Fault diagnosis of photovoltaic array based on SE-ResNet

Shaofeng Fu, Fenghuang Cai*, Wu Wang

Department of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China

*caifenghuang@fzu.edu.cn

Abstract. In recent decades, fault diagnosis of photovoltaic (PV) arrays has become more and more important for the power generation of PV systems. Many traditional artificial intelligence methods have been successfully applied to use fault data samples to build fault diagnosis models, but most rely on manual feature extraction or expert knowledge to build diagnosis models, which is inefficient and may ignore some potentially useful features. Therefore, this paper proposes an SE-ResNet PV array fault diagnosis algorithm based on Residual Network (ResNet) and Squeeze-and-Excitation Network (SENet), and uses Bayesian Optimization (BO) to optimize the parameters. In order to verify the effectiveness of the proposed fault diagnosis algorithm, a comprehensive fault experiment was carried out on the PV platform. Experimental results show that the model has achieved high overall performance in terms of accuracy, generalization performance, reliability and training efficiency, and has high practicality.

1. Introduction

In recent years, PV array fault diagnosis methods have also been rapidly developed. Buerhop et al. [1] detect the mismatch and shading fault of the PV array by calculating the power loss, as well as the sudden change of the DC side power and irradiance. However, when the module parameters are changed, it may cause deviation between the simulation and the actual measurement room, thereby affecting the accuracy of fault identification. Rajasekar et al. [2] proposed an improved algorithm based on bacterial foraging to estimate PV parameters. By determining the value of $V_{oc}$ and $I_{mp}$, the parameters of three different PV models were extracted to accurately simulate PV characteristics. But its reliability depends excessively on the accuracy of the parameter extraction algorithm. Yi et al. [3] proposed a PV system DC side fault detection method based on multiple signal decomposition (MSD) and fuzzy inference system (FIS), using MSD-based signal processing methods to extract features, and applying fuzzy logic rules to detect line faults and ground fault detection. But its reliability still depends on the accuracy of the threshold setting. Kaplaniet al. [4] used infrared thermal imaging technology to identify whether there are hot spots, aging and other faults based on the temperature of the monitoring array in different states. However, the equipment cost is too high, and it is not easy to maintain, and cannot be applied to actual PV systems on a large scale. Chen et al. [5] used a random forest ensemble learning algorithm to identify PV faults. The training process is not prone to overfitting, but the algorithm requires manual feature extraction and the training convergence time is too long.

In response to the above problems, this paper proposes a fault diagnosis model based on SE-ResNet, which considers four different fault conditions, namely partial shading (PS), line-line fault (LLF), open circuit fault (OCF) and normal condition (NORMAL). And use the experimental data set to...
evaluate the performance of the model. Experimental results show that this method can achieve high fault diagnosis accuracy.

2. Experiment and Methods

2.1. PV experiment

Then in order to diagnose the fault of the PV array, an experimental platform with a PV capacity of 6.48kWp was established. The PV array consists of 2 PV strings connected in parallel, and each PV string consists of 12 PV modules in series. The detailed parameters of the PV system are as follows Table 1 shows.

| Table 1 Parameters of the laboratory PV power system |
|-----------------------------------------------|
| PV components | Parameters |
| PV module | $P_{mpp}$ 270.119W |
| | $V_{mpp}$ 31.3V |
| | $I_{mpp}$ 8.63A |
| | $V_{oc}$ 38.5V |
| | $I_{sc}$ 9.09A |
| PV array | 12 (Series)× 2 (Parallel) |

The PV array faults mainly studied in this paper include line-line fault (LLF), open circuit fault (OCF), and partial shading (PS). LLF refers to accidental connection of two points with a potential difference or low impedance. This fault may occur in the same PV string or between different PV strings. When a fault occurs, the fault string generates a reverse current, thereby reducing the output power, which is a very common fault type in PV arrays. In this paper, a failure experiment was carried out on one and two modules in the same string, which were denoted as LL-1 and LL-2. OCF refers to the failure of output power drop caused by the breakpoint of one or more strings in the array. This fault is usually caused by blown fuses or failure of solder joints. This article simulates PV string disconnection by flipping the combiner box air switch. PS means that one or several modules in the PV array are blocked, causing the bypass diode of the PV module to turn off, thereby causing power loss or hot spot effects. This paper uses cardboard to quickly shield the operation of PV modules to simulate the sudden drop in irradiance.

Among them, each group of experiments was collected at a temperature of 15~30°C and an irradiance of 200~800W/m². The data collection equipment was a Yokogawa-WT500 power analyzer. The set sampling frequency was 1Hz, and the sampling time of each sample was 60s. In the end, a total of 500 samples were collected. Among them, 100 are NORMAL samples, 100 are LLF-1 samples, 100 are LLF-2 samples, 100 are OCF samples and 100 are PS samples.

2.2. Fault diagnosis of PV array based on proposed model

Due to the complex diagnosis of PV arrays and the few features, the calculation efficiency is low, and even over-fitting problems occur. Using the ability of deep learning to automatically extract features, an SE-ResNet[6] model that can effectively diagnose PV faults is established. Fig. 1 shows the proposed PV array fault diagnosis model, and the fault diagnosis process is as follows:

On the established PV experimental platform, the WT500 power analyzer was used to collect 500 groups of fault samples with a sampling frequency of 1 Hz and a sampling time of 60s.

After data collection, the sample data is normalized according to Eq.1-3 and transformed into a 60x3 two-dimensional feature matrix.

$$I_{norm} = \frac{I_{pp}}{P_{mpp}} \frac{P_{mpp}}{V_{mpp}}$$

(1)
\[
V_{\text{norm}} = \frac{V_{\text{pv}}}{S V_{\text{oc}}} \quad (2)
\]
\[
P_{\text{norm}} = \frac{i_{\text{pv}}}{P t_{\text{sc}}} \cdot \frac{V_{\text{pv}}}{S V_{\text{oc}}} \quad (3)
\]

Where: \(i_{\text{pv}}\) and \(V_{\text{pv}}\) are the working current and working voltage; \(I_{\text{sc}}\) and \(V_{\text{oc}}\) are the short-circuit current and open-circuit voltage of the PV array under Standard Test Conditions (STC); \(P\) is the number of parallel PV strings in the array, and \(S\) is the number of PV modules connected in series in the PV string; \(I_{\text{norm}}, V_{\text{norm}}\) and \(P_{\text{norm}}\) are the three new normalized features. The values of these parameters are shown in Table 1.

Divide 64% of the total sample into the training set, 16% into the validation set, and 20% into the test set. The proposed model is pre-trained on the training set and the validation set to complete the weight update of the model.

Use Bayesian optimization algorithm (BO) to search for the minimum value of the objective function in the parameter space of preset values. After 300 iterations of calculation, the corresponding optimal parameters of the model are obtained.

The test set is input to the optimal model after training and optimization, one-dimensional feature vector is obtained through the fully connected layer, and the fault classification is performed through the Softmax function to test the performance of the model.

### 3. Experiment and Methods

#### 3.1. Parameter optimization

In order to make the model show its superior performance and obtain the best classification accuracy on the test set. The text uses Bayesian optimization (BO) algorithm to optimize the five parameters of the model. After pre-setting the type, range and quantity of the parameters, the objective function is calculated for 300 iterations to find the minimum value. Among them, the objective function is set as the difference between Loss and Accuracy. The parameter range and best value are shown in Table 2.

| Parameter          | Range           | Best value |
|--------------------|-----------------|------------|
| Batch Size         | \{8,16,32\}    | 16         |
| Num of dense       | \{32,64,128\}  | 64         |
| Activation Function| \{Sigmoid,Tanh,ReLu\} | ReLu |
| Learning Rate      | \{0.0001,0.2\} | 0.001      |
| R                  | \{2,4,8\}      | 4          |

In the table: Batch size is the number of samples selected for training; Num of dense is the output feature dimension of the fully connected layer; Activation function is the activation function of fully connected layer; Learning rate is a hyperparameter when updating the weights in the gradient descent process; \(R\) is the dimensionality reduction coefficient of the SE module.
3.2. Feature visualization

In order to prove the reliability of the model, use t-Distributed Stochastic Neighbor Embedding (t-SNE) to reduce the high-dimensional spatial data set to two dimensions, and visualize the original vector before feature extraction. The results are shown in Fig. 2. It can be seen that, except for OCF, such fault samples have their own regions in the feature space, which are easy to distinguish. However, the other three types of failure samples all overlap with each other. In particular, it should be pointed out that PS samples are the most difficult type of fault to identify, and their samples are mixed with NORMAL and LLF. The reason is that the waveform trends of LLF and PS at the moment of failure are the same. In the case of low irradiance, the amplitude of the two types of faults decreases so little that they are confused with NORMAL. The various feature vectors after feature extraction have their own regions in the feature space, with clear boundaries, and no other types of samples. It shows that the model is efficient and reliable in feature extraction.

![Feature visualization](image)

3.3. Classification result

In order to make the test more convincing, deep learning models such as 1D-CNN, AlexNet, VggNet, and ResNet were introduced as comparison algorithms, and 20 independent training and testing were performed. And the average classification accuracy is statistically analyzed, and the results are shown in Table 3. The method proposed in this paper has a classification accuracy of 99.56% on the training set, and a test accuracy of 99.67%, which is higher than its training accuracy. Its residual structure solves the problems of gradient disappearance and gradient explosion to a certain extent, which makes the model training fast and gradient transmission easy. The attention mechanism of the SE module improves the sensitivity of the model to the channel by changing the channel weight. It can be seen that compared with other deep learning models, this model has better generalization performance, and there is no over-fitting problem.

![Classification results](image)

| Model       | 1D-CNN | AlexNet | VggNet | ResNet | SE-ResNet |
|-------------|--------|---------|--------|--------|-----------|
| Training accuracy | 93.74% | 95.74% | 97.71% | 98.35% | 99.56%    |
| Test accuracy | 92.29% | 95.29% | 97.50% | 98.61% | 99.67%    |

4. Conclusion

Aiming at the shortcomings of traditional machine learning's weak ability to extract features from two-dimensional graphics data, this paper builds a fault diagnosis model for photovoltaic arrays based on SE-ResNet, and analyzes several common faults of PV arrays, including LLF, OCF, PS and
NORMAL. First, collect the current, voltage and power of the DC side at the time of failure on the already-built PV platform as sample data. After the samples are collected, they are normalized and converted into 60x3 two-dimensional image data. Then, 64% of the total samples are used as the training set, 16% as the validation set, and the remaining 20% as the test set. Secondly, use the training set to pre-train the model, and use BO to automatically search for the optimal parameters of the model. Finally, based on the verification set and test set, the model's fit and generalization accuracy were verified, and several other deep learning models were introduced for comparative verification.

Experimental results show that the model does not have over-fitting phenomenon, and the generalization accuracy is high, and its verification accuracy and test accuracy have reached 99.56% and 99.67%, respectively. It is proved that the model has practical engineering significance and can realize effective fault diagnosis of PV array.

Acknowledgments
This work is supported by Fuzhou Science and Technology Plan Projects(2019-G-44).

References
[1] Buerhop C, Schlegel D, Niess M, et al. Reliability of IR-imaging of PV-plants under operating conditions[J]. Solar Energy Materials & Solar Cells, 2012, 107:154-164.
[2] Rajasekar N, Kumar N K, Venugopalan R. Bacterial Foraging Algorithm based solar PV parameter estimation[J]. Solar Energy, 2013, 97:255-265.
[3] Yi Z, Etemadi A. Fault Detection for Photovoltaic Systems Based on Multi-resolution Signal Decomposition and Fuzzy Inference Systems[J]. IEEE Transactions on Smart Grid, 2017:1-1.
[4] Kaplani, E. Detection of Degradation Effects in Field-Aged c-Si Solar Cells through IR Thermography and Digital Image Processing[J]. International Journal of Photoenergy, 2012:1-11.
[5] Chen Z, Han F, Wu L, et al. Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents[J]. Energy Conversion and Management, 2018, 178:250-264.
[6] Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, PP-99.