Ice Detection for Wind Turbine Blades Based on PSO-SVM Method

Zhou Guangfei1,a, Tan Wen1 and Zhang Da1

1School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

a Email: 1138814841@qq.com

Abstract. Detection of ice formation on wind turbine blades is still in the research and development stage. Generally, wind turbines have to be shut down to deice the blades when the icing is serious. This paper adopts a data-driven approach to detect the icing blades by analyzing a lot of actual wind field data collected by SCADA system to create a prediction model. Considering the characteristics of industrial big data, firstly, we preprocess the field data by eliminating the singular data, and then select suitable features through mechanism analysis. Finally, the characteristic data are fed into the support vector machine (SVM) which uses the particle swarm optimization (PSO) algorithm to optimize the parameters. Experimental results show that the proposed approach can achieve good performance and provide a new idea for fault diagnosis of ice formation on wind turbine blades.

1. Introduction

Ice formation on wind turbine is a global problem. Blade icing caused by low temperature environment, changes in materials and structural properties, changes in load, etc. pose a greater threat to the performance and safe operation of wind turbines [1]. At present, studies on the icing of wind turbines have focused on finding ways to effectively prevent icing and remove icing on the blades, such as wind tunnel experiments and numerical simulation studies [2]. On the other hand, if we can accurately predict the icing process so as to be able to open the deicing system as soon as possible, it is also an important means to prevent the icing of the blades. The literature [3] collected the temperature data of the blade through the infrared sensor, and judged whether icing or not according to the preset icing threshold. Huidong Gao et al. proposed an ultrasonic guided wave detection technique [4]. However, the above methods require additional equipment or blade modification resulting in rising costs. The literature [5] studied environmental temperature and humidity, wind turbine power loss, etc. to establish a corresponding icing detection model, with the help of the researcher's expertise. It can determine the health status of the blade efficiently. But this detection method also has defects: more dependent on professional knowledge; lack of generality. The detection accuracy will be greatly discounted when the operating environment changes.

The SCADA system has been widely used in the wind power industry and it generates a large amount of monitoring data everyday in which often contain abnormal operating status. If we departure from the perspective of data-driven and use the data for mining and modeling, it can realize the prediction and diagnose of some serious faults, so that the past emergency maintenance mode can be transformed into active predictive maintenance. Also, the method can effectively improve the utilization rate and reduce maintenance cost of wind power equipment.

This paper analyzes the SCADA data of a wind field, selects the characteristic variables that can
reflect the icing state of the blades, and proposes a data-driven fan blade icing detection method. The specific process is shown in Figure 1.

![Flow chart of ice detection on wind turbine blades](image)

**Figure 1 Flow chart of ice detection on wind turbine blades**

### 2. Data pretreatment

This paper uses two data sets and each set includes twenty hundreds sample collected from two wind turbines in a wind farm for one month. The detection system has monitored a total of 26 variables, as shown in Table 1. Data preprocessing becomes more important because of large samples and characteristic variables.

| Number | Variable |
|--------|----------|
| 1      | wind speed |
| 2      | generator speed |
| 3      | power |
| 4      | wind direction |
| 5      | wind mean |
| 6      | yaw position |
| 7      | yaw speed |
| 8      | pitch 1 angle |
| 9      | pitch 2 angle |
| 10     | pitch 3 angle |
| 11     | pitch 1 speed |
| 12     | pitch 2 speed |
| 13     | pitch 3 speed |
| 14     | pitch 1 motor temperature |
| 15     | pitch 2 motor temperature |
| 16     | pitch 3 motor temperature |
| 17     | acceleration x |
| 18     | acceleration y |
| 19     | environment temperature |
| 20     | internal temperature |
| 21     | ng5 1 temperature |
| 22     | ng5 2 temperature |
| 23     | ng5 3 temperature |
| 24     | pitch 1 ng5 DC |
| 25     | pitch 2 ng5 DC |
| 26     | pitch 3 ng5 DC |

**Table 1** Wind turbine detection variables

| Turbine number | Total sample | Normal sample | Ice sample | Invalid sample |
|----------------|--------------|---------------|------------|----------------|
| 1              | 200000       | 182135(91.1%) | 9753(4.9%) | 9112(4.6%)     |

2.1 Data analysis

According to the indicated time stamp, table 2 shows the distribution of the training set.

| Turbine number | Total sample | Normal sample | Ice sample | Invalid sample |
|----------------|--------------|---------------|------------|----------------|
| 1              | 200000       | 182135(91.1%) | 9753(4.9%) | 9112(4.6%)     |

First, we should delete the invalid samples in training set. Then use the box plot method to remove singularities. Its basic principles and steps are:

1. Sort the samples based on a specific variable in descending order;
2. Calculate the upper quartile $Q_2$, the lower quartile $Q_1$, and the median of the variable;
3. Calculate the upper and lower limits:

   \[ UCL = Q_2 + k \cdot \Delta, LCL = Q_2 - k \cdot \Delta \]  \( k \)

4. Delete the sample including the singularity of the variable;
5. Continue to calculate the next variable until all variables have been tested. After the above operation, approximately 6,000 abnormal samples are eliminated.
2.2 Balance data set
By analyzing the data distribution in Table 2, we can find that the ratio between normal and abnormal data is as high as 18:1. This is a typical set of unbalanced data. To solve this problem, this paper adopts a large sample undersampling and a small sample oversampling methods to improve both normal and icing data to 20000 samples. However, a lot of useful information will be lost from the undersampling of 180000 sets of normal data. Therefore, in combination with the actual blade icing conditions, several filtering rules are proposed. Some samples that are not significantly iced will be eliminated firstly shown in Figure 2 where the red part is the normal sample, the blue part is the faulty sample, and the samples inside the ellipse are the apparently non-icing data.

![Figure 2: Four rules for removing apparently no icing data](image)

A total of 50000 samples in the norm data are removed after screening and left about 140000 samples which still remains high. Actually, the values of the corresponding feather variables between many samples are almost the same, that is, many samples represent the same information. An undersampling algorithm based on the similarity function is proposed to delete redundant samples. The function is defined as:

$$R_{ij} = \exp\left(-\frac{1}{\delta} \|x_i - x_j\|^2\right)$$  \hspace{1cm} (2)

Where $x_i, x_j \in \mathbb{R}^m (i, j = 1, 2, \ldots, n)$, $m$ is the sample dimension, $n$ is the sample number, $R_{ij}$ represent the similarity between $i$th and $j$th sample, $\|\|$ is matrix norm, $\delta$ is normalization factor, defined as:

$$\delta = \sum_{j=1}^{m} (\max_{j} L_j - \min_{j} L_j)$$  \hspace{1cm} (3)

Where $L_j$ is the $j$th row of the matrix $X$

The specific algorithm process is:
1) Normalize the original samples and calculate $\delta$;
2) Calculate the similarity between $i$th and $j$th ($j > i$) samples;
3) Define a suitable threshold $\varepsilon$ ($\varepsilon \in [0, 1]$), if $R_{ij} > \varepsilon$, remove the sample $x_j$;
4) Repeat the above steps until all samples have been checked;

Figure 3 shows the proportion of retained samples in different $\varepsilon$. The sample is about 60% remaining when we take $\varepsilon = 0.999$. 140000 reduced to about 80000.
Finally, we use fixed-length sampling to collect 20000 normal samples from 80000 samples. At the same time, SMOTE algorithm are implemented to double icing fault data which only have about 10000 sets. The main idea of algorithm is to insert new sample between a few much closed samples to achieve the goal of balancing the data sets. It is not simply to copy samples according to the method of random oversampling, but to add new non-existent examples. This idea can effectively avoid overfitting of the classifier. The specific algorithm is:

1) Calculate each minority class sample $x$’s $k$ nearest neighbor samples ;

2) According to sampling rate $N$, we randomly choose $N$ samples denoted as $y_1, y_2, \cdots, y_N$ from $k$ nearest neighbor samples; then, new samples $p_j$ are constructed by performing random linear interpolation between $x$ and $y_j (j = 1, 2, \cdots, N)$;

$$p_j = x + \text{rand}(0,1) \cdot (y_j - x), j = 1, 2, \cdots, N \quad (4)$$

After the above series of operations, two kinds of data sets can reach a ratio of 1:1.

3. Feather selection

Wind turbine blade icing is a complex process that is affected by a variety of environmental variables. Therefore, it is very difficult to establish a rigorous mathematical model. It will cause a huge amount of calculations if we bring in all feathers into data-driven methods. On the basis of fully understanding the operating principle and icing process of wind turbine, we can exploit the icing sensitive characteristics to achieve an effective prediction model.

The essence of the wind generator is that the wind turbine drives the blades through wind, then; the blades drive the generator through the gear box to generate electrical energy. Power will drop because icing can changes the aerodynamic shape of the blades. Therefor, the relationship between wind speed and power can indicate that icing is or not. Figure 4(a) gives power scatter of normal and abnormal data and we take the curve obtained by fitting normal data as the theoretical power curve. The residuals of actual power and the theoretical power can be a feather. It can be seen that if the residual is large, it may indicate that the blade is already frozen.

Similarly, the icing of the wind blades will also affect the corresponding relationship between wind speed and generator speed. We can fit normal speed curve and use the rotating speed residual as the second feature, as shown in Figure 4(b).
Environment temperature is the key factor that causes the icing of wind turbine blades. Furthermore, the residuals between Environment and internal temperature are also sensitive to blade icing. As shown in Figure 5(a)(b), most of icing data, shown as blue scatter in the ellipse are mostly located above at the same wind speed.

The use of principle component analysis (PCA) proposed by literature [7] to analyze the projection of the sample in the dominant and non-dominant direction can construct sensitive feathers. Figure 6(a)(b) shows that in non-principal component direction of wind speed and power, it can be seen that there is a clear distinction between normal data and icing data.

Finally, wind speed, power, environment temperature, internal temperature, and four derived variables are selected as total features.

4. Classifier construction and parameter optimization
In this paper, support vector machine (SVM) is used and particle swarm algorithm (PSO) is introduced to find the optimal parameters. The algorithm is described below.

4.1 Fundamental of SVM
SVM is a machine learning method proposed by Vapnik on the basis of statistical learning theory. By introducing a kernel function, SVM can effectively solve the problem of linear inseparability. The algorithm can be simply described as:

There are a set of training samples \((x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\), where \(x_i \in \mathbb{R}^n\), \(y_i \in \{1, -1\}\), \(l\) is the number of samples. The decision function of SVM is:

\[
 f(x) = sign \left( \sum_{i=1}^{l} y_i a_i k(x_i, x) + b \right)
\]
number of training set. The SVM algorithm is to find optimal hyperplane in the n-dimensional data space, and its equation can be expressed as:

$$w^T \cdot x + b = 0$$  \hspace{1cm} (5)

Where $w, b$ are the two parameters to be solved. To maximize the geometric interval between the hyperplane and the two types of support vectors, the objective function and constraints are written:

$$
\begin{align*}
\min & \frac{1}{2} \|w\|^2 \\
\text{s.t.} & \ y_j \cdot (w^T \cdot x_j + b) \geq 1, i = 1, \cdots, n
\end{align*}
$$  \hspace{1cm} (6)

In dealing with linear inseparable and outlier samples, SVM introduces the concepts of kernel function K and penalty factor C respectively. Ultimately, the dual problem of the entire objective function can be written as:

$$
\begin{align*}
\max & \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j y_i y_j K(x_i, x_j) \\
\text{s.t.} & \sum_{i=1}^{n} a_i y_i = 0, \quad 0 \leq a_i \leq C, i = 1, \cdots, n
\end{align*}
$$  \hspace{1cm} (7)

In this paper, RBF function is chosen as the kernel function. Its expression is as follows:

$$K(x_i, x_j) = \exp(-g \|x_i - x_j\|^2), g > 0$$ \hspace{1cm} (8)

Where $g$ is parameter of the function.

After the type of kernel function is selected, parameters $C$ and $g$ are optimized in combination with the PSO algorithm.

4.2 Fundamental of PSO

PSO is a random search and parallel optimization algorithm starts from a random solution and finds the optimal solution through iteration. It has the advantages of easy to understand and implement and few adjustment parameters. The quality of the solution is evaluated by fitness. Let a $E$ dimensional space be composed of a population of $n$ particles and their positions in the dimension are $x_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T$, flying speed $v_i = [v_{i1}, v_{i2}, \ldots, v_{id}]^T$. The best positions in the individual are $p_i = [p_{i1}, p_{i2}, \ldots, p_{id}]^T$ and the best positions for the whole group are $p_g = [p_{g1}, p_{g2}, \ldots, p_{gd}]^T$. In each iteration, the particle’s position and velocity is as follows:

$$
\begin{align*}
V_{id}^{k+1} &= wV_{id}^k + c_1 \eta_1 (p_{id}^k - X_{id}^k) + c_2 \eta_2 (p_{gd}^k - X_{id}^k) \\
X_{id}^{k+1} &= X_{id}^k + V_{id}^{k+1}
\end{align*}
$$  \hspace{1cm} (9)

Where $w$ is inertia weight, $d = 1, 2, \cdots; k$ is the number of iteration; $c_1, c_2$ are acceleration factors; $\eta_1, \eta_2$ are random numbers in the range $[0,1]$.

4.3 Experimental verification and result analysis

20% of the training data are selected for PSO optimization and the parameters of PSO are set to $c_1 = 1.5, c_2 = 1.7, w = 1$. The maximum number of initial populations is 20; the maximum number of iterations is 200 and the number of cross validation is 5. The optimization process is shown in Figure 7. The optimized parameters are set to $C = 0.25$ and $g = 2$. 
At last, normalize the chosen feather samples and bring them into the algorithm to generate the model.

Table 3 shows the prediction results of icing and non-icing on the validation set and the test set. It can achieve high Classification accuracy under the poor quality of data set and using a single classification model.

### Table 3. Results of blade icing prediction

|                | Predicted non-icing | Predicted icing |
|----------------|---------------------|-----------------|
| **Validation set** |                     |                 |
| Actual non-icing | 0.969               | 0.031           |
| Actual icing    | 0.020               | 0.980           |
| **Test set**    |                     |                 |
| Actual non-icing | 0.935               | 0.065           |
| Actual icing    | 0.195               | 0.805           |

As a comparison, the test sets are substituted into logistic regression, KNN and ANN classifiers. In order to facilitate the comparison result, matthews correlation coefficient (MCC) is selected as the evaluation criterion:

\[
MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}
\]  
(10)

Where TP is the correct number of positive sample prediction, TN is the number of negative sample prediction, FP is the error number of positive sample prediction and FN is the error number of negative sample prediction. MCC range is [-1, 1]. The closer to 1 the value is, the more accurate the model prediction is. The results are shown in Table 4. It shows that the model presented in this paper is better than others.

### Table 4. Comparison of performance between four predictive models

| Model     | LR       | KNN      | ANN      | PSO-SVM  |
|-----------|----------|----------|----------|----------|
| MCC       | 0.5259   | 0.6981   | 0.6672   | 0.7463   |

### 5. Conclusions

The data-driven learning methods have a wide range of application prospect in the industrial fields. This paper realizes the prediction of blade icing through analysis of actual field data. However, there are still many improvements can be done in the future.

1) The data from the industrial site has low quality. Therefore, the data preprocessing is very important. What this article has done is not good enough.

2) Combining knowledges of different industrial scenarios and extracting effective feature variables can improve the accuracy and generalization ability of the model. Therefore, it is believed that the prediction result will be more accurate if we can further analyze the principle of blades icing.

![Figure 7. Particle swarm optimization parameter curve](image)
Reference:
[1] Alsabagh A S Y, Tiu W, Xu Y and Virk M S. A review of the effects of ice accretion on the structural behavior of wind turbines [J]. Wind Engineering, 2013, 37(1):59-70
[2] Dong Qiaonian, Jin Zhiegang and Yang Zhigang. A summary of research on wind turbine icing [J]. Machinery Design and Manufacture, 2014, 52(10):269-272
[3] Muñoz C Q G, Márquez F P G and Tomás J M S. Ice Detection Using Thermal Infrared Radiometry on Wind Turbine Blades [J]. Measurement, 2016, 93:157-163
[4] Gao H and Rose J L. Ice detection and classification on an aircraft wing with ultrasonic shear horizontal guided waves [J]. IEEE Transactions on Ultrasonics Ferroelectrics & Frequency Control, 2009, 56(2):334-344
[5] Davis N N, Byrkjedal Ø and Hahmann AN. Ice detection on wind turbines using the observed power curve [J]. Wind Energy, 2016, 19(6):999-1010
[6] Li Ningbo, Yan Tao and Li Naipeng. Ice detection method based on SCADA data [J]. Power Generation Technology, 2018, 39(1):58-62