Solar Energy 164 (2018) 139–148

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Local outlier factor-based fault detection and evaluation of photovoltaic system

Hanxiang Ding[a], Kun Ding[a,*], Jingwei Zhang[b], Yue Wang[a], Lie Gao[a], Yuanliang Li[a], Fudong Chen[a], Zhixiong Shao[c], Wanbin Lai[a]

[a] College of Mechanical & Electrical Engineering, Hohai University, Changzhou, Jiangsu 213022, China
[b] Changzhou Key Laboratory of Photovoltaic System Integration and Production Equipment Technology, Changzhou, Jiangsu 213022, China
[c] Sunshore Solar Energy Co., Ltd, Jiangsu Province 213022, China

ARTICLE INFO

Keywords:
Photovoltaic system
Fault detection
Fault degree evaluation
Modified local outlier factor

ABSTRACT

Aiming at monitoring photovoltaic (PV) systems, evaluating the degree of fault and locating the fault automatically under different outdoor conditions, this paper discusses a new procedure of fault detection and evaluation of fault degree of the PV system. For the PV array connected by PV modules in series and parallel, each string shares the same voltage. The value of current can be used to identify the underperformed strings. In addition, considering the non-stationary stochastic characteristics of current of PV strings, the local outlier factor (LOF) is applied to detect the fault in PV system by evaluating the deviation between the observed data and the whole data. Nevertheless, the LOF method is more suitable for large samples and the LOF value varies with the value of string current. Hence, the conventional LOF method is not suitable for evaluating the fault degree. In order to apply this method to different scale PV systems to detect the fault accurately and evaluate the fault degree, a modified algorithm is proposed in this study. The simulations and experiments based on the model of PV array in MATLAB/Simulink and the 10 kWp PV power plant built on the campus of Hohai University are implemented. The results of experiments reveal that the modified LOF has good performance in fault detection and fault degree evaluation in different scales of the PV systems.

1. Introduction

Compared to the traditional thermal power plants, photovoltaic (PV) power plants utilize solar energy to produce electricity without any environmental pollution. In recent decades, with the great progress of the technology for PV systems and conversion efficiency of PV modules, the PV system attracts more attention all over the world. More and more countries tend to use this method to produce electricity (Shafiei et al., 2009; Raturi et al., 2016; Kim et al., 2016). According to the global status report in 2015, the overall installed capacity of PV systems is 227GW (Datas and Linares, 2017; Pérez-Higueras et al., 2011). With the increment of the total installed capacity of PV power plants, the faults in PV system are more and more reported (Falvo and Capparella, 2015). Those faults cause not only power loss, but also the fire hazards (Capparella and Falvo, 2014; Chen and Li, 2016). However, the faults often occur after PV system operated for a long time. To avoid the accidents and enhance the efficiency of the PV system, the degradation (Ndiaye et al., 2013, 2014; Laronde et al., 2010), fault and performance of PV power plants are focused on by researchers.

At present, many methods have been presented. On the one hand, in the field of fault detection of PV system, some methods are based on the comparison between the measured value and the output of simulation model of PV array (Chao et al., 2008; Hu et al., 2013; Firth et al., 2010; Chouder et al., 2013). The method needs an accurate model of PV module. Corresponding calculation of simulation costs much time. In addition to this, deviation of threshold value is not discussed. So the real-time performance is worse. Some statistical methods are also applied, e.g. methods based on the 3 sigma criteria and the boxplot outlier rule (Zhao et al., 2013; Zhao et al., 2014). Moreover, some intelligent algorithms, e.g. neural network and fuzzy logic, are used to identify the faults (Bonsignore et al., 2014; Karaköse and Firildak, 2015; Mekki et al., 2016). However, for reconfiguration technique, it requires massive number of switches as the array size increase, which is expensive and it detects only the module level faults and bypasses the faulty module (Stellbogen, 1993). For power comparison method, it needs a complex simulation model of PV system and calculation of array, it is not suitable for large PV systems, since various modules experience discrete irradiance value losses (Hariharan et al., 2016). On the other

* Corresponding author.
E-mail address: dingk@hhu.edu.cn (K. Ding).

https://doi.org/10.1016/j.solener.2018.01.049
Received 25 July 2017; Received in revised form 21 November 2017; Accepted 15 January 2018

hand, in order to evaluate the fault degree of PV system, the performance ratio (PR) is widely used (Ueda et al., 2009; Ketjoy et al., 2013; Rehman and El-Amin, 2012). Nevertheless, the PR is affected by ambient temperature and is not sensitive to slight faults. Above all, it cannot perform well under different weather conditions. Commonly, the higher the PR is, the better the system performance is. However, the PR at low ambient temperatures is higher than that at high ambient temperatures, which may lead to false detection at high ambient temperatures (Kuroda et al., 2014). Additionally, some other methods of evaluation in PV system, e.g. the methods based on machine learning technique, have been presented (Hernández-Moro and Martínez-Duart, 2013; Limmanee et al., 2017). These methods proposed cannot locate the fault position, identify fault reason, and evaluate fault degree, simultaneously. In order to deal with this problem, a novel algorithm is required.

Generally, in PV system, if one PV module is fault, e.g. is shaded, short-circuit, bypassed, or open-circuit, the current of the fault branch will reduce. So the measured current of each string can present the fault degree by comparing the value of current in different strings (Chao et al., 2008; Hu et al., 2013; Firth et al., 2010; Chouder et al., 2013). In order to detect faults and evaluate fault degree simultaneously, the statistics and data mining theory are used in this paper, i.e. the local outlier factor (LOF) method. According to this method, the abnormal data will present some mathematical characteristics (Liu and Deng, 2013). These characteristics can reveal the degree of deviation. In PV system, the degree of deviation of the abnormal data represents the fault degree. However, experimental results show that the LOF method has good performance for large samples (more than 20 samples), but not for small samples (less than 10 samples). In order to apply this method for different scales of PV system, especially the small-scale PV system (the rated power is less than 10kWp, and the number of parallel strings is less than 10), the LOF should be modified to realize the fault detection, fault location and evaluation of fault degree under complex weather conditions, even in the cloudy and rainy days.

This paper is organized as follows: in Section 2, the principle of the LOF is introduced and the main disadvantages of the conventional LOF for fault detection and evaluation of fault degree are illustrated; in Section 3, the modified LOF is proposed and the effectiveness and performance of the modified LOF are verified; in Section 4, experiments of the modified LOF are implemented to validate the performance of the proposed modified LOF; in Section 5, the conclusion is summarized.

2. Principle of the local outlier factor

2.1. Local outlier factor

Outlier detection is an important method for data mining. Generally, the outlier detection can be roughly divided into the following categories, including distribution-based, depth-based, distance-based, clustering-based and density-based outlier detection (Huang et al., 2016). The outlier detection based on distribution, depth or distance uses overall criteria to diagnose the fault. They are not suitable for the special observation set. The deviation between different observations is significant. Hence, the outlier detection based on reachability density has better performance than that based on depth or distance (Pokrajac et al., 2007). The local outlier factor belongs to a density-based method, which is first proposed by Breunig et al. (2000). Accordingly, if the observation deviates from the whole observations, it will be considered as an outlier.

The current of PV array is mainly affected by the irradiance, which is characterized by non-uniform distribution in a day. Thus, the currents of each string are different in different time. If overall criterion is applied for detecting the fault of the current, the misjudgment may occur and the accuracy of fault detection decreases. However, the local criterion has better performance in PV system.

In this section, in order to explain the concept of the LOF, some definitions are introduced first. The set $X \in R^{n\times m}$ is a sampled dataset, n represents the number of samples in the set X, m is the number of variables (Ma et al., 2013).

Definition 1 (*Euclidean distance*).

$$d(p,o) = \sqrt{(p-o)^2} \tag{1}$$

where p, o represent two different observations in X, respectively.

Definition 2 (*k-distance*). The k-distance of p is denoted as $d_k(p)$. If the following two conditions are satisfied:

(1) In the X, at least k observations $o' \in X/\{p\}$, satisfy $d(p, o') \le d(p, o)$.
(2) In the X, no more than $k-1$ observations $o' \in X/\{p\}$, satisfy $d(p, o') < d(p, o)$.

the following equation holds:

$$d_k(p) = d(p,o) \tag{2}$$

where k is a positive integer, which represents the number of neighboring observations.

Definition 3 (*k-distance neighborhood of p*). The k-distance neighborhood of observation p is denoted as $N_k(p)$. It represents the observation within k-distance neighborhood. The observations in the k-distance neighborhood form a new set Q in X. Hence $N_k(p)$ satisfies the following equation:

$$N_k(p) = \{Q \in X/\{p\} | d(p,Q) <= d_k(p)\} \tag{3}$$

where Q is a new dataset, belonging to data set X, but not to p. $d(p, Q)$ represents the Euclidean distance between dataset p and dataset Q.

Definition 4 (*reach-distance of p*). The reach-distance of p with respect to o, which is denoted as $reach\text{-}dist_k(p, o)$, is given by:

$$reach-dist_k(p,o) = \max\{k-dist(p), d(p,o)\} \tag{4}$$

Definition 5 (*local reachability density of p*). The local reachability density of p, i.e. $lrd_k(p)$, is given by:

$$lrd_k(p) = \frac{k}{\sum_{o\in N_k(p)} reach-dist_k(p,o)} \tag{5}$$

Definition 6 (*local outlier factor of p*). The local outlier factor of p ($LOF_k(p)$) is given by:

$$LOF_k(p) = \frac{1}{k} \sum_{o\in N_k(p)} \frac{lrd_k(o)}{lrd_k(p)} \tag{6}$$

Eq. (6) shows that the definition of LOF is the ratio of the average local reachability density of neighborhoods to its local reachability density. Obviously, if the sample is normal, the average local reachability density of neighborhoods approaches to corresponding local reachability density of the sample. As a result, the value of LOF approaches to 1. On the contrary, if the sample deviates from the overall observations, the value of LOF will be much greater than 1.

2.2. Case study

For the fault detection of PV system, the value of output current of PV string is regarded as observations in LOF. In order to verify the performance of this method, a small dataset and a large dataset of measured current are selected respectively. For the small dataset, the dataset x is set as {2.25, 2.23, 1.23, 2.26}, and the number of neighboring observations k is 3. Obviously, in dataset x, the third current value is abnormal. The LOF value of x is shown in Fig. 1(a). From Fig. 1(a), the LOF of the third observation is close to 2.5 and the LOF of other samples approach to 1. For the large data set, the dataset y is set as {2.25, 2.23, 1.23, 2.26, 2.25, 2.20, 2.30, 2.26, 2.20, 2.26, 2.20, 2.26,