan Zhang, and Huan Shi

Hebei University of Technology, College of Artificial Intelligence and Data Science, Tianjin, China

Correspondence should be addressed to Liang Tao; 54008214@qq.com

Received 29 March 2019; Revised 20 June 2019; Accepted 25 June 2019; Published 7 August 2019

Guest Editor: Michael Z. Q. Chen

Copyright © 2019 Liang Tao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to reduce the curse of dimensionality of massive data from SCADA (Supervisory Control and Data Acquisition) system and remove data redundancy, the grey correlation algorithm is used to extract the eigenvectors of monitoring data. The eigenvectors are used as input vectors and the monitoring variables related to the unit state as output vectors. The genetic algorithm and cross validation method are used to optimize the parameters of the support vector regression (SVR) model. A high precision prediction is carried out, and a reasonable threshold is set up to alarm the fault. The condition monitoring of the wind turbine is realized. The effectiveness of the method is verified by using the actual fault data of a wind farm.

1. Introduction

SCADA system changes the operation mode of wind farm systems with healthy working environment and reduces the costs of operation and maintenance. However, a large set of high dimensions and many types of data are not fully utilized or developed; only stay on real-time data and historical data reporting statistics are typically monitored or gathered. Therefore, it is important to make full use of the data collected by wind power centralized control center to collect the data of the massive wind turbines, conduct state monitoring of the turbines, and predict their condition and life [1, 2]. Several surveys of WT failures have been conducted in the last two decades to identify failure rates and associated downtime for different subassemblies. However, the different taxonomies used by different turbine manufacturers, wind farm operators, and researchers make comparisons between these surveys challenging. The evaluation of 15 years of data from the German “250 MW Wind” programme [3] and >95% of all the turbines operating between 1997 and 2005 in Sweden [4] gave first insights into the reliability of the first onshore WTs. The German turbines had an average availability of about 98%. An average failure rate of 0.4 failures per turbine per year resulted in an average downtime of 130 hours per turbine per year for the Swedish turbines. A distinctive difference between failure

rate and downtime distribution in subassembly groups was identified. The electrical and electronic control systems were identified as the most failure-prone, but gearbox and generator failures caused the longest downtime.

Many scholars have researched large power wind turbine monitoring and fault diagnoses [5], based on statistical learning method to detect abnormal situations through the wind turbine response model of the weighted least squares support vector-based wind power generator and external regression conditions [6]. The results show that the model is better than conventional forecasting methods. Pandit and Infield [7] used an in-depth analysis of commonly used stationary covariance functions in which wind turbine power curve was used, where GP-based power curve was constructed using different stationary covariance functions, and after that, a comparative analysis was carried out in order to identify the most effective covariance function. The commonly used squared exponential covariance function is taken as the benchmark, against which other covariance functions are assessed. The results show that the performance (in terms of model accuracy and uncertainty) of GP fitted power curve models based on rational quadratic covariance functions is almost the same as for the most commonly used squared exponential function. The studies of Astolfi et al. [8, 9] are a catalog of generalizable methods

for studying wind turbine power curve upgrades. In particular, from the study of the selected test cases, it arises that complex wind conditions might affect wind turbine operation such that the production improvement is non-negligibly different from what can be estimated under the hypothesis of ideal wind conditions. Wan et al. [10] proposed wind form using wavelet based on energy function for asymmetrical fault detection in doubly fed induction generator. The proposed method not only detects the fault within one and half cycle of fundamental wave but also reveals the effectiveness under time-varying conditions. Turbine condition monitoring (TCM) through vibration analysis has pros and cons: basically high diagnostic power against high cost and high complexity for elaborating the information [11] from the data stream into knowledge. Chun-yu et al. [12] put forward a dynamic prediction model of wind turbine blade failure based on the grey theory. The relative error between prediction and field investigation data is less than 5%, meeting the actual needs of engineering and verifying the effectiveness and applicability of the proposed algorithm. The main contribution of Chakkor et al. [13] is designing an intelligent wireless remote monitoring and control system according to features and requirements of wind turbines. This system based on IP communication combines Web and database client/server technology to copy data measurements received from the different sensors installed in the wind turbine machines. Eggers et al. [14] used *Hotelling T2* control chart and an automatic relevance neural network to analyze the wind turbine power to identify wind turbine detection faults. Zhang [15], combined with AHP and variable weight theory, used a wind turbine performance evaluation model based on *Grey Theory* and a variable weight fuzzy comprehensive evaluation. However, these studies did not consider the correlation and coupling between the components of the unit, which makes the model inaccurate. Zhang et al. [16], based on SVR prediction model, helped to establish a prediction model, which takes the amount of SCADA systems as input and the active power of the unit as output. The disadvantage of this model is that the feature extraction of the high-dimensional input vectors is not easy, and the power is used as the only standard to diagnose the state of the unit. BP neural network is used to model and predict gearbox and generator [17], and multiagent method is used to synthetically analyze the diagnosis results of different components, giving the overall operation status of the unit. However, the use of neural network modelling requires time-consuming learning process, and the selection of learning samples lacks basis.

Being based on statistical analysis, it commonly requires vast datasets for providing meaningful indications: the most common opinion therefore is that SCADA can detect incipient faults at a late stage. Astolfi et al. [18] employed artificial neural networks, for their capability in reconstructing nonlinear dependency between inputs and outputs, and formulated simple models for the diagnosis of occurring faults at the level of gearbox. The datasets employed have the 10-minute sampling time of the common SCADA control systems; the gearbox vibrations and the gearbox temperatures are selected as target output to model.

It will be shown that the time resolution of SCADA is too coarse for reliable vibration analysis, which should be rather observed at its proper time scale (several Hz). At present, data mining methods such as clustering and statistical model are widely used in domestic and foreign enterprises, but their cleaning process is complicated and the cleaning conditions are harsh [19]. Therefore, in order to make a reliable analysis of the power generation performance of wind turbines, an efficient and versatile cleaning method is urgently needed.

In view of this, this paper firstly extracts the features from the massive and high-dimensional data collected by the SCADA system, removes the irrelevant and redundant parameters of the operation state of the unit, and improves the monitoring accuracy of the wind turbine by improving the model input. The reasonable threshold is selected to alarm the abnormal state of wind turbine to avoid false alarm and untimely alarm.

The paper is organized in three sections. Section 2 discusses feature selection and sparse learning technology to reduce the dimension of the operation parameters of the SCADA system, remove the independent and redundant parameters of the operating state of the wind turbine, and retain the related characteristic parameters. In Section 3, the multi-input and multi-output SVR model, which takes the active power, the speed of the blade, and the pitch 1 angle as the output vector and the characteristic parameter as the input vector, is established. Cross validation (CV) is combined with a genetic algorithm (GA) for parameter optimization. In Sections 4 and 5, the proposed method is applied to the industrial data. Performance of the proposed model is also discussed. Section 6 concludes the paper.

2. Data Mining of Characteristic Parameters for Wind Turbines

The data collected and recorded by the SCADA system of the wind turbine has high-dimensional characteristics. In this paper, 74 digits of the wind turbine components are selected. The method of feature parameter data mining reduces the number of features and dimension disaster so that the generalization ability of the model is stronger and the overfitting phenomenon is reduced. The commonly used methods for selecting characteristic parameters include principal component analysis (PCA) [20], the Pearson correlation coefficient [14], and the random forest method [21]. When the data are high-dimensional vectors, the calculation of PCA is complicated and it is most suitable for linear data. Pearson correlation coefficient is only sensitive to the disadvantages of the most obvious linear relationship. The random forest method is prone to an overfitting phenomenon. Therefore, in this paper, a data mining algorithm based on the grey correlation degree [22] is proposed to overcome the above shortcomings and to improve the accuracy and effectiveness of wind turbine operation state assessments.

2.1. Extraction of the Characteristic Parameters Based on Grey Relational Grade. There are 74 variables in wind turbine information recorded by the SCADA system. The acquisition interval is 10 minutes, as shown in Figure 1. Figure 1

FIGURE 1: Monitoring variables and numbers of SCADA system.

shows the monitoring variables collected by the wind turbine SCADA system and its response code. The object of study in this paper is that the wind turbines are in the condition of unlimited power and healthy operation. They have some monitoring quantities such as the control mode and alarm of some parameters recorded in the SCADA system, speed mode, state of shaft 1, shaft 2. and shaft 3 converters, etc. Variables in an invariant state can be ignored. Table 1 is part of the parameter alarm information of GE wind turbine manufacturer. To preprocess these eigenvectors, we must remove these eigenvectors to avoid the disaster of dimensionality caused by too many features.

Wind turbine operation is mainly reflected in active power, rotor speed, and pitch angle; these three parameters are used as input vectors. We take the pitch angle 1 as an example to monitor the pitch angle. The grey correlation with other variables is calculated to reduce the wind turbine data dimension while ensuring the smallest loss of information. The concrete steps of extracting the operating characteristic parameters using the grey correlation degree are as follows:

- (1) The characteristic set of the wind turbine operating state D is $D = [p_1, p_2, \dots, p_m]$.
- (2) According to the parameters of the primary wind turbine D , the corresponding parameters are extracted from the SCADA system as the sample set of the grey correlation degree Ω :

$$\Omega = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}, \quad (1)$$

where x_{mn} is the n parameter of the m samples. The degree of correlation between the calculated parameters is found as follows.

TABLE 1: Alarm information of some parameters of wind turbine.

Alarm information
Latest(=n) alarm coming from the frequency converter
$n-1$ alarm coming from the frequency converter
$n-2$ alarm coming from the frequency converter
$n-3$ alarm coming from the frequency converter

Determine the reference sequence X_0 and the comparison sequence $X_i (i = 1, 2, \dots, m)$ according to the training sample Ω .

The absolute difference X_i between X_0 and sequences Ω is calculated from the sample set as

$$\Delta_i(k) = |x_0(k) - x_i(k)|, \quad i = 1, 2, \dots, n. \quad (2)$$

The absolute difference Δ_i is used to calculate the maximum difference and minimum difference of level two, respectively, as

$$\begin{aligned} M &= \max_i \max_k \Delta_i(k), \\ N &= \min_i \min_k \Delta_i(k). \end{aligned} \quad (3)$$

X_0 is calculated, and the sequence X_i in the moment of k for correlation coefficient $r_{0i}(k)$ is compared by

$$\begin{aligned} r_{0i}(k) &= \frac{N + \xi M}{\Delta_i(k) + \xi M}, \quad \xi \in (0, 1); k = 1, 2, \dots, n; i = 1, 2, \dots, m, \\ r_{0i} &= \frac{1}{n} \sum_{k=1}^n r_{0i}(k), \quad i = 1, 2, \dots, m. \end{aligned} \quad (4)$$

2.2. Mining of Characteristic Parameter Data of Wind Turbines. The grey correlation analysis mentioned above is

applied to 1.6 MW wind turbines. By preprocessing the eigenvectors, 23 variables related to the operating state of the wind turbine are selected, and the values of these variables are extracted from the SCADA system. Based on operational data from January 1, 2015, to January 1, 2017, the sample set features selected by grey correlation analysis and the characteristic parameters of the grey correlation matrix for color map are as shown in Figure 2. According to Figure 1, the grey correlation degree between each primary parameter of the power, rotor speed, and pitch angle is different, which makes it feasible to excavate the characteristic parameters of the generator set. In this paper, variables with correlation degree greater than 0.5 are selected as input of monitoring variables. The set of characteristic parameters is shown in Tables 2–4.

3. Prediction Model Based on Support Vector Regression

After the data effectiveness analysis and dimensionality reduction are conducted, the parameters of the wind turbine are regressed. SVR [23] algorithm of structural risk minimization criterion solves the practical problems of small sample, non-linearity, and high dimension and overcomes the shortcomings of the indetermination of the network structure and local minima, over learning and under learning. Therefore, this paper chooses SVR algorithm to build regression prediction model. The specific algorithm is as follows.

Set a given sample training set for $(x_i, y_i) i = 1, 2, \dots, l, x \in R^l, y \in R, x_i$ is the i input vector, and y_i is the i output vector. Nonlinear mapping should be used for the nonlinear SVR model $\varphi(\cdot)$. The mapping sample sets are used to feature spaces $\varphi(x_i)$. The optimal decision function is as follows:

$$f(x) = w^T \cdot \varphi(x) + b, \quad (5)$$

where w^T is the characteristic space weight coefficient vector and b is the bias. It is assumed that all training samples can be in precision with linear functions at ε accuracy. According to the principle of structural risk minimization, the problem can be formulated as

$$\begin{aligned} \min \quad & \frac{1}{2} w^T w + c \sum_{i=1}^l (\xi_i + \hat{\xi}_i), \\ \text{s.t.} \quad & \begin{cases} y_i - f(x_i) \leq \varepsilon + e_i, \\ f(x_i) - y_i \leq \varepsilon + e_i, \\ \xi_i \geq 0, \\ \hat{\xi}_i, \quad i = 1, \dots, l, \end{cases} \end{aligned} \quad (6)$$

where $\hat{\xi}_i$ is the relaxation factor. Introducing the Lagrange function, the optimization problem in the dual space is used to obtain the following formula:

$$\begin{aligned} L(w, b, e, \alpha) = & \frac{1}{2} w^T w + c \sum_{i=1}^l (\xi_i + \hat{\xi}_i) - \sum_{i=1}^l \alpha_i [\xi_i + \varepsilon - y_i + f(x_i)] \\ & - \sum_{i=1}^l \hat{\alpha}_i [\hat{\xi}_i + \varepsilon - y_i + f(x_i)] - \sum_{i=1}^l (\xi_i \gamma_i + \hat{\xi}_i \hat{\gamma}_i), \end{aligned} \quad (7)$$

where $\alpha_i, \hat{\alpha}_i, \gamma_i$, and $\hat{\gamma}_i$ are the Lagrange multipliers and c is the penalty factor respective to w, b, ξ , and α . Find the partial derivative and make it equal to 0 and bring the derivative into the Lagrange function:

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1, j=1}^l (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j)(\varphi(x_i) \cdot \varphi(x_j)) + \sum_{i=1}^l (\alpha_i - \hat{\alpha}_i) y_i - \sum_{i=1}^l (\alpha_i + \hat{\alpha}_i) \varepsilon, \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1, j=1}^l (\alpha_i - \hat{\alpha}_i) = 0, \\ \alpha_i - \hat{\alpha}_i \in [0, C]. \end{cases} \end{aligned} \quad (8)$$

Using the positive definite matrix theorem, the $\varphi(x_i) \varphi(x_j)$ inner product is replaced by a kernel function $k(x_i, x_j)$. Therefore, the SVR function can be obtained as follows:

$$f(x) = \sum_{i=1}^l (\alpha_i - \hat{\alpha}_i) k(x_i, x_j) + b. \quad (9)$$

In the kernel function, the structure of the radial basis function (RBF) kernel is simple, and its generalization ability

is better. Based on this, the kernel function of the model is selected as the radial basis function. $k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$, where σ is the kernel width.

3.1. Genetic Algorithm (GA). In this model, the penalty coefficient and parameter of the kernel function affect the SVR precision. Therefore, GA is used to optimize the parameters of the SVR model, which is based on the natural selection and genetic mechanism of the theory of biological

FIGURE 2: Feature parameters of the grey correlation degree color map.

TABLE 2: Feature parameters' code of power prediction.

Pitch system	Average pitch angle
Gearbox	Temp. gearbox, temp. gearbox bearing, temp. shaft bearing
Power grid	Reactive power
Generator	Temp. generator 1, temp. generator 2, generator speed (CCU), generator speed (PLC), temp. generator cooling air
Environment variable	Temp. ambient, wind speed
Nacelle	Temp. nacelle, nacelle revolution
Rotor system	Rotor speed (PLC), temp. hub
Yaw system	Wind deviation 10 sec.

TABLE 3: Feature parameters' code of rotor speed prediction.

Gearbox	Temp. gearbox, temp. gearbox bearing, temp. shaft bearing
Power grid	Reactive power
Generator	Temp. generator 1, temp. generator 2, generator speed (CCU), generator speed (PLC), temp. generator cooling air
Environment variable	Temp. ambient, wind speed
Nacelle	Temp. nacelle, tower acceleration, nacelle position
Rotor system	Rotor speed (PLC), temp. hub
Wind deviation	Wind deviation 10 sec.

TABLE 4: Feature parameters' code of pitch 1 angle prediction.

Pitch system	Collective actual (measured) pitch angle
Gearbox	Temp. gearbox, temp. gearbox bearing, temp. shaft bearing
Power grid	Reactive power
Generator	Temp. generator 1, temp. generator 2, generator speed (CCU), generator speed (PLC), temp. generator cooling air
Environment variable	Temp. ambient, wind speed
Nacelle	Temp. nacelle, nacelle revolution
Rotor system	Rotor speed (PLC), temp. hub
Yaw system	Wind deviation 1 sec., wind deviation 10 sec.

FIGURE 3: Flow chart of the model.

evolution by Darwin [24] to find the optimal solution. The main process is to encode the solution to the problem. There are two ways to code the solution individually, including binary coding and real number coding, which essentially maps the solution space to the chromosome space. Then, a reasonable initial population is generated in these solution spaces, and individuals are selected according to fitness function, genetic selection, crossover, and mutation operation. The individual with high fitness value is kept and vice versa. This new generation of offspring retains the advantages of the previous generation, whereas the last generation did not. This process is iterated many times until the optimal solution is obtained.

3.2. Cross Validation (CV). In machine learning, CV is mainly used for model performance evaluation and learning. The basic principle is that the original sample is divided into a training set and a validation set, and then the training set is used to train the model. The model validated by the test set is obtained from the training model. As a performance index evaluation model, CV considers the training error as well as the generalization error. The most common CV method is k -folding cross validation (K -fold CV), and the specific algorithms are as follows:

- (1) The sample S are divided into k subsets that are not intersected, and the number of samples is m/k . $S_1, S_2, \dots, S_k$ are remembered.

- (2) For each model $M_i (i = 1, \dots, k)$, do the following. Training set $S_1 \cup \dots \cup S_{i-1} \cup S_{i+1} \cup \dots \cup S_k$ and model M_i . Get the corresponding hypothesis function h_{ij} . Validation set S_i is used to calculate the generalization error $\hat{\epsilon}_{S_i}(h_{ij})$.
- (3) The average generalization error of each model is calculated, and the model with the least generalization error M_i is selected.

3.3. SVR Parameter Selection Based on CV and GA. The selection of the parameters of the SVR model is essentially the optimization of the model. The algorithm of K -fold CV and GA is used to optimize the parameters of SVR, and $k = 3, 5, 8, 10$. The model as shown in Figure 3 is as follows:

- (1) The SVR parameters (c, σ) are coded to form the initial population.
- (2) For population decoding, we calculate the fitness of individuals based on the K -fold CV method. In this paper, $\hat{\epsilon}_{S_j}(h_{ij})$, the minimum mean square error MSE of samples (c, σ) , is chosen as the fitness function value of the GA algorithm.
- (3) Judge whether or not to meet the terminating condition if it is satisfied to turn (5); otherwise, proceed to (4).
- (4) Update the population by selection, cross, and variation; then, return to (2).
- (5) The optimal (c, σ) and optimal model is output.

FIGURE 4: (a) 3-CV MSE of SVR model. (b) 5-CV MSE of SVR model. (c) 8-CV MSE of SVR model. (d) 10-CV MSE of SVR model.

4. Condition Monitoring Based on SVR Parameter Optimization

4.1. Data Processing. In this paper, power, rotor speed, and pitch angle are taken as the output vectors and other feature parameters are taken as input vectors, and then multiple-input and multiple-output SVR model is established. The accuracy of the proposed model is verified by running data of the wind turbine for four months.

- (1) According to the fault information recorded by the SCADA system, the samples of the maintenance shutdown due to the failure of the wind turbine and the samples of the less power operation are eliminated.
- (2) Consider that the cutting wind speed is 3 m/s, the rated wind speed is 12 m/s, and the cutting wind speed is 25 m/s. According to the actual power curve of wind turbine, the wind speed range selected in this paper is 3 to 25.
- (3) To eliminate the magnitude of interference between the parameters, the parameters are normalized to $[0, 1]$, according to the dimensions.

4.2. Model Establishment. In order to prove the validity of the model, this paper selects four months effective data of wind turbine to predict. In this model, the principle of cross validation selection first considers the minimum error MSE. According to the errors of MSE, to avoid the occurrence of the learning state, a group of smaller penalty parameters is selected as the best parameter. From Figure 4, we can see that 5-CV and GA model are the best. The average fitness curve in Figure 4 indicates the average fitness of all the individuals in each generation. The best fitness curve represents the maximum fitness of all individuals in each generation. The convergence of the fitness curve is very fast, and the convergence level of the final fitness curve is relatively consistent, which reflects the optimization of SVR parameters. When the power, rotor speed, and pitch angle are output, the best parameter is applied to the SVR model. The comparison between the actual and predicted values of the SVR model is shown in Figure 5. Table 5 lists the power, rotor speed, and pitch error. The mean relative error values indicating the good prediction accuracy and stability of the SVR model CV-GA algorithm is shown in Table 5.

FIGURE 5: (a) Predicted and actual values of power. (b) Predicted and actual values of rotor. (c) Predicted and actual values of pitch 1 angle.

4.3. Discriminating the Health Status of Wind Turbine Based on Threshold. According to the trained SVR model, the observed values of power, rotor speed, and pitch angle can be obtained from the current input vector, and the distance between the measured value and the observed value extracted from SCADA system is δ . According to the distance between the measured value and the observed value,

TABLE 5: Errors of the SVR model.

SVR model	Power (%)	Rotor speed (%)	Pitch angle (%)
Maximum relative error	10.36	11.87	11.95
Minimum relative error	0.49	0.91	0.51
Mean relative error	3.01	4.95	3.73

TABLE 6: Operational conditions for the wind turbine with different distance thresholds.

Threshold ν	Leakage rate (%)	False positive (%)
0.04	0	92
0.06	3	86
0.08	7	72
0.1	10	41
0.12	12	33
0.14	15	24
0.16	16	12
0.18	28	11
0.2	37	8
0.22	52	5

FIGURE 6: Distance between the predicted and actual values.

the threshold ν is compared with the set threshold to distinguish the operating state of the wind turbine. The distance from the observed value to the measured value in the SCADA system is defined as

$$\delta = \frac{|\alpha_j - \alpha'_j|}{\alpha_j}. \quad (10)$$

If ν is selected too low, the algorithm is too sensitive to the change of the operating state of the wind turbine and it is prone to misjudge the results. If ν selection is too large, the prediction time will be reduced and the detection rate of the abnormal operating state will be affected. To solve this problem, it is necessary to select the appropriate threshold ν . In this paper, two-year SCADA data of the 1.6 MW wind turbine is analyzed, and the appropriate threshold is determined. From Table 5, it can be seen that threshold selection will cause the operating state of the wind turbine to not be normally recognized. To find a lower detection rate and the misjudgment rate, the threshold is selected as $\nu = 0.16$ (in Table 6).

5. Example Analysis

At 5:32 on August 7, 2016, a wind turbine in Hebei Province went into shutdown due to the SCADA system failure alarm. After checking the pitch gear of blade 1 of the wind turbine, the wind turbine went into shutdown due to failure. 950 sets of data were extracted from the SCADA system before the blower alarm stopped. 950 sets of collected data are input

into the model, as shown in Figure 6. As can be seen from the figure, close to 150 data points were detected as abnormal points before the wind turbine shut down. The situation of power restriction indicates that some of the wind turbines have begun to deteriorate at these times, and the unit has issued abnormal alarms. The reduced power generated by the unit indicates that the model can give a hint before the failure occurs. Therefore, the proposed model is effective for the state monitoring and fault prediction of the wind turbine, and it can avoid the continued deterioration of the fault and the influence on the safe operation of the power grid.

6. Conclusion

In order to extract relevant state from massive and high-dimensional data of the SCADA system, realizing the monitoring of the state of the wind turbine, a grey correlation degree is proposed based on data mining technology to extract characteristic parameters of the wind turbine's operating state, which reduces the data dimensions and computation. To improve the precision, GA and CV are combined to optimize the parameters of the regression model. To verify the validity of the model, the threshold of the SVR model is analyzed, and the model is applied to wind farm. The results show the following:

- (1) By establishing a data mining model of the characteristic parameters based on the grey correlation analysis, we extract parameters that are more related to the power, rotor speed, and pitch angle, effectively avoiding "dimension disaster."

- (2) By comparing the power, rotor speed, pitch angle regression model, and the measured values, the results showed that the average relative error of the SVR model is low. The regression model has high accuracy and generalization ability; it can be applied to wind turbine anomaly distinguishing analysis.
- (3) Applying the model to practice, the analysis of the model results and the SCADA system can be used to record the measured values. The results show that when using the distance threshold to choose the appropriate conditions, the wind turbine condition monitoring can reflect the operating status of the wind turbine to provide technical references for online monitoring of wind turbines.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Liang Tao provided data and ideas as well as optimization algorithm; Qian Siqi mined and analyzed data; Zhang Yingjuan used GA-optimized SVR algorithm to predict monitoring variables; and Shi Huan analyzed and summarized the experimental results.

References

- [1] T. Liang, Z. Yuan, C. Mei et al., "Wind power remote centralized intelligent monitoring system design and management," *Automation and Instrumentation*, vol. 30, no. 11, pp. 50–53, 2015.
- [2] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. Tameghe, and G. Ekemb, "Wind turbine condition monitoring: state-of-the-art review, new trends, and future challenges," *Energies*, vol. 7, no. 4, pp. 2595–2630, 2014.
- [3] B. Hahn, M. Durstewitz, and K. Rohrig, "Reliability of wind turbines experience of 15 year with 1500 WTs," in *Wind Energy*, J. Peinke, P. Schaumann, and S. Barth, Eds., pp. 329–332, Springer, Berlin, Germany, 2007.
- [4] J. Ribrant and L. M. Bertling, "Survey of failures in wind power systems with focus on Swedish wind power plants during 1997–2005," *IEEE Transactions on Energy Conversion*, vol. 22, no. 1, pp. 167–173, 2007.
- [5] Lingwu Energy, H. Naishan, and R. Li, "Short-term wind power prediction based on cloud support vector machine model," *Power Automation Equipment*, vol. 33, no. 7, pp. 34–38, 2013.
- [6] H. H. Yang, M. L. Huang, and P. C. Huang, "Detection of wind turbine faults using a data mining approach," *Journal of Energy Engineering*, vol. 142, no. 3, article 04015021, 2016.
- [7] R. K. Pandit and D. Infield, "Comparative analysis of Gaussian process (GP) power curve models based on different stationary covariance functions for the purpose of improving model accuracy," *Renewable Energy*, vol. 140, pp. 190–202, 2019.
- [8] D. Astolfi, F. Castellani, and L. Terzi, "Wind turbine power curve upgrades," *Energies*, vol. 11, no. 5, p. 1300, 2018.
- [9] D. Astolfi and F. Castellani, "Wind turbine power curve upgrades: part II," *Energies*, vol. 12, no. 8, p. 1503, 2019.
- [10] S. Wan, J. Wan, and C. Zhang, "Comprehensive evaluation of wind power unit performance evaluation based on grey theory and variable weight fuzzy mathematics," *Acta Energiæ Solaris Sinica*, vol. 36, no. 9, pp. 2285–2291, 2015.
- [11] K. Bassett, R. Cariveau, and D. S.-K. Ting, "Vibration analysis of 2.3 MW wind turbine operation using the discrete wavelet transform," *Wind Engineering*, vol. 34, no. 4, pp. 375–388, 2010.
- [12] Y. Chun-yu, G. Jian-ying, and X. Shi-guang, "Failure mechanism analysis and failure number prediction of wind turbine blades," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 12, no. 3, pp. 533–540, 2014.
- [13] S. Chakkor, M. Baghoury, and A. Hajraoui, "Wind turbine fault detection system in real time remote monitoring," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 4, no. 6, 2014.
- [14] J. J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, 2003.
- [15] C. Zhang, "To evaluate the performance of Waanje, wind turbine fuzzy comprehensive evaluation based on Grey theory and variable weight," *Journal of Solar Energy*, vol. 36, no. 9, pp. 2285–2291, 2015.
- [16] X. Zhang, S. Hao, J. Li et al., "Grey correlation based analysis on impacting factors of maximum power point tracking control of wind power generating unit," *Power System Technology*, vol. 39, no. 2, pp. 445–449, 2015.
- [17] A. Zaher, S. D. J. McArthur, D. G. Infield, and Y. Patel, "Online wind turbine fault detection through automated SCADA data analysis," *Wind Energy*, vol. 12, no. 6, pp. 574–593, 2009.
- [18] D. Astolfi, L. Scappaticci, and L. Terzi, "Fault diagnosis of wind turbine gearboxes through temperature and vibration data," *International Journal of Renewable Energy Research (IJRER)*, vol. 7, no. 2, pp. 965–976, 2017.
- [19] Y. Feng, Y. Qiu, C. J. Crabtree, H. Long, and P. J. Tavner, "Monitoring wind turbine fault detection," *Wind Energy*, vol. 16, no. 5, pp. 728–740, 2013.
- [20] W. Liao, R. Zhang, W. Yu et al., "PV output power forecast based on similar samples and PCA," *Solar Journal*, vol. 37, no. 9, pp. 2377–2385, 2016.
- [21] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [22] X. Zhang, H. Sipeng, J. Li et al., "Analysis of influence factors of MPPT control based on grey correlation degree," *Power Grid Technology*, vol. 39, no. 2, pp. 445–449, 2015.
- [23] Wind Turbine Ruiming, *Support Vector Machine Analysis Theory and Application*, China Power Press, Beijing, China, 2007.
- [24] Y.-Y. Hong and Y. Y. Chen, "Placement of power quality monitors using enhanced genetic algorithm and wavelet transform," *IET Generation, Transmission & Distribution*, vol. 5, no. 4, pp. 461–466, 2011.

Submit your manuscripts at
www.hindawi.com