Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 9, No. 3, 2024

Fault Diagnosis Method for Photovoltaic Arrays Based on Support Vector Machines

Li Wen[1], Gongquan Tan[1, 2, *]

[1] School of Automation and Information Engineering, Sichuan University of Science & Engineering, Zigong, China
[2] Artificial Intelligence Key Laboratory of Sichuan Province, Yibin, China
[*] Corresponding author: GongQuan Tan (Email: 43180229@qq.com)

Abstract: A fault diagnosis method for photovoltaic arrays based on Support Vector Machines (SVM) is proposed to address four typical faults of photovoltaic arrays (short-circuit, open-circuit, aging, shadowing). MATLAB is used to simulate the faults in the photovoltaic array, obtaining four characteristic parameters under different faults: short-circuit current (I_{sc}), open-circuit voltage (U_{oc}), current at the maximum power point (I_m), and voltage at the maximum power point (U_m). These parameters are used as training samples to establish a classification model through the SVM algorithm for training and verification. Simulation results show that this method can accurately diagnose the typical faults of the photovoltaic array, and the fault diagnosis accuracy is high.

Keywords: Photovoltaic array; Support vector machine; Fault diagnosis.

1. Introduction

With the rapid development and widespread application of solar power generation technology, the monitoring and fault diagnosis of photovoltaic arrays, which are a crucial part of photovoltaic power generation systems, have become particularly important. Due to the characteristics of the photovoltaic modules themselves and their constant exposure to harsh environments, they are prone to faults such as short-circuits, open-circuits, and aging. These not only affect the lifespan of the photovoltaic modules but also severely damage the power generation efficiency of the photovoltaic power station. Therefore, quickly and accurately detecting and diagnosing faults in photovoltaic arrays is crucial for improving power generation efficiency and reducing unnecessary losses [1].

In response to the direct current side fault problem of photovoltaic arrays, scholars from home and abroad have proposed numerous solutions, which are mainly divided into traditional diagnostic methods and intelligent diagnostic methods. In terms of traditional diagnostic methods, Peizhen Wang and others proposed a method for diagnosing faults in photovoltaic arrays using infrared imaging [2]. However, the infrared imaging equipment required by this method is expensive and not suitable for use in large-scale photovoltaic power stations. In addition, some scholars have proposed multi-sensor detection methods and ground capacitance measurement methods, but both of these methods are offline detection, which severely reduces the efficiency of photovoltaic power generation. In terms of intelligent diagnostic methods, researchers such as Yuanzhang Wang [3] and Guangyu Liu [4] have attempted to diagnose faults in photovoltaic arrays using neural network models. However, neural network algorithms have disadvantages such as randomness, non-replicability, and slow convergence speed, so they are not suitable for quick and stable fault diagnosis of photovoltaic arrays.

Support Vector Machine (SVM) is based on the principle of structural risk minimization, aiming to improve the generalization ability of the classification model. This method has been widely used in various fields such as pattern recognition, real-time prediction, and fault diagnosis [5-6]. In this paper, MATLAB is used to simulate the faults of photovoltaic arrays to obtain the characteristic parameters of the photovoltaic power station during operation. These parameters are used for fault diagnosis of the uninterrupted operation of the power station after model training. The experimental results fully verify that the method proposed in this paper has excellent recognition capabilities under various fault conditions and is a practical fault diagnosis method.

2. Fault Detection Method Based on SVM

2.1. The principle of SVM

Support Vector Machine (SVM) can effectively solve practical problems such as small sample size, non-linear regression, and local minima. It uses kernel functions to map sample data to high-dimensional space for linear regression processing. The principle is shown in Figure 1.

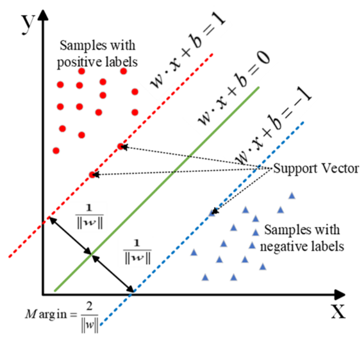

Figure 1. Support Vector Machine binary classification diagram

The circles and triangles represent two different types of samples. In the case of linear separability, the two types of samples can be easily divided into positive and negative categories. The solid line is the dividing line between the two types of samples, that is, the hyperplane; the two dashed lines respectively correspond to the two planes that pass through the two types of samples and are closest to the hyperplane, and the interval between them is the classification interval; the sample data on the dashed line is the support vector of the problem to be classified. The core principle of SVM is to find an optimal classification hyperplane to minimize the classification error rate and maximize the classification interval.

Assume a linearly separable training sample dataset x_i and fault class y_i. The classification hyperplane is $w \cdot x+b=0$, and finding the optimal hyperplane can be achieved by converting it to the constraint problem of finding (1) :

$$\begin{cases} F=\frac{1}{2}\|w\|^2+C\sum_{i}^{n}\zeta_i \\ s.t.y_i(w^T\cdot x_i+b)+\zeta\geq 1 \end{cases} \quad i=1,\dots,n \tag{1}$$

Where: w is the weight vector of the hyperplane; b is offset. In practical problems, linear inseparability will be encountered, which may cause hard interval classification failure, so relaxation variable ζ is introduced to relax the constraints and allow some data sample points to exceed the classification plane. At the same time, a non-negative penalty factor C is introduced to constrain ζ to ensure the accuracy of classification.

By introducing Lagrange function, the quadratic programming problem is transformed into the dual problem of formula (2) :

$$\begin{cases} \max L(\alpha)=-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\left(x_i,x_j\right)+\sum_{i=1}^{n}\alpha_i \\ s.t.\sum_{i=1}^{n}\alpha_i y_i=0 \quad 0\leq\alpha_i\leq C \end{cases} \tag{2}$$

Where: α is a Lagrange multiplier. If the training sample data set is nonlinear and indivisible, in order to obtain accurate classification results, the training sample in the original space should be mapped to a high-dimensional space. The final optimal classification surface function is expressed as formula (3) :

$$f(x)=\operatorname{sgn}\left[\sum_{i=1}^{n}\alpha_i y_i k\left(x_i,x_j\right)+\hat{b}\right] \tag{3}$$

Where: $k\left(x_i,x_j\right)$ is the kernel function. This will directly affect the performance of SVM and determine the accuracy of classification. Commonly used kernel functions include sigmod kernel function and Gaussian kernel function radial basis kernel function (RBF). Because RBF kernel function has good performance [7], this paper chooses RBF kernel function as the kernel function of support vector machine.

2.2. Fault monitoring methods are introduced

The schematic diagram of fault detection method is shown in Figure 2, which is divided into two parts: model construction and fault diagnosis.

Stage 1: Model construction. Firstly, the training set data is collected, which includes the characteristic parameters operating under different fault states, and then the data is normalized for the training of the diagnostic model.

Stage 2: fault diagnosis: collect the characteristic parameters of MATLAB simulation, input the data into the model constructed in stage 1 after normalization processing for state discrimination, and output the discrimination results after processing by the state decision module for state output, the output result is the final result of the fault diagnosis method.

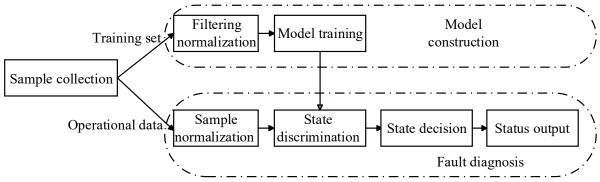

Figure 2. Fault diagnosis method diagram

3. Photovoltaic Array Fault Analysis

3.1. Fault sample extraction of photovoltaic array

In order to study the fault diagnosis of photovoltaic array, this paper uses MATLAB/Simulink software to build a simulation model [8]. Specifically, a simulation model of a photovoltaic array with 4 rows and 3 columns in series and parallel structure was established to simulate the operation of the actual photovoltaic power generation system under the 1000W/m2 light intensity and 25℃ temperature reference standard [9]. As is shown in Figure 3.

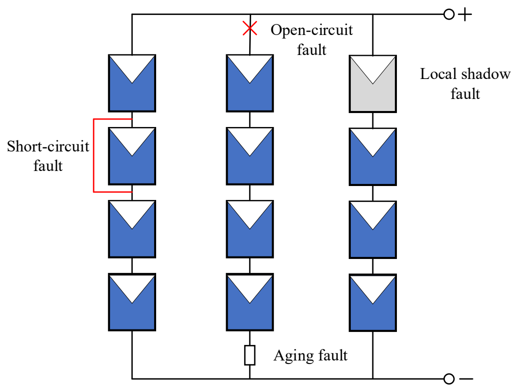

Figure 3. Pv array fault simulation

After the photovoltaic array is built, the series resistance is used to simulate the aging fault. Shading fault is simulated by reducing light intensity. The short-circuit fault is simulated by shorting the positive and negative electrodes of the photovoltaic cell. The open circuit fault is simulated by cutting the connection wire [10], and I-U and P-U curves under different operating states are obtained through simulation, as shown in Figure 4

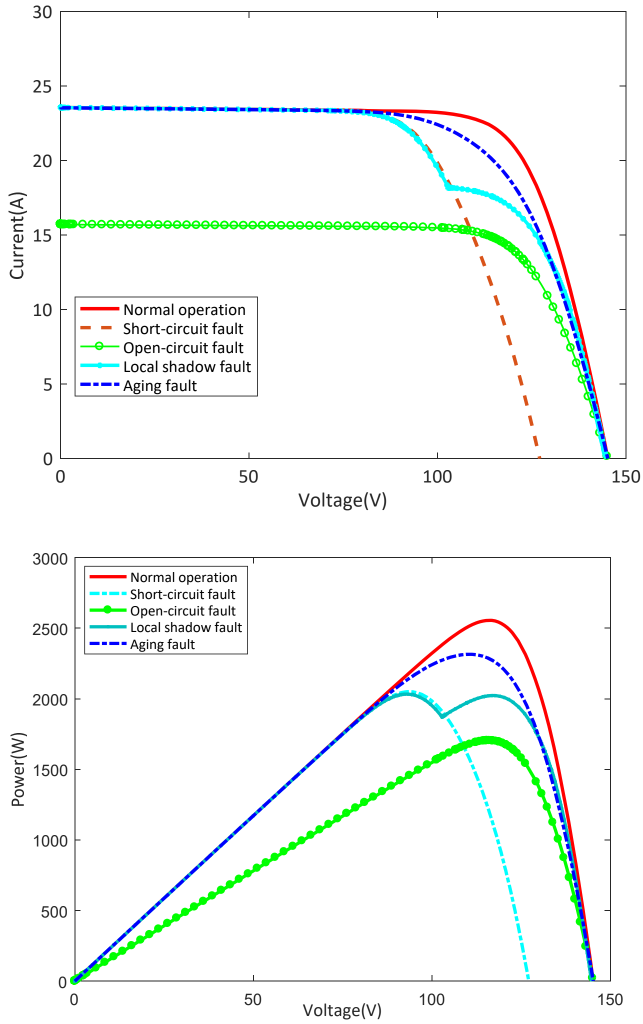

Figure 4. The output I-U and P-U characteristic curves of the photovoltaic array

As can be seen from Figure 4, compared with the normal operation of photovoltaic modules, the open circuit voltage U_{oc} will decrease significantly when a component is short-circuited, while the short-circuit current I_{sc} will remain unchanged. When an open circuit fault occurs, the short circuit current I_{sc} of the PV array decreases obviously. When the abnormal aging fault occurs, the output voltage of the array is between the output voltage of the short-circuit fault with the same number of fault components and the output voltage of the normal operating state. When the local shadow fault occurs, the I-U characteristic curve presents a stepped change, and the P-U characteristic curve presents a "multi-peak" phenomenon, and the voltage Um at the maximum power point of the photovoltaic array decreases obviously. Therefore, short-circuit current I_{sc}, open circuit voltage U_{oc}, maximum power point current I_m and maximum power point voltage Um are selected as input values to accurately distinguish these four fault types.

3.2. Classification of fault samples

Through simulation in MATLAB, four characteristic parameters, I_{sc}, U_{oc}, I_m, and U_m, were obtained and used as training samples for fault diagnosis of photovoltaic arrays. 950 sets of data were selected from these, and 750 sets were randomly chosen as training samples for the Support Vector Machine. The remaining 200 sets were used as test data, as shown in Table 1.

Table 1. Sample classification

Operating status	Training samples/groups	Testing samples/groups
Normal condition	150	40
short circuit fault	150	40
open circuit fault	150	40
aging fault	150	40
Partial shading fault	150	40

4. Simulation Experiment Analysis

4.1. Parameter setting

When training the diagnostic model with sample data using SVM, there are strict requirements for the selection of the kernel function and the setting of the penalty factor. After multiple experiments, this paper selects the optimal penalty factor (C) as 3.6, and the optimal kernel function (g) as 0.01. As can be seen from Figure 5, the fitness value is 8.902e-08 and tends to stabilize at the 300th iteration.

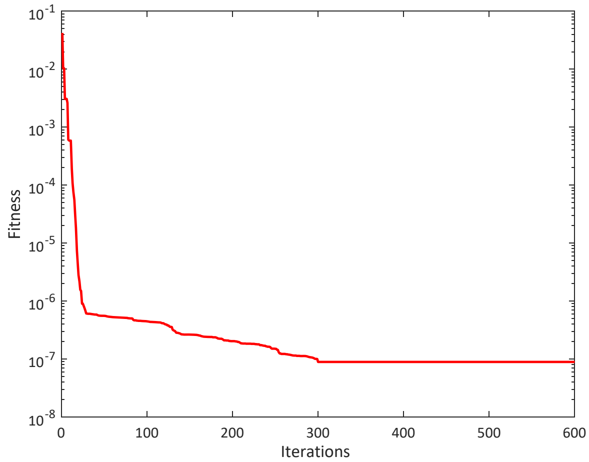

Figure 5. Fitness curve

4.2. SVM diagnostic results

The SVM photovoltaic array fault diagnosis model proposed in this paper, after being trained with 750 sets of fault training data, predicts the accuracy of 200 sets of test data. The fault values on the y-axis correspond to the photovoltaic array operating in normal state, short circuit fault, open circuit fault, aging fault, and local shadow fault. The red stars represent the true values of the actual fault types, and the blue circles represent the predicted values of the predicted fault types. When the two coincide, it indicates a correct diagnosis; otherwise, it indicates a wrong diagnosis. 186 sets were correctly classified, with an accuracy of 93%. The classification results are shown in Figure 6.

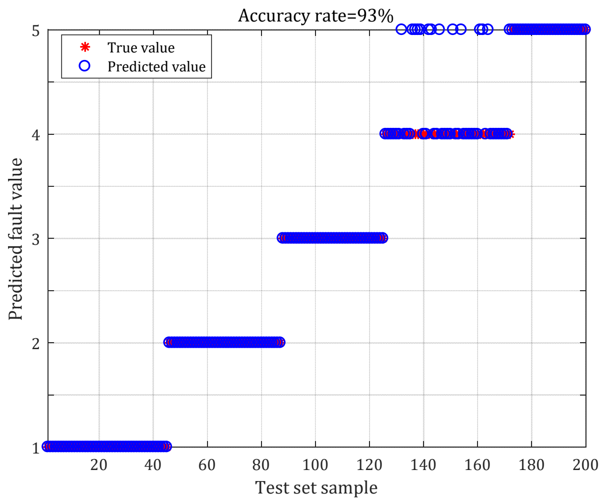

Figure 6. Classification results

5. Conclusion

The simulation results show that the photovoltaic array fault diagnosis method based on SVM proposed in this paper can not only accurately classify the four types of faults in the photovoltaic array, especially the short circuit fault, open circuit fault, and local shadow fault, which can be predicted with 100% accuracy, but also has a faster convergence speed. Therefore, this method has certain research value for improving the fault detection technology of photovoltaic arrays. It is of great significance for achieving reliable supply of sustainable energy.

References

[1] Peisheng Sun, Tangxian Chen, Chen Cheng, Zheng Li. Research on Photovoltaic Array Fault Diagnosis Based on SOA-SVM Model [J]. Journal of Power Supply: 1-13.

[2] Peizhen Wang, Shicheng Zheng. Fault Analysis of Solar Photovoltaic Arrays Based on Infrared Images [J]. Journal of Solar Energy, 2010, 31(02):197-202.

[3] Yuanzhang Wang, Chunhua Wu, Diqing Zhou, et al. Research on Photovoltaic Array Fault Diagnosis Based on BP Neural Network [J]. Power System Protection and Control, 2013, 41(16):108-114.

[4] Liu G , Yu W , Zhu L .Condition classification and performance of mismatched photovoltaic arrays via a pre-filtered Elman neural network decision making tool[J].Solar Energy, 2018, 173(OCT.):1011-1024.

[5] Shengming Zhou, Jianling Qu, Feng Gao. Application of Nonlinear Support Vector Regression in Predicting Performance Parameters of Aircraft Engines [J]. Measurement and Control Technology, 2012, 31(02):35-36+41.

[6] Changfeng Tang, Zhinong Wei, Zhijie Li, et al. Grid Fault Risk Assessment Based on Factor Analysis and Support Vector Machine [J]. Power System Technology, 2013, 37(04):1039-1044.

[7] Haoran Guo, Zetao Li. Research on Photovoltaic Array Fault Diagnosis Optimized by Genetic Algorithm and Support Vector Machine [J]. Intelligent Computing and Applications, 2019, 9(05):58-62.

[8] Ropp M E , Gonzalez S .Development of a MATLAB/Simulink Model of a Single-Phase Grid-Connected Photovoltaic System[J].IEEE Transactions on Energy Conversion, 2009, 24(1): 195-202.

[9] Ding K , Bian X , Liu H ,et al.A MATLAB-Simulink-Based PV Module Model and Its Application Under Conditions of Nonuniform Irradiance[J].IEEE Transactions on Energy Conversion, 2012, 27(4):864-872.

[10] Dongjun Li, Qingkun Li, Zhong Zheng. Fault Diagnosis and Simulation of Photovoltaic Power Generation Array Based on Simulink [J]. Electronic Production, 2016(23):75-76.