Weighted Islanding Detection for DC Microgrid Based on Random Forest Classification

Qingzhu WAN[1], Kaicong WU[1, 2 *]

[1]North China University of Technology, Shijingshan District, Beijing 100043, China
[2]DONGFANG Electric wind power corporation, Jingyang District, Deyang 618000, China

Abstract. At present, the main form of microgrid is AC grid. DC microgrids have received extensive attention and research with the rapid development of various DC power. The operation mode of the DC microgrid is divided into grid-connected operation and islanding operation. Islanding is formed after the circuit breaker tripped, which connects microgrid to large grid. Islanding operation can be divided into planned islanding and unplanned islanding. Unplanned islanding will cause certain harm to users and systems, so it is necessary to detect islanding accurately in the DC microgrid. This paper proposes an islanding detection method for DC microgrid based on random forest classification. Firstly, raw data is cleaned, extracted features and generated feature vector set. The extracted features include six islanding characteristic indexes, which consist of voltage, current, output active power and their first order backward difference on the DC bus side. Then, based on random forest classification, building the islanding detection model. Islanding detection model for DC microgrid can distinguish islanding event successfully and accurately. Based on weighted random forest classification, it can detect islanding event more accurately compared with decision tree classification when processing large amounts of data.

1 Introduction

Currently, DC microgrids are developing rapidly as distributed power supplies and a variety of DC power-using devices are being used in a large number of microgrids[1].

DC microgrid operation is divided into two modes of operation: grid-connected and islanded[2-3], and when the circuit breaker of the connection between the microgrid and the larger grid is tripped, islands are formed. Island operation can be divided into planned and unplanned islands[4]. Unplanned islands can cause some harm to the users or the system[5], so accurate island detection is essential for the safe and stable operation of DC microgrids.

Traditional islanding detection methods mainly include local active, local passive, and remote methods[6]. These methods have large detection blind zones, affecting the power quality of inverter output, high cost and complex design problems. Data mining technology has two main functions: one is to query the historical operating information of the power system, and the other is to establish potential links between query data to address siloed predictions, decision making, etc. Therefore, scholars at home and abroad have conducted different levels of research on data mining techniques. The reference[7] proposes an islanding detection data mining system, which consists of three steps, including key feature identification, base learner and meta-learner, to improve the accuracy and generalization of AC microgrid islanding detection. The reference [8] employs a data mining C4.5 decision tree to solve the problem of islanding detection in AC microgrids. Reference [9] proposed a random forest classifier-based anomaly data detection method to achieve anomaly detection and anomaly data repair function. The reference [10] trains five different classifiers for the problem of islanding detection in AC microgrids, with a random forest classifier capable of Detecting silos with high accuracy and reasonable time. Random forest is an integrated learning method for multiple decision trees, which can overcome some of the shortcomings of a single decision tree, with good scalability and parallelism[11], which is more widely used in many fields, such as remote sensing data analysis[12], biochips[13], speech identification[14], cheating webpage detection and complete network detection[15]. It has a good application in islanding detection in distributed power generation systerms[16].

In this paper, an islanding detection method for DC microgrid based on random forest classification is proposed. First, a DC microgrid simulation model was built to obtain state information such as voltage, current, and output active power on the DC bus side of the system in grid-connected and islanded operation mode, and data cleaning was performed. Then, key features reflecting the islanded operation of the DC microgrid are extracted to generate a set of feature vectors. Finally, a random forest classification-based method for DC microgrid islanding detection is proposed, which is shown to improve the accuracy of DC microgrid

* Corresponding author: jianranc@163.com

islanding detection. Compared with decision tree classification under different scales of data and weighted random forest, the random forest classification model considering the weight can detect islands more accurately, which has certain scalability and practical significance.

2 DC Microgrid Islanding Operation Feature Extraction

DC microgrid islanding detection mainly includes data acquisition, data cleaning, islanding feature extraction and random forest classification. Island features are intrinsic to island operation and are directly related to the accuracy of island detection, so extracting valid island features is key to island detection.

At the moment the islanding occurs, the AC grid changes active power P, reactive power Q, frequency f, voltage U, and power factor $cos\ \varphi$. Therefore, the key features of islanding detection are: P and its rate of change $\frac{dP}{dt}$, f and its rate of change $\frac{df}{dt}$, U and its rate of change $\frac{dU}{dt}$, $cos\ \varphi$ and its rate of change $\frac{d\ cos\ \varphi}{dt}$ [17-18]. To summarize the existing studies, the common features of islanding detection in AC power grids are shown in Table 1.

Table 1. Common features of islanding detection for AC power grid.

Features	Symbol	Features	Symbol
active power	P	active power's rate of change	$\frac{dP}{dt}$
frequency	f	frequency's rate of change	$\frac{df}{dt}$
voltage	U	voltage's rate of change	$\frac{dU}{dt}$
power factor	$cos\ \varphi$	power factor's rate of change power factor	$\frac{d\ cos\ \varphi}{dt}$

The DC microgrid only needs to consider active power balance and stable DC bus voltage, but not phase, frequency, and reactive power [19-20], so the paper selects six islanding characteristics indicators such as DC bus side voltage, current, output active power, and their respective first-order backward differential as detection features, as shown in Table 2.

Table 2. Six key features of islanding detection for DC microgrid.

Features	Symbol	Features	Symbol
voltage	$U(N)$	voltage's respective first-order backward differential	$U'(N)$
current	$I(N)$	current's respective first-order backward differential	$I'(N)$
output active power	$P(N)$	output active power's respective first-order backward differential	$P'(N)$

The extraction of key features of DC microgrid islanding operation includes the following steps, as shown in Figure 1.

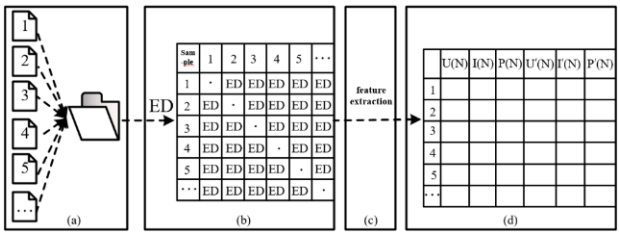

Fig. 1. Islanding features extraction process for DC microgrid.

2.1. Captions/numbering

Build a DC micro-grid simulation model, and the DC bus side of the continuous voltage is $U(t)$, the sampling voltage is $U(NT_d)$, for the sampling period, defined as $U(N) = U(NT_d)$. The voltage data $U(N)$ for grid-connected and islanded operation of the DC microgrid can be obtained. Similarly, you can get state information such as the DC microgrid bus-side current $I(N)$, output active power $P(N)$, AC-side voltage, current, output active power and reactive power and so on. And these data is stored.

2.2. Data cleansing

Estimating the similarity of each pair of samples using (Euclidean distance) [21-22], cleaning and merging the data, and handling duplicates values and outliers to improve the quality of the dataset. The Euclidean distance is calculated as follows

2.3. Captions/numbering

Captions should be typed in 9-point Times. They should be centred above the tables and flush left beneath the figures.

$$ED(A,B) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \quad (1)$$

where, $ED(A,B)$ -the Euclidean distance between the sample A and the sample B; n -the number of sample dimensions; a_i, b_i -the dimensional information of the sample A, B.

2.4. Feature extraction

Six state quantities of voltage, current, output active power and their respective first-order backward differential are extracted from the cleaned data as DC microgrid islanding detection characteristics. The first-order backward differential of voltage, current, and output active power are $U'(N)$, $I'(N)$, $P'(N)$ respectively, and are calculated as follows

$$\begin{cases} U'(N) \approx \frac{\nabla U(N)}{T_d} = \frac{U(N)-U(N-1)}{T_d} \\ I'(N) \approx \frac{\nabla I(N)}{T_d} = \frac{I(N)-I(N-1)}{T_d} \\ P'(N) \approx \frac{\nabla P(N)}{T_d} = \frac{P(N)-P(N-1)}{T_d} \end{cases} \quad (2)$$

where, $U(N)$, $U(N-1)$-the voltage on the DC bus at the moment and moment $(N-1)T_d$, $I(N)$, $I(N-1)$-the current on the DC bus at the moment NT_d and moment $(N-1)T_d$, $P(N)$, $P(N-1)$-the output active power on the DC bus at the moment NT_d and moment $(N-1)T_d$.

2.5. Generation of vector sets

Determine the set of feature vectors, each of which contains $U(N)$, $I(N)$, $P(N)$, $U'(N)$, $I'(N)$, $P'(N)$ etc. six feature indicators.

3 Random forest-based classification DC microgrid islanding detection

The random forest consists of multiple mutually independent decision trees, which is an integrated learning method based on decision trees, and the random forest classification results are determined by the voting of all decision trees, which has a high accuracy and is able to process big data effectively, and therefore has good results in orphan detection.

The paper develops the modeling using random forest algorithm on the basis of obtaining the indexes of DC microgrid islanding characteristics, and the modeling process is shown in Figure 2. First, the sample set containing six indicators $U(N)$, $I(N)$, $P(N)$, $U'(N)$, $I'(N)$, $P'(N)$ of islanding characteristics was split into a training set and a test set. Then use the method randomly selects multiple subsets of training samples $S_k(k = 1, 2, \ldots, n)$ from the training sample set and models a decision tree for each subset separately. Combining the decision results from multiple trees, the final orphan detection model is derived by voting[23].

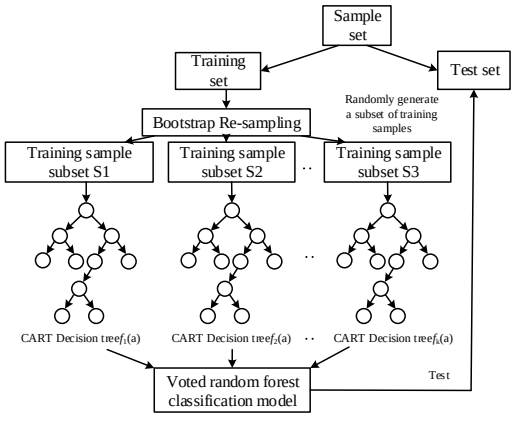

Fig. 2. Modeling process of islanding detection for DC microgrid based on random forest

3.1. Random selection of a subset of training samples

The raw sample set consists of two types of data, DC microgrid grid-connected operation and islanded operation. The training set containing six indicators of islanding characteristics is divided into a training set and a test set according to 7:3[24-25]. Firstly, a n sub-sample set was randomly selected from the training set by sampling method, and then a n classification regression tree was constructed (classification and regression tree，CART)[26].

The construction of each decision tree requires the random selection of F of the M island detection traits as a random feature variable to participating in the decision tree nodes splitting process. This method solves the overfitting problem arising from the construction of a decision tree and ensures the randomness of the construction of the decision tree.Where

$$0 \le F < \log_2(M+1) \quad (3)$$

M is the total number of island detection features. When M is 6, randomly choose F as 1 or 2. When decision tree randomly chooses 1 characteristic, F=1; and when decision tree randomly chooses 1 characteristics, F=2.

3.2. Building the CART decision tree

Based on the principle of Gini coefficient minimization, n decision trees is constructed for each subset of random training samples using a CART algorithm, creating a "forest"[27]. According to the reference[28] the number of decision trees is about 100 when the random forest classification performance is close to optimal, so in this paper 99 trees are selected decision tree to form a random forest.

3.3. Voting on the results of silo testing

The random forest model consists of a set of n CART decision trees that are used to validate the accuracy of the model using test set data. The test samples a are used as inputs to the random forest and the output of the k tree is the

$$f_k(a) = i, \ i = \begin{cases} 0, & \text{non-islanded} \\ 1, & \text{islanded} \end{cases} \quad (4)$$

then the output of the stochastic silvicultural classification model is

$$f_{RF}(a) = \arg\max_{i=0,1} \left\{ \sum_{k=1}^{n} \left[f_k(a) = i \right] \right\} \quad (5)$$

where, f_k-single decision tree classifier model; f_k-single decision tree classification result, i=0 indicating that the prediction result is non-islanded, i=1 indicating that the prediction result is islanded; f_{RF}-random forest classification model output; n-total number of decision trees included in the random forest model.

4. Experimental results and analysis

Application of MATLAB/Simulink establishes a simulation model for DC microgrid. The model consists of five parts: AC main grid, inverter controller, line impedance model, DC circuit breaker, distributed generation, and a DC microgrid system and load model, is shown in Figure 3.

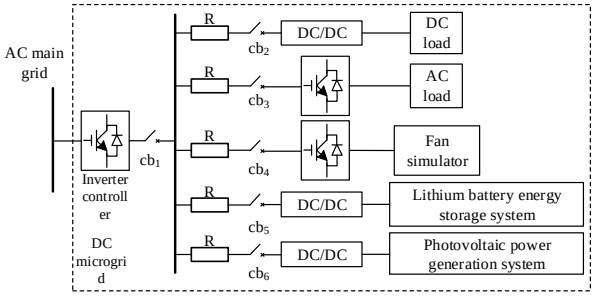

Fig. 3. Topological of the typical DC microgrid

The wind turbine simulator is used as the object of study. On the basis of guaranteeing the maximum power output of wind power generation and system power balance, the DC microgrid grid-connected operation, large island event, and small island event are simulated by switching action, respectively. As shown in Table 3.

Table 3. Simulation events

Type	Switching Action
Grid-Connected	—
Large Island	disconnect cb1：the DC microgrid is disconnected from the AC main power grid
Small Island	disconnect cb1、cb2： the DC microgrid is disconnected from the AC main power grid and DC load disconnect cb1、cb3：the DC microgrid is disconnected from the AC main power grid and AC load disconnect cb1、cb5：the DC microgrid is disconnected from the AC main power grid and lithium battery energy storage system disconnect cb1、cb6：the DC microgrid is disconnected from the AC main power grid and photovoltaic power generation system

For grid-connected operation, large islanding events, and small islanding events, three operating states are simultaneously collected, including DC bus side voltage, current, output active power, AC side voltage, current, output active power, and output reactive power of the wind turbine simulator.

Take the grid-connected DC microgrid operation and large island operation as an example, the blue part of the system simulation diagram shows the grid-connected operation state and the red part shows the large island operation state. When the simulation runs to 0.7s, the cb1 switch is disconnected and the DC microgrid enters large island operation. After 0.2s or 0.9s moment, the large island operation state enters steady state operation and collects the DC bus side voltage of the wind turbine simulator. Current, output active power changes, as shown in Figure 4.

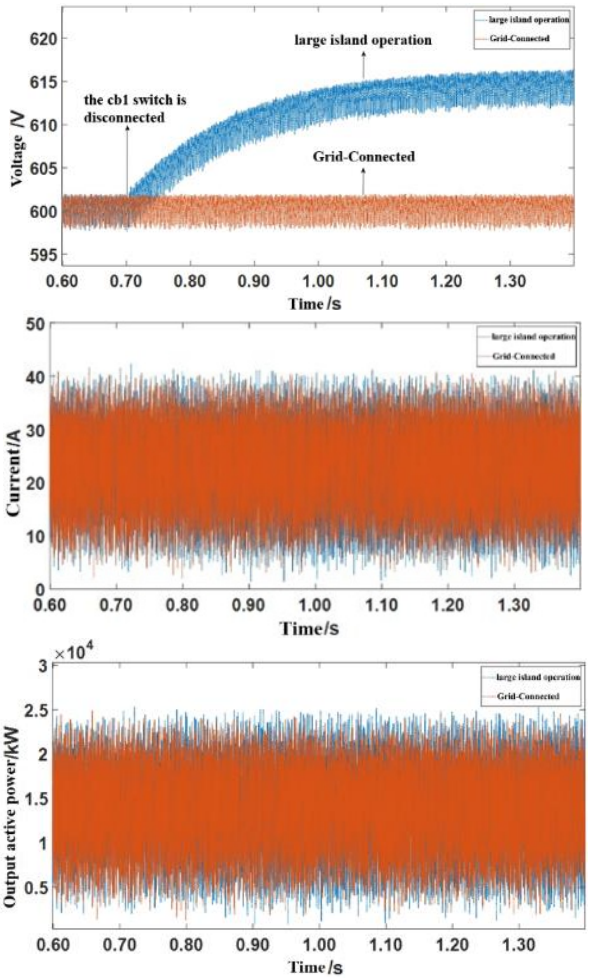

Fig. 4. Scatter diagram of voltage, current and output active power

Compare the changes of each parameter in grid-connected operation and large island operation, and select the state of the DC bus side in the interval T_d=5μs from 0.7s to 0.9s. By simulating Table 4 simulation events, 240,000 sets of simulation data can be obtained. Each set of data contains status information such as fan simulator DC bus side voltage, current, output active power, AC side voltage, current, output active power, output reactive power and so on.

Select the DC bus-side voltage, current, and output active power as raw data, and clean these data in MATLAB. KNN classification results of the classifier learner as an evaluation of the quality of the dataset indicators. The confusion matrix has 0 for non-islanded and 1 for islanded, and the 240,000 sets of data are cleaned to get the 223,780 groups data. Misclassified data were significantly reduced and the quality of the dataset improved, as shown in Figure 5.

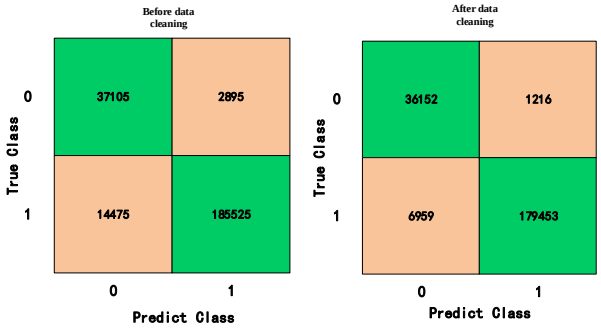

Fig. 5. Confusion matrix before and after data clean

The AUC values before and after data cleaning are 0.93 and 0.97 respectively, so the cleaned data can be classified better, i.e. the data quality is higher. . The ROC curves before and after data cleaning are shown in Figure 6.

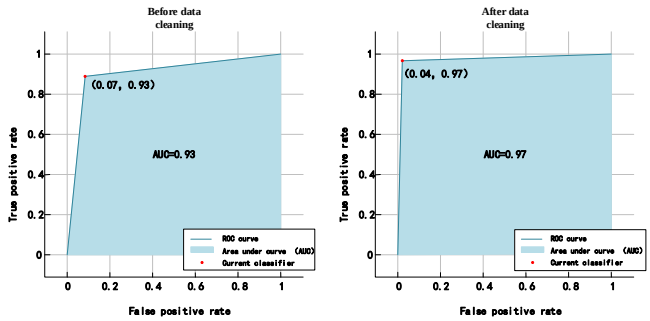

Fig.6. ROC curves before and after data clean

Select $U(N)$, $I(N)$, $P(N)$, $U'(N)$, $I'(N)$, $P'(N)$, $0 < N \le 240000$ from the cleaned data. And these six sets of features are selected and combined to generate feature vector sets. The sample set contains 223,780 feature vectors, of which 70% are randomly selected for the training set and the remaining 30% for the test set. The training set is used to train the random forest classifier and the test set is used to test the recognition rate of the completed trained model. The number of samples is shown in Table 4 and the test results are shown in Table 5.

Table 4. Partitioning the sample set

Set	Number Of Samples
Training Set	156646
Test Set	67134
Total	223780

Table 5. Classification results

Type	Total Number	Correct	Wrong	Accuracy
Islanded	55875	54880	995	98.22%
Grid-Connected	11259	10562	697	93.81%

From Table 5, it can be seen that the detection accuracy for islanding and parallel operation is 98.22% and 93.81%, the total detection accuracy of random forest classification is 97.48%. respectively, methods based on random forest classification can be effective in detecting orphan islands.

5. Experimental Comparison

5.1 Comparison of random forest classification method and decision tree method

Compare and analyze the random forest-based classification method with the decision tree method. Select sample datasets containing 180,000, 120,000, and 60,000 groups for the test analysis and repeat the above steps, and the results of the analysis are shown in Table 6.

Table 6. Performance comparison between the proposed algorithm and decision tree

The Initial Data Set Contains The Amount Of Sample Data	Accuracy %	
	Random Forest Method	Decision Tree Method
240000	97.48	95.68
180000	96.88	96.22
120000	96.69	96.74
60000	94.85	97.45

As shown in Table 6, the larger the sample size of the dataset, the higher the predictive accuracy of the random forest classification detection model compared to the decision tree approach.

5.2 Weighted random forest classification method

In the random forest classification method, calculate the decrease value of the Gini coefficient of a feature index at the corresponding node, which is called D_{Gini}. The weight of the characteristic index can be obtained through solving and calculating the reduced value of all Gini coefficients in the forest, and taking the average. Calculated as follows

$$Q_k = \frac{\sum_{i}^{n}\sum_{j=1}^{i} D_{Gkij}}{\sum_{k}^{m}\sum_{i}^{n}\sum_{j=1}^{i} D_{Gkij}} \tag{6}$$

Where, Q_k is the weight of the k indicator in the m indicators, which is the percentage of importance; m is the total number of characteristic indicators, m=6 in this article; n is the number of overall trees in the construction of the random forest, n=99 in this paper; t is the number of nodes in a single classification tree; D_{Gkij} is the reduction value of the Gini coefficient of the k-th indicator at j nodes on the i-th tree.

The above method of calculating the weights of feature indicators can effectively improve the generalization ability of the data set, and the data set can handle multiple types of data without standardization.

Bring the complete training set sample data into the random forest model for training, use the self-detection function of the random forest model, and obtain the

results of the importance of each indicator according to the formula (8) at the same time.

Table 7. The average Gini coefficient reduction value and importance degree

Features	Average Gini coefficient reduction	Importance percentage/%
voltage	90.15	32.6
current	15.76	5.7
output active power	46.39	15.8
voltage's respective first-order backward differential	77.71	28.1
current's respective first-order backward differential	15.21	5.5
output active power's respective first-order backward differential	34.01	12.3

According to Tabel 7, voltage is the most important indicator, accounting for 32.6% of the importance of all indicators, which is shown in the status information comparison chart of grid-connected operation and island operation. The first order of voltage, the first-order backward difference of output active power, and output active power account for 28.1%, 15.8%, and 12.3%, indicating that these three indicators have a greater effect on islanding detection and can affect the detection result to a certain extent. It can be found that the influence of current on the detection result is much lower than the voltage characteristic index.

For the generated feature vector group, each electrical feature index is given a weight value, and the weight distribution is

$$W = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6) \quad (7)$$

Where, $\omega_1 = 0.326$, $\omega_2 = 0.057$, $\omega_3 = 0.158$, $\omega_4 = 0.281$, $\omega_5 = 0.055$, $\omega_6 = 0.123$.

After assigning weights to the feature vector set obtained before, the feature vector set is obtained.

Similarly, the sample set contains 223780 feature vectors, of which 70% are randomly selected as the training set, and the remaining 30% are used as the test set. The training set is used to train the weighted random forest classifier, and the test set is used to test the recognition rate of the trained model. The test results are shown in Table 8.

Table 8. Classification results

Type	Total Number	Correct	Wrong	Accuracy
Islanded	55875	54903	972	98.26%
Grid-Connected	11259	10553	706	93.73%

As shown in Figure 7, it can be seen that in the weighted random forest model, the detection accuracy of isolated islands and grid-connected operation are 98.26% and 93.73%, respectively. Compared with the unweighted random forest, the islanding situation's detection accuracy of the weighted random forest model has increased by 0.04%.

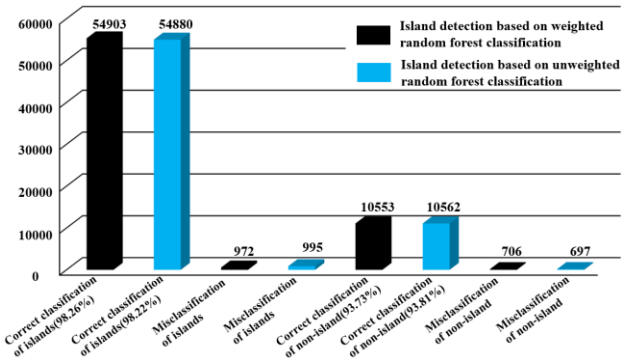

Fig.7. Comparison between weighted random forest and unweighted random forest

6 Conclusion

This paper proposes an island detection method for DC microgrid based on random forest classification and verifies the effect through simulation, resulting in the following main conclusions.

1) The method proposed in the paper is capable of accurately detecting orphaned islands with an accuracy of 98.26%. This ensures the safe operation of the distribution network containing a large number of distributed power supplies.

2) Compared to decision trees and random forest, the weighted random forest classification model has better generalization ability and can better detect siloed situations when dealing with large amounts of data.

The follow-up research focus of the paper will further consider the influence of more factors including the number of random forest containing trees, the proportion of transform training set and test set on the islanding detection accuracy, and continuously improve the proposed islanding detection method for DC microgrids based on random forest classification.

References

1. M Meng, Y N Yuan, M W Guo. (2014) "Research on grid-connected coordinated control strategy for DC micro-network." Heilongjiang Electric Power 36.2 (2014): 291-295.
2. S P Gao, K C Wang, W H Li. (2018) "A new islanding detection method for grid-connected photovoltaic system based on harmonic impedance component." Acta energiae solaris sinica 39.5 (2018): 1335-1341.
3. L J Yan, H Q Li, X Y Yu, et al. (2018) "Multi-objective day-ahead optimal scheduling of isolated microgrid considering flexibility." Power systems technology 42.5 (2018): 1432-1440.
4. K Tao, B B Qi, L HUANG. (2016) "Research on the protection scheme of the intentional island with distributed power supply." Electronic design engineering 24.6 (2016): 53-63.
5. W P Zhu, L Chen, C S Dai, et al.(2016) "Study and application of an intelligent anti-islanding device."

Distributed power and grid connection technology 5(2016): 47-52.

6. C He, M Wang, G Z Chen. (2015) "Optimal selection of UHVDC connection mode to multi-infeed HVDC system." Electric power automation equipment 35.6(2015): 87-92.
7. P C Zhang, X F Tan, P X Yang. (2014) "Critical feature identification and meta-learning method for islanding detection." Automation of electric power systems 38.18(2014): 72-77.
8. X Lin, X F Dong, Y P Lu, et al. (2011) "Application of data mining in island detection of distributed generation." Proceedings of the CSU-EPSA 23.2(2011): 38-44.
9. X L Yang, X F Tao, F X Huang, et al.(2018) "The method for detecting and repairing abnormal electric energy data based on random forest and time series model." Proceedings of the 3rd Smart Grid Conference Computer Engineering and Application, Nan Jing, (2018): 224-225.
10. Bataineh H-A, Kavasseri R G. (2005) "Islanding Detection with data mining methods-a comparative study." Ninth Annual IEEE Green Technologies Conference. Denver, Colorado, USA, IEEE, (2005): 104-109.
11. Y S Wang, S T Xia. (2018) "A survey of random forests algorithms." Research & development (2018): 49-55.
12. ZHANG C X, GUO G. (2011) "Research of the applications of Out-of-bag sample."Software 3(2011): 1-4.
13. X Li, T W Zhang, S Q Rao, et al. (2004) "Decision forest for extracting feature genes." Journal of harbin institute of technology 36.4(2004): 480-483.
14. G L Li, Y Tie, L Qi. (2019) "Multi-feature speech emotion recognition based on random forest classifiction optimization." Microelectronics&computer 36.1(2019): 70-73.
15. X N Fang, H X Zhang, S Gao. (2013) "Web spam detection based on SMOTE and random forests." Journal of Shandong University (Engineering Science) 1(2013): 22-27.
16. D W Wang, Z W Sun. (2015) "Big Data Analysis and parallel load forecasting of electric power user side." Proceedings of the CSEE 35.3(2015): 527-537.
17. Redfern M A, Usta O. (1995) "A new microprocessor based islanding protection algorithm for dispersed storage and generation units." IEEE trans on power delivery 10.3(1995): 1249-1254.
18. Salman S K, King D J, Weller G. (2002) "New loss of mains detection algorithm for embedded generation using rate of change of voltage and changes in power factors." Seventh International Conference on Developments in Power System Protection Amsterdam, the Netherlands (2002): 82-85.
19. Ito Y, Z Q Yang, Akagi H. (2004) "DC microgrid based distribution power generation system." IEEE International Conference on Power Electronics and Motion Control 3(2004): 1740-1745.
20. Schonberger J, Duke R, Round S D. (2006) "DC-bus signaling: A distributed control strategy for a hybrid renewable nano grid." IEEE transactions on industrial electronics 53.5(2006): 1453-1460.
21. Y P Deng, X Yang, M X Zhang, et al. (2015) "Discrimination of magnetizing inrush and short circuit current of converter transformer based on euclidean distance." Power system technology 39.9(2015): 2671-2676.
22. J Y Zhao, X Q Han, C Liang, et al. (2017) "Optimal distribution network reconfiguration combining fuzzy membership with euclidean distance." Power system technology 41.11(2017): 3624-3631.
23. T Zhao, L T Wang, Y Zhang, et al. (2016) "Relation factor identification of electricity consumption behavior of users and electricity demand forecasting based on mutual information and random forests." Proceedings of the CSEE 36.3(2016): 604-614.
24. H F Wu, Y N Liu, T Sun, et al. (2011) "Classification of malignant and benign pulmonary nodules in CT image based on curvelet transformation." Beijing biomedical engineering 30.5(2011): 471-473.
25. X L Chen, L M Yang. (2009) "Partitioning machine learning sample set using similarity to mean vector." Journal of Central South University 40.6(2009): 1636-1641.
26. J Liu, Y Wu, L Tan. (2006) "An improvement to the resampling procedure of Bootstrap method." Mathematical theory and plications 1.26(2006): 69-72.
27. X L Yuan, X Mei, J S Huang. "Feature selection based on random forest algorithm and the applications in fMRI data." Microelectronics&computer 31.8(2014): 132-135.
28. M Liu, R L Lang, Y B Cao. "Number of trees in random forest." Computer engineering and applications 51.5(2015): 126-131.