

A Fast Abnormal Data Cleaning Algorithm for Performance Evaluation of Wind Turbine

Zhongju Wang , Long Wang , Member, IEEE, and Chao Huang , Member, IEEE

Abstract. In this paper a method to remove abnormal data from wind turbines is proposed. This method is divided in data cleaning and data classification. In the former, pixels corresponding to normal data are extracted based on the characteristics of normal and abnormal data. In the later step images are classified based on the existence of the corresponding pixels. The data cleaning is implemented parallelly using CUDA. The proposed method's effectivity is validated using 37 commercial wind turbines. This paper's results show an increase in performance cleaning abnormal data while reducing computational time.

Keywords: Abnormal data cleaning, data-driven approaches, wind power curve (WPC), wind turbines.

1 Introduction

Monitoring wind turbine power generation performance is conducive to arranging maintenance plans reasonably, preventing failure occurrences, and minimizing O&M cost. To measure the wind turbine power generation performance, accurate wind power curve modeling is widely applied. With the development of information technologies, current wind turbines are generally equipped with data acquisition and monitoring systems. Supervisory control and data acquisition systems that accumulate vast historical data of wind turbines are widely used for various condition monitoring tasks of wind turbines, such as fault diagnosis and failure detection. WPC modeling-based methods are another main type of method for abnormal data cleaning. Based on the spatial distribution characteristics, wind power data points distributed outside the boundary of WPC are considered as outliers. Verma utilized a five-year historical data set to fit the WPC and applied the k-means clustering algorithm to distinguish normal wind power data points and outliers. Identified abnormal wind power data points based on the probabilistic WPC and corrected the detected outliers via the spatial correlation characteristics of neighboring wind farms. Considering abnormal data affects the error distribution in WPC modeling, Wang et al. proposed two novel asymmetric spline regression models, mixture of asymmetric Gaussian and mixture of asymmetric exponential power, for solving the problem of complex and asymmetric error distribution in WPC modeling. However, these methods required a selected training data set to fit WPC models so that abnormal data cleaning performance was influenced by the selection of the training data set. In the proposed method, the principal part of the WPC image was extracted via MMO, and the mapping relationship between wind power data points and WPC image pixels was established to mark the normal and abnormal data. Here is proposed an image thresholding method for WPC abnormal data cleaning. The proposed method filtered the abnormal data by thresholding on a gray level feature image. In

conclusion, existing abnormal data cleaning methods for wind power data have the following limitations: 1) high computational overhead is required; 2) cleaning performance is seriously influenced by the selection of training data; and 3) hyperparameters need to be determined for different data sets. Considering the features of WPCs, the data cleaning stage filters abnormal data based on pixel spatial distribution characteristics of abnormal and normal data in WPC images.

2. Proposed Method

In this section, the overall procedures of the proposed method are first introduced. Next, the details of the WPC normal data extracting operation executed on CPU and GPU are described. The main steps of the proposed algorithm are shown in Fig. 1, including:

- 1) Convert the scatter figure of wind power data to a WPC binary image
- 2) Clean abnormal pixels in the WPC binary image using pixel operations.
- 3) Classify the wind power data point based on the existence of corresponding pixels in the processed WPC image.
- 4) Evaluate power generation performance via WPC models based on the extracted normal data.

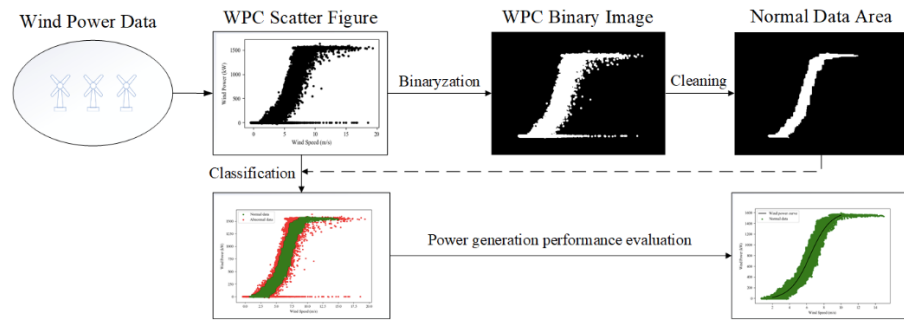


Figure 1

The algorithm used to convert the scatter image into a WPC binary image is as follows in pseudocode:

```

Input: Pixel value of WPC image  $f(x, y)$ 
Output: Pixel value of the extracted image  $f_{\square}(x, y)$ 
Initialize  $g(x, y) := 0, f_{\square}(x, y) := 0$ 
for  $j := 1$  to  $n$  do for  $i := 1$  to  $m$  do
     $U_{k,j} = \{(sk,j,ek,j) | f(i,j)=255, 0 < sk,j < i < ek,j < m\}$  end
     $dbest,j = \max\{e_{1,j} - s_{1,j}, e_{2,j} - s_{2,j}, \dots, e_{t,j} - s_{t,j}\}$  for
 $k := 1$  to  $t$  do
    if  $ek,y - sk,y = dbest,y$  do  $g(sk,y : ek,y, j) := 255$ 
    end end
end
for  $i := 1$  to  $m$  do
for  $j := 1$  to  $n$  do
 $V_{k,i} = \{(sk,i,ek,i) | g(i, j) = 255, 0 < sk,i < j < ek,i < n\}$ 
end
 $dbest,i = \max\{e_{1,i} - s_{1,i}, e_{2,i} - s_{2,i}, \dots, e_{t,i} - s_{t,i}\}$  for
 $k := 1$  to  $t$  do
    if  $ek,i - sk,i = dbest,i$  do  $f_{\square}(i, sk,i : ek,i) := 255$ 
    end end
end

```

3. Conclusion

This paper proposed a GPU-accelerated WPC abnormal data cleaning algorithm based on image processing for wind turbine power generation performance evaluation. The proposed algorithm included two stages, data cleaning and data classification. To further improve the execution speed, the data cleaning operation was executed in parallel on GPU via CUDA. The normal data points detected by the proposed method were applied to power generation performance evaluation. The main findings of this study are summarized as follows:

- 1) The proposed algorithm outperforms the MMO and LOF algorithms in terms of the highest detection rate and the shortest computing time.
- 2) The proposed algorithm can be further accelerated by using CUDA on GPUs and at least a 45× speedup ratio is obtained based on an entry-level GPU (Nvidia GTX 1050Ti).
- 3) Wind power data filtered by the proposed algorithm can be utilized to accurately measure the power generation performance.

In this study, only wind speed and power outputs are employed to evaluate wind turbine power generation performance. In future work, more operational parameters of wind turbines will be considered to develop a generation performance indication model. Meanwhile, the method will be further developed to eliminate the influences of outlier holes in WPC images.

References

1. Zhongju Wang , Long Wang , Member, IEEE, and Chao Huang , Member, IEEE.: A Fast Abnormal Data Cleaning Algorithm for Performance Evaluation of Wind Turbine. (2021)