International Core Journal of Engineering Volume 5 Issue 9, 2019
ISSN: 2414-1895 DOI: 10.6919/ICJE.201908_5(9).0004

Analysis of Fault Data of Fan Blades in Power Grid

Yanan Yang [a], Youchan Zhu [b]
School of North China Electric Power University, Baoding 130600, China.
[a]1258620225@qq.com, [b]yc_hd@sina.com

Abstract

Wind energy is a clean renewable energy source that plays an important role in power generation. The blade is the main connecting component of wind turbines to convert wind energy into electrical energy, and its health status is highly valued by related industries.The light gradient boosting machine (LightGBM) is used to analyze the SCADA data in the grid to find out the characteristics related to the fan blade failure in the SCADA data.Comparing the normal data and fault data of these features, it is found that the fan blade state is different at normal time and fault time, which lays a foundation for wind turbine blade failure warning.

Keywords

Fan blade failure; Data analysis; LighGBM; Feature detectione.

1. Introduction

As a renewable and clean energy source, wind energy has the advantages of no pollution and unlimited utilization. With the maturity of wind power technology, more and more wind turbines have been put into use. At present, there are nearly 30 wind farms under construction in China. The installed capacity of wind power in 2020 is planned to be 20 million to 30 million kW [1].Wind turbines have a harsh working environment, such as excessive wind, snow, ice, thunder, etc. These extreme weather and climate will affect the service life of wind turbine blades.Existing wind turbine blade fault monitoring often obtains fan blade data by collecting fan blade vibration frequency [2], collecting aerodynamic noise around the blade [3], and other methods.However, these methods often require additional sensors placed around the wind turbine blades, which invisibly increases the cost of fan blade failure monitoring. The grid uses the SCADA system to obtain timely and comprehensive information on the operating status of the wind turbine unit [4]. The SCADA system stores a large amount of data information during the operation of the wind turbine.

This paper analyzes the existing SCADA data in the power grid. Use the LightGBM method to analyze features related to wind turbine blade failures and rank these features by importance. And analyze the difference between the normal data and the fault data of the fan in these characteristics.

2. Causes of Fan Blade Failure

Blade is an important component of wind turbine to convert wind energy into mechanical energy. The blade is driven by wind energy to transfer mechanical energy to the transmission system in the engine room, and then to the generator through the transmission system. Finally, the wind energy is converted into electricity by the generator.

Blade failures of wind turbines are classified into observable faults and unobservable faults. Observable faults are mainly visible faults. These faults have caused serious damage to the blades of the wind turbines, such as damage caused by lightning strikes. The unobservable faults need to be

ISSN: 2414-1895
DOI: 10.6919/ICJE.201908_5(9).0004

detected by instrument inspection, such as the aging of the blade material and the internal corrosion of the blade.

Because the blade is exposed to high altitude for a long time, it is easy for the blade to be struck by lightning at some high altitudes. Lightning will cause blade burns, which will cause blade breakage in severe cases. The wind in nature is uncertain, and its speed and direction will change. The uneven force acting on the blade and the excessive impact of wind speed on the blade are the reasons for the failure of the blade.In addition, in some coastal areas, humid air and salt in the air will cause corrosion to the blades and aggravate the damage of the blades.

3. LightGBM Technology

LightGBM is an efficient gradient decision tree [5]. In January 2017, Microsoft released the first stable version of LightGBM. Compared to traditional gradient boosting decision tree(GBDT), such as XGBoost and pGBRT, LightGBM introduces two new technologies: gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) [6].These two new technologies make LightGBM nearly 20 times faster than conventional GBDT in processing data with high feature dimension and large amount of data.The traditional distributed gradient descent tree traverses all data for each feature and calculates information gain for every possible segmentation point.

The traditional gradient boosting decision tree traverses all the data of each feature and calculates the information gain of each possible segmentation point when processing the feature.

When GOSS sampled the data, it did not apply all the sample points to the calculation, but chose the sample points with large gradient for sampling calculation, and made random sampling calculation for the sample points with small gradient according to the proportion. Because GOSS randomly sampled samples with small gradient, the processed data deviated from the original data. In order to compensate for this deviation, a constant is enlarged for random sampling points when calculating information gain. This can increase the proportion of random sample points in the calculation. The sampling method of GOSS reduces the amount of data needed to be analyzed, while ensuring more accurate information gain, reduces the computational complexity.

EFB binds mutually exclusive features in sparse feature space to form a new feature on the premise that high-dimensional data are always sparse data in practical application. This method reduces the feature dimension, reduces the memory loss and reduces the time complexity of sample processing when searching for the best segmentation point. In practical applications, EFB can get a smaller number of feature bindings by allowing a small number of collisions because it can't get complete mutually exclusive features in a short time. This method further improves the effectiveness of the calculation.

4. Important Feature Extraction of Fan SCADA Data

4.1 SCADA Data Set Preprocessing.

In the SCADA system arranged in the national grid, the environmental parameters, working conditions parameters, state parameters and control parameters of wind turbines are collected in real time. This experiment mainly collects the environmental parameters, working conditions and state parameters of No. 19 wind turbine, each kind of parameters contains a variety of data characteristics. The parameter information of wind turbine is stored in CSV file, and each CSV file contains more than 400 data. There are about 1400 unprocessed CSV files.

In real life, database is vulnerable to noise and some irresistible external factors, resulting in data missing and data errors. Correct data processing can reduce the interference of wrong data to experiment, facilitate computer calculation, and improve the efficiency and accuracy of data mining. The data collected in this experiment are processed as follows:

1) Data consistency check.

ISSN: 2414-1895 DOI: 10.6919/ICJE.201908_5(9).0004

Check whether the data meets the actual requirements, whether the range of feature values is reasonable, and whether it meets the relationship between features. The data with air atmospheric pressure of 0 is deleted, and some data that fail to meet the requirements of fan start-up are deleted.

2) Data cleaning.

Clean incomplete data, error data and invalid data in data sets. Delete some abnormal data of full behavior 0.

Feature transformation.

Feature transformation includes mathematical transformation of features, construction of new features and deletion of invalid features.Delete invalid features that are all listed as 0. The average value, maximum value, minimum value, median value and standard deviation of each feature in the CSV file are calculated. Statistical changes before and after the data in CSV files, and statistical changes in the average data, maximum, minimum, median, standard deviation.

After cleaning, the number of CSV files of No. 19 fan is about 1200, and the number of CSV files is reduced by about 200. This shows that the number 19 fan is in a state of shutdown for a period of time during data acquisition, and the data of fan shutdown is of little significance as an experimental reference.

4.2 Selection of LightGBM Parameters.

First, the selected learning rate is 0.1. When the learning rate is 0.1, the learning speed of the model is higher. Then we use the CV () function of LightGBM to find the best iteration number. The best iteration number n_estimators is 917, with a score of 0.1. Then, the grid SearchCV () function in sklearns is used to find the optimal tree depth and the number of leaf nodes. Finally, the depth value of Max_depth is 5, and the number of leaf nodes num_leaves is 16. The test score is -0.0438. In order to reduce the over-fitting settings, we are going to find the parameters suitable for feature_fraction and bagging_fraction. What we need to note here is that these two parameters need to be set at the same time. The grid SearchCV () function in sklearns is also used. The test results show that the feature_fraction parameter is 0.6, and the bagging_fraction parameter is 0.7. It is divided into - 0.266. The best number of iterations for the new test is 1000 and the score is 0.12. The final parameters are shown in Table 1.

Table 1 LightGBM parameter settings

LightGBM Parameters	parameter values
boosting_type	gbdt
objective	binary
learn_rate	0.1
Max_depth	5
num_leaves	16
bagging_fraction	0.7
features_fraction	0.6
n_estimators	1000
min_child_weight	1

The important features that are ultimately obtained using lightGBM are arranged in order , see Fig. 1. The top 15 characteristics are selected, among which the characteristics related to the pressure of the inverter are most related to the blade failure, followed by the fan blade angle and the wind direction angle.

ISSN: 2414-1895
DOI: 10.6919/ICJE.201908_5(9).0004

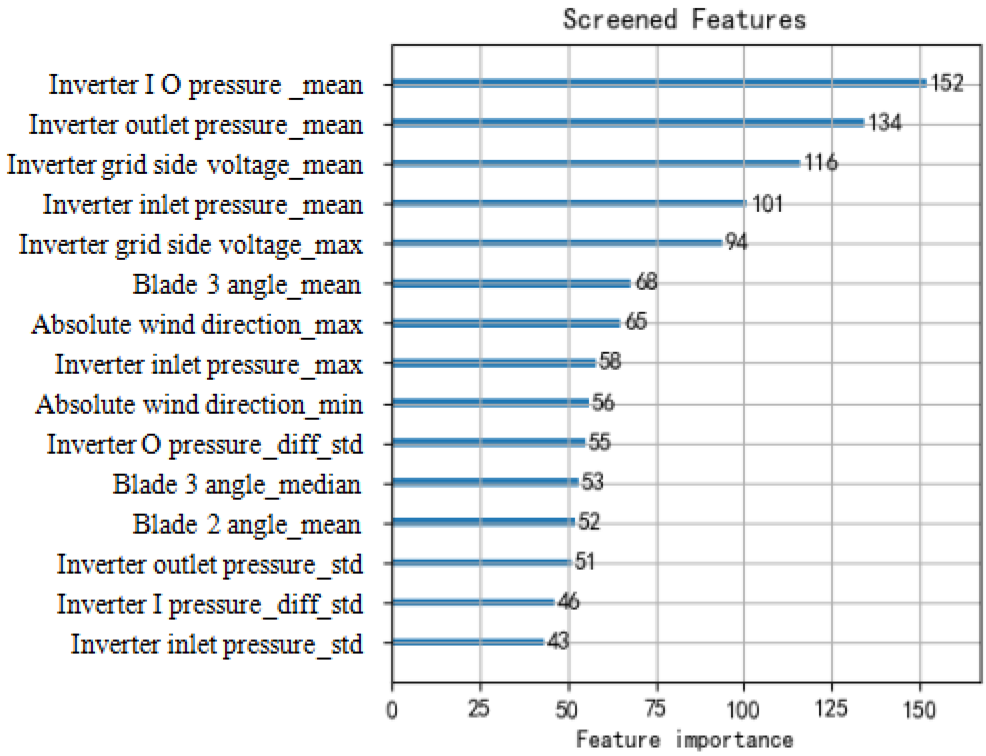

Fig. 1 Feature importance ranking

5. Comparison of Normal Data and Fault Data

Taking the inverter inlet and outlet pressure characteristic and the inverter grid side voltage characteristic as examples, look at the difference between the fault data and the normal data in the features selected by LightGBM. Fig. 2 shows the inverter inlet and outlet pressure data. In Fig. 2, inverter I O pressure means inverter inlet and outlet pressure. The inverter inlet and outlet pressure is obtained by subtracting the value of inverter outlet pressure from that of inverter inlet pressure. From Fig. 2, it can be seen that the fluctuation range of normal data is significantly larger than the fault data, and the fault data values are mostly concentrated near -2.1. Fig. 3 shows the inverter gride side voltage data. In Fig. 3, it can be found that the value of the fault data mostly exceeds the value of the normal data. The data in Fig. 2 and Fig. 3 have a good degree of discrimination.

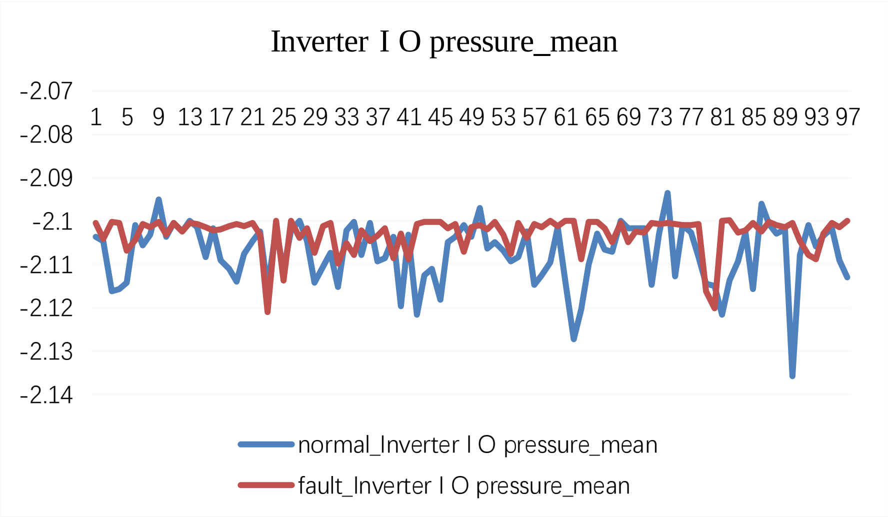

Fig. 2 Inverter inlet and outlet pressure data

ISSN: 2414-1895
DOI: 10.6919/ICJE.201908_5(9).0004

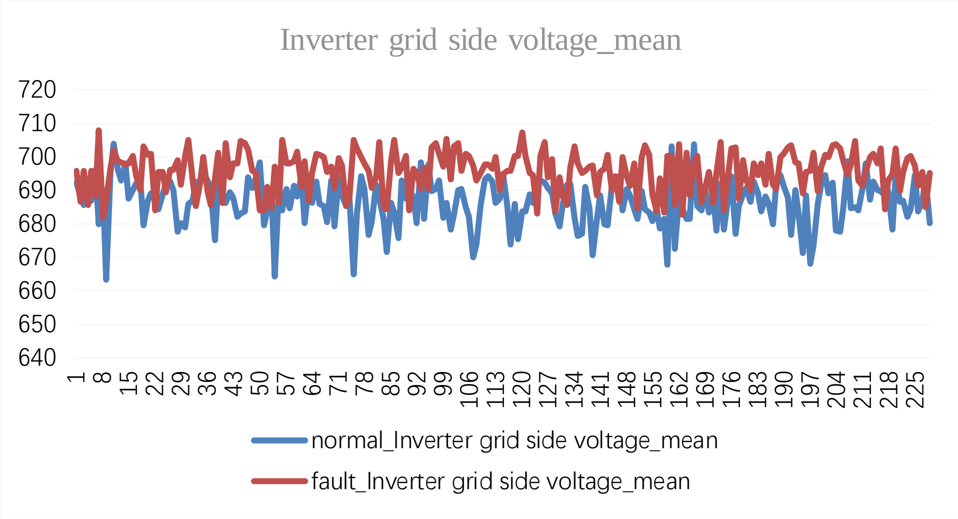

Fig. 3 Inverter gride side voltage data

6. Conclusion

In this paper, LightGBM was used to process the fan generator SCADA data, and 15 features related to fan blade failure are screened out. By analyzing the top-ranked feature data, it was found that the fault data is different from the normal data under the feature. The fault data under these characteristics has a good distinction from the normal data. Using these features as input features of the wind turbine blade fault diagnosis model will be more conducive to wind turbine blade fault diagnosis.

References

[1] Wang Zhijie, Xu Yufa, Liu Sanming, et al. [M] State monitoring and intelligent fault diagnosis of large-scale fan-generator units. Shanghai: Shanghai Jiaotong University Press, 2013.

[2] Zhang Baoqin, Lei Baozhen, Zhao Linhui, et al. Vibration method for fan blade fault prediction [J]. Journal of Electronic Measurement and Instrument, 2014, 28 (3): 285-291.

[3] Yan Jingzhong, Zhou Jin, Ji Cuina, et al. Preliminary study on detection of blade cracking defects of wind turbines based on aerodynamic noise [J]. Instruments and analysis monitoring, 2018, 2:1-5.

[4] Chen, N., Yu, R.R., Chen, Y.:Comparison on wind turbine condition monitoring methods using SCADA data[J]. Wind Integration Workshop, 2015,no. 025.

[5] Guolin Ke , Qi Meng , Thomas Finley ,el. LightGBM: A Highly Efficient Gradient Boosting Decision Tree[C].31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

[6] Zhang Danfeng, Prediction of Wind Blade Icing Based on LightGBM, XGBoost and ERT Mixed Model [D]. Shanghai Normal University, 2018.