A Data-Mining Approach for Wind Turbine Power Generation Performance Monitoring Based on Power Curve

Jianlou Lou¹, Heng Lu¹, Jia Xu² and Zhaoyang Qu¹,

¹ Department of Information and Engineering, Northeast DianLi University, Jilin 132012, China
² Long Yuan(Beijing) Wind Power Engineering Technology CO.,LTD, Beijing 100034, China
loujianlou@qq.com

Abstract. A data-mining approach is proposed to investigate the power generation monitoring of wind turbine based on power curve profiles in this paper. The weakened power generation performance could be identified by this method through assessing the wind-speed power datasets. Shapes of wind power curve profiles over consecutive time intervals are constructed by fitting power curve models into wind-speed power datasets. In this research, an optimal constraint in each sub-dataset is developed for governing the data-driven wind-power generation method based on distance-based outlier detection and variance analysis model. The Auto-adapt Optimal Interclass Variance algorithm realize the self-optimization of the threshold parameter and achieves a high degree of robustness to the variations in wind-power generation performance monitoring. The blind industrial researches are conducted to validate the effectiveness of the approach, and shows the decrease of error rates when detecting weakened power generation performance or causing financial loss.

Keywords: Wind turbine; Power curve; Data-mining; Performance monitoring;

1 Introduction

The power curve reflects the operation performance of the wind turbine [1]. It is often influenced by air density, system control, ambient temperature so that raw data collected by SCADA system usual contains lots of anomalies. But traditional data-mining methods which detect weakened performance through building model training of the datasets with certain ratio generally have inevitable prediction error, so the identification of the poor power generation performance of turbines is not accurately for the real-time data [3-5].

¹ Project supported by the National Natural Science Foundation of China (No. 651277023) and Jilin Province Science and Technology Development Project (20150204084GX).
Corresponding author: Jianlou Lou.
Email: loujianlou@qq.com.

ISSN: 2287-1233 ASTL
Copyright © 2016 SERSC
In this paper, a data-mining approach is proposed to assess the wind power generation performance through analyzing the variation of wind power curves rather than individual data points. The data is partitioned into sub-datasets based on consecutive equal time-intervals. Then the outlier-detection approach and variance analysis model are used to realize the self-optimization of threshold in the Auto-adapt Optimal Interclass Variance (AOIV) algorithm. The effectiveness of this approach is demonstrated through some blind industrial studies.

2 Data-Mining Based on Power Curve

2.1 Data Preprocessing

Before mining the power curve profiles, the data are preprocessed according to the methodology used in wind power generation enterprise of China. That encompasses the three following steps: 1) validity check, 2) data range check, and 3) missing data processing. Moreover, these anomalies should be removed to make a separate analysis if necessary. Actually, the simplified rules as follows are often used when taking into account time and high-efficiency.

a) wind speed < cut-in speed—the cut-in speed is the wind speed value at which the turbine starts operating; pitch angle about 90°.

b) wind speed between cut-in speed and cut-out speed—power output zero or negative;

c) wind speed > cut-out speed—the cut-out speed is the wind speed value at which the turbine stop generating available power; pitch angle about 90°.

2.2 OIV Algorithm

The optimal interclass variance (OIV) algorithm based on the power curve is a simple and efficient data-mining method for the power curve analysis and detect anomaly by combined with the initial variance threshold. The algorithm is as follows.

Given a sample dataset of power curve $U = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, and satisfy $y_i < y_{i-1}$, $i \in (2, n)$, $x$ represents wind speed, $y$ is expressed as power, and $n$ is the total number of sample points. The profile for depicting the characteristic of the power curve contained in $U$ is $\lambda$ if and only if $\lambda$ satisfies (1).

$$\lambda = \arg \max \left\{ \sum_{i=2}^{n} \left( \frac{1}{\lambda} \sum_{j=1}^{\lambda} (y_i - \overline{y_\lambda})^2 \right) < S \right\}$$

(1)

Where $y_j$ is the $j$th power value; $\overline{y_\lambda}$ is the average of first $\lambda$ power value; $\lambda$ is constant; $S$ is initial threshold.
Each sub-dataset could be classified by $S$. And the results of similar dataset are finally summarized into normal and abnormal.

### 2.3 AOIV Algorithm

The auto-adapt optimal interclass variance (AOIV) algorithm is proposed to enhance the accuracy with the combination of the outlier detection and the optimization of the variance threshold in this research.

#### A. The Detection of Outliers

The outlier detection approach based on the distance is effective to detect power curves with different curvature. And the detected outlier should be removed and belongs to the category of the abnormal dataset. $u_i = (x_i, y_i)$ represent a data point in the sample set $U$, The distance between two points is defined:

$$d_k(u_i, u_{i-1}) = (|x_i - x_{i-1}|^p + |y_i - y_{i-1}|^p)^{1/p}$$  \hspace{1cm} (2)

**Definition 3.1** For any point $u_i = (x_i, y_i)$ in $U$, given a relative small positive number $\varepsilon$, if any point $u_i = (x_i, y_i)$ in data set $U$ satisfy with condition: $d_k(u_i, u_{i-1}) < \varepsilon$, so $u_{i-1}$ is the $\varepsilon$-proximal point of $u_i$, and the set of all $\varepsilon$-proximal points is $\varepsilon$-neighborhood of $u_i$.

**Definition 3.2** For any point $u_i = (x_i, y_i)$ in $U$, given a relative small positive number $\varepsilon$, select an empirical critical value $N_0$. Assume that the number of $\varepsilon$-neighborhood of $u_i$ is $N_i$, if $N_i < N_0$, the $u_i$ is called an isolated point of $U$.

#### B. The Optimization of Variance Threshold

The improved model of variance analysis achieves the optimization of threshold in each sub-dataset. Given a set of sliding variance samples is $Z = \{z_1, z_2, \ldots, z_k\}$, $k$ represents the number of sample points. To define the range of $S$ is $[s_1, s_2]$, and satisfy with $S \in N_0$. The default value is $s_1 = 1$, $s_2 = 500$. The sample set $Z$ is divided into two groups by selecting different values of $S$ in turn, and the following calculation is performed if and only if exist two groups, otherwise reselect next $S$ value. Assume that the two groups of data in an specific $S$ are $Z_1 = \{z_1, z_2, \ldots, z_{1}\}$ and $Z_2 = \{z_2, z_{k-1}, \ldots, z_k\}$, and the internal error and external error between $Z_1$ and $Z_2$ is calculated in (3) and (4).

$$\sigma = \sum_{j=1}^{k} (z_j - \overline{Z_1})^2 + \sum_{j=1}^{k} (z_j - \overline{Z_2})^2$$  \hspace{1cm} (3)

$$\omega = (\overline{Z_1} - \overline{Z})^2 \lambda + (\overline{Z_2} - \overline{Z})^2 \ast (k - \lambda)$$  \hspace{1cm} (4)

Where $\sigma$ presents external error; $\omega$ presents internal error; $z$ presents sliding variance; $\overline{Z_1}$ presents the mean value of $Z_1$; $\overline{Z_2}$ presents the mean value of $Z_2$;
Z̄ presents the mean value of Z; S is the optimal variance threshold for the sub-dataset when only satisfy (5).

\[ S = \arg \max_{S} \left( \sum_{\omega_k \in S} \sigma_{\omega_k} / \omega_k \right) \]  

(5)

In order to avoid the error detection in the normal operation mode of the wind turbine, S usually need to join a certain threshold supplementary quantity \( \mu \).

The block diagram of the AOIV algorithm is as follows:

![Block Diagram of AOIV Algorithm](image)

Fig. 1. Block diagram of AOIV algorithm.

### 3 Industrial Studies

#### 3.1 Data Preparation

In order to verify the effectiveness of the approach and the application of the power generation performance monitoring, this paper investigates the 10-min SCADA data of 33 wind turbines collected from May 31, 2013 to May 8, 2013 in a 100MW Class wind farm in China. The collected data contained values of wind turbine performance parameters, wind conditions, as well as the fault logs.

#### 3.2 Contrastive Analysis between Algorithms

To compare the data-mining effect between two algorithms in chapter 2, this section presents some industrial studies based on turbines’ 10-min SCADA data which collected randomly from a wind farm in China. The data-mining results with two algorithms are shown in figure 2 and the performance comparisons are shown in table 1.
Fig. 2. Power curve with normal data (cross) and abnormal data (circle).

Left: OIV algorithm; right: AOIV algorithm.

Figure 2 shows the result of two algorithms. Obviously, the AOIV algorithm has better ability of data-mining.

Table 1. Contrastive Analysis Table

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Artificial</th>
<th>OIV</th>
<th>AOIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND</td>
<td>1341</td>
<td>1165</td>
<td>1370</td>
</tr>
<tr>
<td>LD</td>
<td>426</td>
<td>662</td>
<td>397</td>
</tr>
<tr>
<td>HD</td>
<td>938</td>
<td>938</td>
<td>938</td>
</tr>
<tr>
<td>EDoN</td>
<td>0</td>
<td>32</td>
<td>30</td>
</tr>
<tr>
<td>EDoL</td>
<td>0</td>
<td>268</td>
<td>12</td>
</tr>
<tr>
<td>EDoH</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1 shows the data-mining results among the artificial statistics results, OIV algorithm and AOIV algorithm. ND, LD and HD represent the detected data size of normal, limited-power and halt. EDoN, EDoL and EDoH represent the error-detected data size of normal, limited-power and halt. We can see both of normal data and limit data have different degrees of error detection with OIV algorithm. In addition, the method of detecting halt-data are same law in chapter 2.1 so that the amount of error detected halt-data is zero. But after improving the algorithm, data-mining results has a lower error rate and higher accuracy.

3.3 Performance Monitoring of Turbines

A. The Data-Mining with AOIV Algorithm

Taking into account the number of test turbines, here is only given some data-mining results of randomly selected turbines from 15081501 to 15081533. The results are shown in figure 3.
From Figure 3 it can be seen that this approach can effectively detect the various power curve profile. Among them, the turbines 15081505 and 15081507 processing results are better, but there are error-detection in the rated wind speed zone, which is due to the amount of data is too small to be mistaken for an isolated point, it can be neglected. According to the above, this approach algorithm has better versatility and accuracy.

B. The Analysis of Detected Faults

Through the analysis of identified anomalies by data-mining, the root cause of poor power generation performance of turbine will be found. Taking turbine 15081515 for example to describe the detailed analysis procedure. The irregular profiles are shown in Figure 4.

In Figure 4, the turbine 15081515 does not generate power normally so that power curves display abnormal curvatures. Actually, this is because of the pitch faults which usually seriously affect power generation performance of turbines. The turbine should be maintained as soon as possible.

4 Conclusion

In this paper, we proposed a data-mining method to identify impaired power generation performance by analyzing the curvature and shape of the wind power curve. The outlier detection based-on distance were applied to eliminate the impact of
isolated points in the power curve analysis, while the variance analysis model realize
the self-adaptive threshold to enhance the accuracy rating of data-mining. Compared
to other methods, this way could more accurately and fleetly detect the anomalies of
turbines just by analyzing power curve without complex sample training. It can also
process the real time data for online monitoring and identify the weakened
performance of generating unit.

Some industrial studies were conducted to prove the effectiveness and accuracy of
the algorithm. We have mined the power curve of 33 sets of wind turbines, and the
great data results are obtained. The future research will investigate an intelligent
method to identify the fault type and the root cause with poor performance of
turbines. In addition, the relationship between wind power and multiple parameters
will be considered in the data mining.

References

1. Schlechtingen, M., Santos, I. F., & Achiche, S.: Wind turbine condition monitoring based on
   Computing, 13(1), 259-270 (2013)
   Energy, 34(6), 1487-1493 (2009)
turbine power curve monitoring: a comparative study. Sustainable Energy, IEEE
5. Sanchez, H., Escobet, T., Puig, V., & Odgaard, P.: Fault Diagnosis of Advanced Wind
   Turbine Benchmark using Interval-based ARRs and Observers. In World Congress, vol. 19,
   No. 1, pp. 4334-4339 (2014)