Ice Detection on Fan Blades

Ruonan Jia, M.S.

Nima Zahadat, Ph.D.

The George Washington University

1. **Background**

Fan Blade Icing is a big problem for wind electricity industry all over the world. Although China’s wind power resources are abundant, they are unevenly distributed. The wind field is mainly the wilderness wind field distributed in high latitudes. The geographical location of the wind place and its climate determine that the fan blades are prone to icing, which can cause great obstacles to the safe and efficient operation of the fan. In addition, as the power of the wind turbine design continues to increase, the height of the wind turbine tower is also growing. Therefore, even in the northern coastal areas and low-latitude mountainous areas, a large number of wind turbines will touch lower clouds in the winter, and it is very easy to freeze in such low temperature and humid environments.

Figure 1 Distribution of wind energy

Once the surface of the fan blade is covered with ice, it will bring a series of problems. For Instance, Complete shutdown and serious loss of power generation；The blade is reduced in power generation due to the original aerodynamic characteristics of the icing loss itself；Blade overload causes blade stall delay；Uneven ice coating，Thereby exacerbating fatigue damage and even directly causing blade breakage；Large ice cubes falling off the blades can cause uncontrollable serious threats and injuries to people and facilities around them.

SCADA, Supervisory Control And Data Acquisition, has become an important system for wind farm equipment management, monitoring and control。Real-time collection of environmental parameters fan operation, the working parameters, status and control parameters, the wind farm managers real-time visibility and operational health of wind power equipment assets. Therefore, the deviation between the actual power of the fan and the theoretical power can be compared through the monitored device parameters. When the deviation reaches a certain value, the alarm and shutdown of the fan will be triggered. However, most current SCADA systems are still limited to alarms of serious faults that have occurred. These faults are often severe when they reach the alarm phase, and the turbines need to be shut down and repaired, resulting in huge power generation losses and huge maintenance costs.

For the blade icing failure, when the existing detection means triggers the alarm, the blade has already formed a large area of icing, and in such a case, the risk of blade breakage is increased. Although many new wind turbines are designed with automatic de-icing systems, the challenge in practical applications is that it is difficult to turn on the de-icing system in the early stages of icing. When the SCADA system issues an icing failure alarm, the fan has to be shut down and installed. The significance of the de-icing system is greatly reduced. In other words, the speed at which the fan blades freeze, the efficiency of the deicing system, the loss of efficiency of the fan, and the risk of fan operation are determined.

Therefore, the project focuses on the rapid and timely detection of blade icing failures by mining and modeling the big data environment generated by the SCADA system, and performing deicing operations in the early stage of icing, thus making the past stress type. The maintenance mode is changed to the active maintenance mode to effectively improve the utilization rate and operation and maintenance cost of wind power equipment.

* 1. **Solution Plan**

1. Exploring important monitoring parameters that can characterize fan blade icing

Data exploration: to study the data variation characteristics of monitoring parameters such as power, pitch angle and temperature during the icing stage of the blade. In particular, it is necessary to pay attention to the response of different parameters of the blade icing stage when the fan reaches the design power and the design power is not reached. .

1. Identify external factors related to blade icing

Combining the mechanism of fan blade icing, collect meteorological data such as air humidity and precipitation in the wind field in the event of blade icing failure, and provide external information for blade icing detection.

1. Establish a highly reliable blade icing detection method to achieve early warning of icing accidents

Feature extraction of data, construction of input variables, selection of appropriate machine learning algorithms and modeling, comprehensive calculation of cost, detection accuracy and detection speed (identification of the speed of icing events) and other indicators to optimize and evaluate the model.

1. Develop a wind turbine blade icing warning system for practical application.

Based on the blade icing detection model, develop an early warning system to realize the functions of reading and preprocessing of SCADA system data, online detection of blade icing failure and data visualization.

1. **Model building** 
   * 1. Data Overview

3.2.1.1 Fan data overview

The data set of this project is from the SCADA system of a wind farm in Inner Mongolia, with data of three wind turbines. Remove the fan number, fan fault number, fan availability status and time, including 26 monitoring parameters such as wind speed, power generation, pitch angle, and ambient temperature. The data collection interval is 7 seconds. The data set covers multiple dimensions of the operating parameters, environmental parameters and status parameters of the wind turbine. The names and descriptions of the variables are shown in Table 1:

Table 1 Features of dataset

|  |  |  |
| --- | --- | --- |
| **#** | **Features** |  |
| 1 | wtid |
| 2 | wtur\_flt\_main |
| 3 | wman\_state |
| 4 | time |
| 5 | wind\_speed |
| 6 | generator\_speed |
| 7 | power |
| 8 | wind\_direction |
| 9 | wind\_direction\_mean |
| 10 | yaw\_position |
| 11 | yaw\_speed |
| 12 | pitch1\_angle |
| 13 | pitch2\_angle |
| 14 | pitch3\_angle |
| 15 | pitch1\_speed |
| 16 | pitch2\_speed |
| 17 | pitch3\_speed |
| 18 | pitch1\_moto\_tmp |
| 19 | pitch2\_moto\_tmp |
| 20 | pitch3\_moto\_tmp |
| 21 | acc\_x |
| 22 | acc\_y |
| 23 | environment\_tmp |
| 24 | int\_tmp |
| 25 | pitch1\_ng5\_tmp |
| 26 | pitch2\_ng5\_tmp |
| 27 | pitch3\_ng5\_tmp |
| 28 | pitch1\_ng5\_DC |
| 29 | pitch2\_ng5\_DC |
| 30 | pitch3\_ng5\_DC |

Since the fan provides a stable power supply for the sensor and the data is transmitted by the optical fiber, the data quality is good, and there is little data missing or repeated. However, in addition to the icing failure, the data also contains some abnormal data and other types of fault data caused by equipment maintenance. At the same time, not all data is available for icing failure, so the wind turbine will be used next. The data is preprocessed.

Four of the above 30 fields are special, they are fields 1, 2, 3 and 4, which are not used as features in the actual training model. The No. 1 feature is the fan number, which is used for identification. The No. 2 field indicates the number corresponding to the fault. If there is no fault currently, the value of this field is 0. If an ice fault occurs, the value of this field is 133 or 132. When the field No. 3 is 1, it indicates that the data is valid. When it is 0, it indicates that the data collected from the sensor is invalid due to a fault or the like; the field No. 4 determines the sequence of the sequence.

3.2.1.2 Meteorological data

The meteorological environment has a decisive influence on the working state of the wind turbine. The unfavorable weather conditions are the root cause of the icing of the wind turbine blades. Therefore, the interpretation of historical weather information on the icing of wind turbines should be relatively strong. However, the SCADA data for the turbine itself includes only temperature and does not include critical information such as humidity and pressure that have a significant impact on icing. Therefore, it is necessary for the project to obtain historical weather information from the outside and to find out its relationship with the icing of wind turbine blades.

There are many ways to obtain historical meteorological data, mainly the Meteorological Data Center of the China Meteorological Administration, the National Oceanic and Atmospheric Administration (NOAA), and the European Union Meteorological Bureau. The data center of China Meteorological Administration does not open the data of fine-grained ground stations in the long-term. The European Meteorological Bureau does not provide hourly temperature and humidity data. Only NOAA provides three-hour granular meteorological data of the global international ground exchange. Based on the geographic location of the wind farm (as shown in Figure 3), download the 2015 and 2016 data from the nearest 53083 and 53192 weather stations.



Figure 3 Wind farm and weather monitoring station location

According to the manual provided by NOAA, it can be seen that the meteorological information is divided according to the collection time. Each row is divided into different dimensions according to the length of the field, including the acquisition time, ground station number, geographic coordinates, data source (including ground station and aerial high-altitude station), and air. Temperature, dew point temperature, air pressure, wind direction, visibility, data check digits and other additional information. Because the two ground stations are nearly 60 kilometers away from the wind field, the wind direction data cannot be used for interpolation, that is, only the temperature, dew point temperature, and air pressure are available.

* + 1. Data Pre-processing

3.2.2.1 Fan Data Pre-processing

For the icing failure studied in this project, the data with the fault number 133 or 132 is filtered and marked, and then the invalid data with the data availability status of 0 is removed from the original data set. In actual situations, when the fault actually occurs, the data collected by the sensor is generally invalid, and there should be an advance amount when predicting the fault. Therefore, for the specific application scenario, we will use the 2 hours before the occurrence of the icing fault (corresponding to about 1000 data points) as the marking interval of the fault occurrence; 2 to 4 hours before the fault can be considered as a fuzzy zone and not participate in the evaluation; 4 hours outside is the normal operating range. Specifically, we processed the data as follows:

1) Arranging the sequence data in order according to the number 4 field and incrementally marking the index for each data point;

2) Perform first-order differential processing on field 2;

3) The data point of the difference after the 2nd field is 133 or 132. The field is assigned the value 1, which is the first data point when the icing failure occurs. If the field number 2 is 0 (normal) or other fault number, the field is assigned a value of 0, and the data of other fault numbers will be removed in the next step;

4) The 1000 data points before the data point with the field number 2 is also marked as 1, which is marked as the fault point; the first 1000~2000 data points are marked as 2, which will not be considered in the later model training; The data points remain unchanged.

5) The invalid data points with the field number 3 being 0 are removed, including all data with the non-zero field number 2.

After the above data processing, we found that some of the icing events have a small time interval and can be regarded as the same event. Therefore, the data is further processed, and the icing event is marked as no more than half an hour. The same icing event. Finally, the distribution of fault markers in the entire data set is shown in Figure 4:

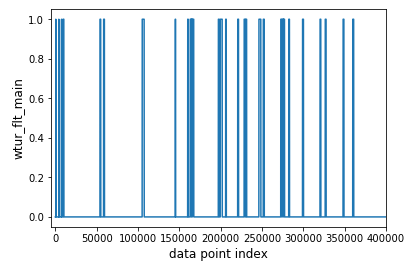


Figure 4

It can be seen from the figure that the fault time period is small. Like other industrial fault diagnosis data, the fan fault data also has strong non-uniformity for the classified positive and negative labels, so further equalize the processing and remove a part of the normal. The data is such that the normal data is approximately twice the icing failure data.

* + 1. Data Features Selection

Using the pre-processed data, the statistical analysis method is used to analyze the change of the same parameters before and after the fault, and the Pearson correlation coefficient is used to extract the characteristics that are more correlated with the icing fault. On this basis, the normalized data statistics method is used to describe the dynamic characteristics of the relevant factors of the fan normal/fault, and the important feature set of the fan icing fault monitoring is obtained for the subsequent deep learning algorithm. Research reference.

1) Significant analysis of various factors before and after failure

The difference between the monitoring parameters before and after the fault was judged by the ttest2 function in Matlab at 5% significance level. Two faults were randomly selected. The results of the significance analysis are as follows:

Table 2 Significant analysis results of two failures

| **Monitoring parameter** | **T-test result of fault (1)** | **T-test result of fault (2)** |
| --- | --- | --- |
| wind\_speed | 1 | 1 |
| generator\_speed | 1 | 1 |
| power | 1 | 1 |
| wind\_direction | 1 | 1 |
| wind\_direction\_mean | 1 | 1 |
| yaw\_position | 1 | 1 |
| yaw\_speed | 0 | 0 |
| pitch1\_angle | 0 | 1 |
| pitch2\_angle | 0 | 0 |
| pitch3\_angle | 0 | 1 |
| pitch1\_speed | / | / |
| pitch2\_speed | / | / |
| pitch3\_speed | / | / |
| pitch1\_moto\_tmp | 1 | 1 |
| pitch2\_moto\_tmp | 1 | 1 |
| pitch3\_moto\_tmp | 1 | 1 |
| acc\_x | 0 | 1 |
| acc\_y | 0 | 1 |
| environment\_tmp | 1 | 1 |
| int\_tmp | 1 | 1 |
| pitch1\_ng5\_tmp | 1 | 1 |
| pitch2\_ng5\_tmp | 1 | 1 |
| pitch3\_ng5\_tmp | 1 | 1 |
| pitch1\_ng5\_DC | 0 | 0 |
| pitch2\_ng5\_DC | 0 | 0 |
| pitch3\_ng5\_DC | 0 | 0 |

In the above table, a t-test result of 1 indicates that the monitored parameter has a significant difference at the 5% significance level before and after the icing failure. By sorting out the calculation results, it can be concluded that the relationship between each monitoring parameter and the icing fault is as follows:

Irrelevant factors: Yaw Speed, Pitch ng5 DC (ng5 charger DC current), Pitch Speed (blade speed)

Related factors: Generator Speed, Power (Net Side Active Power), Yaw Position (Yar Position), Temperature (Wind), Wind Speed (Wind Speed), Wind Direction (Wind Angle), Wind Direction Mean ( 25 second mean wind angle)

2) Correlation analysis between various factors

Furthermore, the Pearson correlation coefficient is used to study the correlation of each factor before and after the fault. In Matlab, the rate of change matrix of the monitoring parameters before and after the fault is calculated, and then the corrcoef function is called to extract the fault (1) and fault ( 2) Several sets of parameters with high correlation rates before and after failure:

Table 3

|  |  |  |
| --- | --- | --- |
|  | Under fault (1)  Pierce correlation coefficient | Under fault (1)  Pierce correlation coefficient |
| yaw\_position& pitch1\_moto\_tmp | -0.740108375 | 0.72406932 |
| yaw\_position& pitch2\_moto\_tmp | -0.549229942 | 0.65828732 |
| yaw\_position& pitch3\_moto\_tmp | -0.596499174 | 0.689253171 |
| pitch1\_moto\_tmp& pitch2\_moto\_tmp | 0.509324928 | 0.746233502 |
| Pitch1\_moto\_tmp& pitch3\_moto\_tmp | 0.610628114 | 0.545749774 |
| Pitch2\_moto\_tmp& pitch3\_moto\_tmp | 0.586951306 | 0.302947918 |
| pitch1\_ng5\_DC& pitch3\_ng5\_DC | -0.830840522 | 0.629916764 |

In addition, the correlation coefficient between some parameters is around 0.3~0.4, and the correlation coefficient between the parameters is 0.01 orders of magnitude or less. Therefore, by obtaining the Pearson correlation coefficient, the parameters with higher correlation can be obtained. The set of parameters listed in the table clearly conforms to the basic operating mechanism of the actual fan.

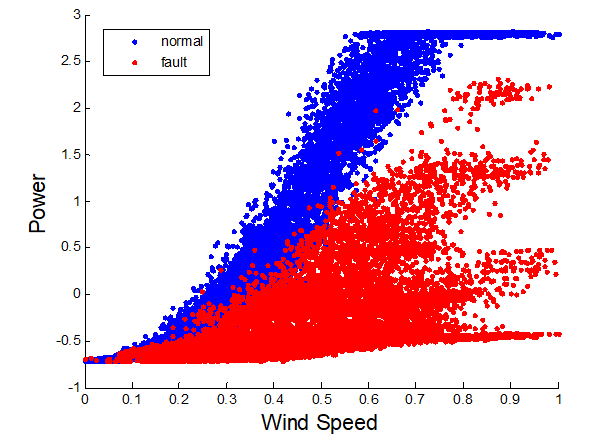
****

图 6 风机正常/结冰时风速——功率关系图

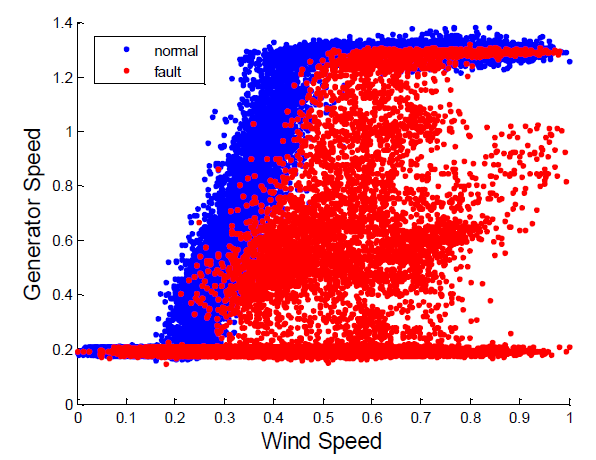


图 7 风机正常/结冰时风速——发电机转速关系图

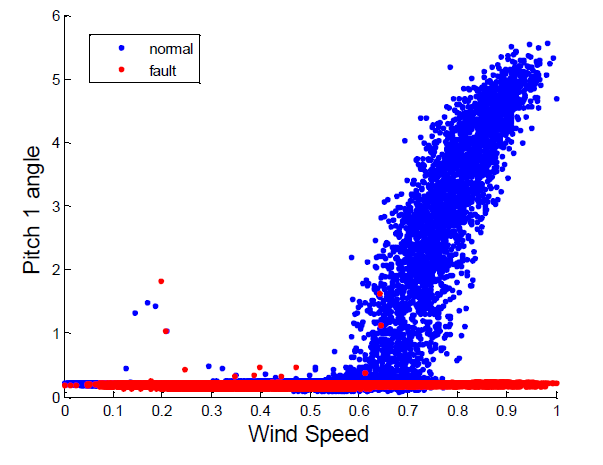


图 8 风机正常/结冰时风速——桨距角关系图

Using LSTM for training, you can solve the problem of long-term sequence dependence and extract the patterns related to the failure from a large amount of time series data. These historical information and states are then processed by the fully connected layer to finally output a classification result vector indicating that the predicted input data points belong to different classes of probabilities. In the application scenario of this paper, the classification result vector is a two-dimensional vector representing the probability of failure and the normal probability. By comparing the two probability values, the one with the higher probability is selected as the classification result, and 1 and 0 respectively represent the fault and normal. The fault prediction result corresponding to the sequence data is obtained. Finally, the real warning point is extracted according to the change pattern of the fault prediction result.Model optimization

1) Adjust and analyze the LSTM deep network hyperparameters to determine the best network for this problem

2) Pre-processing the input with an automatic encoder to analyze the difference in the effect of different types of auto-encoders

3) Combine the mechanism analysis and experimental results, adjust the input of the model to determine the best combination of features

4) Reselection of the traceback time parameter.

Only the current time is more than 10% higher than the accuracy of the past ten minutes Take the results of the optimized feature selection scheme as an example to illustrate the optimized effect:data. Because the judgment of icing is mainly determined by the power difference at the current time, a large amount of past data input causes interference.

Take the results of the optimized feature selection scheme as an example to illustrate the optimized effect：

We use the 26 dimensions (26 monitoring parameters) of all the original data as the input feature vector as the reference feature selection scheme, and compare it with the final feature selection scheme. The final feature selection includes 19 raw data features manually selected and a newly added 6-hour temperature difference feature, a total of 20 features. Compared with the benchmark scheme, the final feature selection scheme reduces the input dimension, thereby reducing the amount of computation and prediction. It has a higher accuracy rate with similar accuracy, thus reducing the probability of false alarms under normal operating conditions. The accuracy ratios of different feature selection schemes are shown in Table 5.

table 5 Experimental results of different feature selection schemes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Accuracy | Accuracy Rate | Recall | F1 |
| Original feature selection（26 Features） | 0.8939 | 0.8763 | 0.9540 | 0.9135 |
| Final feature selection（20Features） | 0.8857 | 0.9410 | 0.8591 | 0.8982 |

* + 1. Early Warning Rules

The two-category results output from the model and the predicted values of the failure probability are not directly output as prediction results because there is a certain degree of jitter. In order to avoid alarm volatility, it is necessary to anti-shake the alarm signal. To this end, we have designed a set of early warning rules to deal with the predicted output of the model. Specifically, it takes the past period of time as the observation window. If the alarm point exceeds a certain percentage, the alarm is performed, similar to the voting mechanism. Alarm points are those that exceed the classification threshold. If a certain percentage is exceeded, it is considered to be a real need for failure warning.

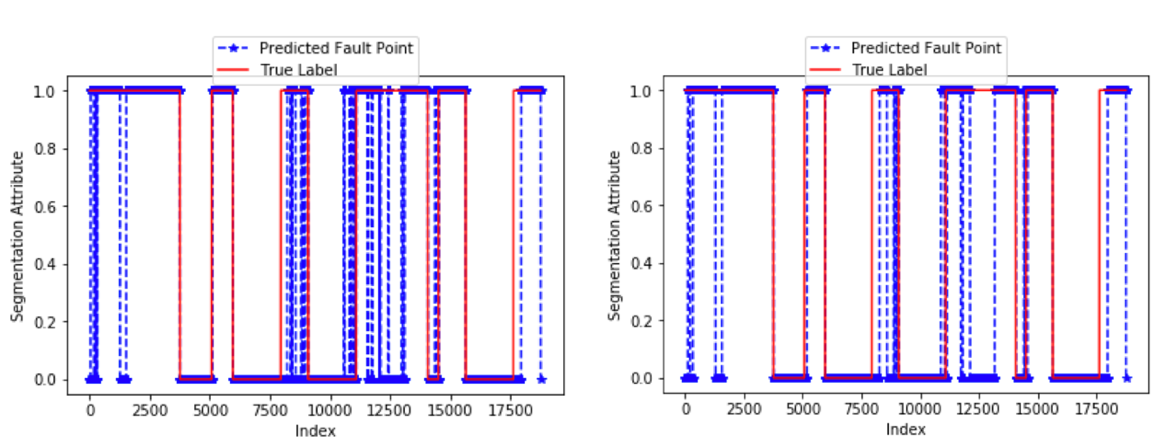
Below we use two examples to illustrate, the observation window of these two examples are 5min and 15min, respectively, the ratio is 90%. Observe the effect of the size of the window on the anti-shake effect: take the past 5 minutes (left), the past 15 minutes (right). The ratio is set to 90%. It can be seen that the window time becomes longer and the jitter is reduced.

Figure 17 Prediction results of different observation windows

In order to reduce the false positive rate, we also set a classification threshold. By appropriately increasing the classification threshold, the false alarm rate can be reduced, that is, the probability threshold for classifying the data as a fault is appropriately increased, and the recall rate is sacrificed to improve the accuracy rate. The experimental results of the prediction of different failure probability thresholds are shown in Table 6.

Table 6 Prediction experiment results of different failure probability thresholds

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Threshold | Accuracy | Accuracy | Recall | F1 |
| 0.5 | 0.8939 | 0.8763 | 0.9540 | 0.9135 |
| 0.55 | 0.8933 | 0.9144 | 0.9028 | 0.9085 |
| 0.6 | 0.8719 | 0.9440 | 0.8311 | 0.8840 |

1. **References**
2. 2017年中国风电行业发展现状分析及未来发展趋势预测[EB/OL]. 2017. <http://www.chyxx.com/industry/201703/502764.html>.
3. 禹迅. 风力发电机叶片结冰检测系统研究开发[D]. 华中科技大学, 2013.
4. 陈阳. 风电叶片除冰方法理论研究[J]. 天津科技, 2015, 42(7): 29-30.
5. 蒋立波. 憎水性涂料对风机叶片结冰影响研究[J]. 涂料工业, 2015, 42(8): 61-66.
6. 王鹏. 基于超声导波方法的风机叶片覆冰检测[D]. 哈尔滨工业大学, 2016.
7. Kusiak A, Verma A. A data-mining approach to monitoring wind turbines[J]. IEEE Transactions on Sustainable Energy, 2012, 3(1): 150-157.
8. Andrew Kusiak, Li W. Virtual Models for Prediction of Wind Turbine Parametes[J]. EnergyConversion IEEE Transactions on, 2010, 25(1): 245-252.
9. AnoopVerma, Andrew Kusiak. Predictive analysis of wind turbine faults: a data mining approach[J]. Proc.ind.eng.res.conf, 2011.
10. 杭俊, 张建忠, 程明, 等. 风力发电系统状态监测和故障诊断技术综述[J]. 电工技术学报, 2013, 28(4): 262-271.
11. 刘辉海,赵星宇,赵洪山,宋鹏,邓春. 基于深度自编码网络模型的风电机组齿轮箱故障检测[J]. 电工技术学报, 2017, 32(17): 156-163.
12. World Meteorological Organization, 1988. General meteorological standards and recommended practices, Appendix A, WMO Technical Regulations, WMO-No. 49.
13. Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
14. Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. science, 2006, 313(5786): 504-507.
15. Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.