The Journal of Engineering

The 14th IET International Conference on AC and DC Power Transmission (ACDC 2018)

Islanding detection method of distribution generation system based on logistic regression

eISSN 2051-3305
Received on 3rd September 2018
Revised 15th October 2018
Accepted on 17th October 2018
E-First on 12th December 2018
doi: 10.1049/joe.2018.8794
www.ietdl.org

Chen Haoran[1] ✉, Xu Jianyuan[1], Li Hui[2], Chen Guoqiang[3], Jiang Yuanyu[1], Cai Shengwei[2], Chen Jiangbo[2]
[1]*School of Electrical Engineering, Shenyang University of Technology, No. 111 Shenliaoxi Road, Shenyang, People's Republic of China*
[2]*State Key Laboratory of Power Grid Environmental Protection, China Electric Power Research Institute, No. 143 Luoyu Road, Wuhan, People's Republic of China*
[3]*Economic Research Institute of State Grid Hubei Electric Power Supply Co. Ltd, No. 225 Xudong Street, Wuhan, People's Republic of China*
✉ *E-mail: smileland@126.com*

Abstract: Islanding operation will harm the safety and stability of the power system, in the occurrence of island operation which must be in the specified time to monitor it. Here, the logical regression algorithm is introduced into the islanding operation monitoring. First, an accurate distributed generation system model is established in Matlab environment. The model considers various possible island operating states, minimum monitorable area, multiple distributed power supplies, various operating conditions, and different network topologies, and then uses. The data feature quantity under various island operating conditions is trained in the TensorFlow environment. Finally, the trained model is used to monitor the islanding status. The results show the feasibility of the method.

1 Introduction

Distributed Generation (DG) has become an important type of electric power. It is significant to cope with global warming and promote sustainable development of economy. The Electric Power System (EPS) integrates various forms of DG, Increase transmission and distribution system transmission margin, improve system reliability. As such integration also raises some problems, the intermittent nature of renewable energy and its volatility pose many difficulties in the operation, control and protection of power grids. Island operation is one of them.

Islanding occurs when a DG with partial load is disconnected from the main grid, and DG continues to supply power to the isolated power system. The resulting isolated system is an isolated island. Due to the volatility of DG output, DG alone power supply system may endanger the safety of related equipment in the system. As a general rule, once an islanding operation is detected, it is required to immediately suspend power supply to the power system by the DG. Therefore, it is required that DGs have the function of islanding detection [1, 2].

At present islanding detection technology can be divided into three categories, namely, passive, active, based on the communication method [3, 4]. Communication-based methods are not as good as the other two in terms of reliability and cost, so the currently widely used islanding detection methods focus on passive or active detection. This paper studies a passive islanding detection method. Passive method is to determine whether there is an island by detecting whether the output of the inverter is abnormal when the distributed power supply is in an island running state. The relevant parameters for judging include terminal voltage, frequency, phase, and so on. In passive island detection methods, the more widely used voltage frequency detection method and the key rate of change of electricity detection method. Distributed power is in island operation, and the rate of change of power parameters such as power and frequency that are sensitive to system changes that will increase. It can be judged whether the system is in or out by detecting whether the amplitude or rate of change of these parameters exceeds the limit island running status [5]. The main problem of islanding detection and protection is that the threshold setting of various detection methods lacks a definite setting formula and mainly relies on the experience of experts, which makes the detection methods often have undetectable areas, that is, dead zones.

In recent years, some scholars have introduced the idea of machine-learning into the study of islanding threshold setting. The classification algorithm in machine-learning can be used to determine whether the system is in an isolated state and set a reasonable detection threshold. For example, [6] applied decision tree algorithm to islanding detection and detection threshold setting. Literature [7] verified the validity of C4.5 algorithm in islanding detection. By means of machine-learning, regression detection algorithm is used to allocate the detection weight of each detection variable value, to reduce the undetectable area that may be generated using a single variable detection and to increase the redundancy of the detection threshold.

Classification task is a kind of supervised learning problem, in which the output information is discrete classification, that is, given the input system parameters and whether the parameter is islanding or not, classification output is one of the mutually exclusive categories of the problem. Here, the output is islanding and non-islanding. The purpose of a classification task is to discover some form of relationship between input system parameters and output classes so that knowledge of the discovery can be used to estimate whether the system is operating with islanding.

Logistic Regression (LR) can be said to be the most widely used machine-learning [8] binary classification algorithm in the internet field. This algorithm has clear principle, simple code implementation, good classification effect, and strong universality. However, no research has been done in the field of power system islanding detection. If the LR algorithm can be introduced into island state detection, it can provide a new way for islanding detection.

Here, we first set up a distributed generation model, and set up the islanding event according to the possible state of the islanding and the non-islanding fault that may affect the islanding detection. Second, we select the feature quantity that has a high correlation with the islanding detection target based on the set of island operation events, the simulation data were obtained under different island operation conditions and the corresponding feature quantities were extracted. Then, the gradient regression method was used to train the logistic regression model under the environment of TensorFlow (TF). Finally, it is concluded that the

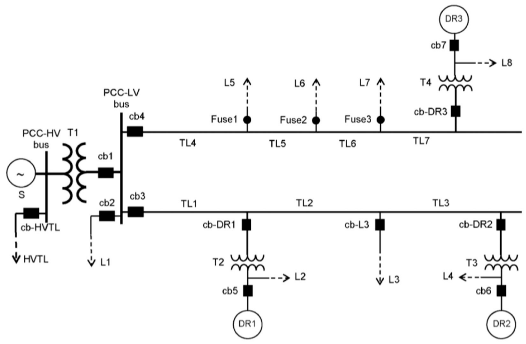

Fig. 1 *Typical DG topology*

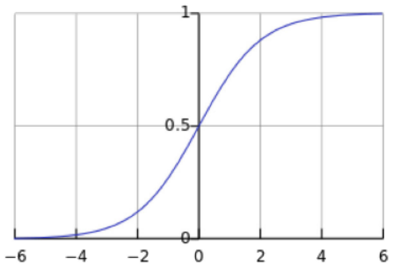

Fig. 2 *Predict function image curve*

islanding detection method based on logistic regression described has high accuracy and certain generalisation ability. Here, we introduce the deep learning framework of TF to introduce deep learning into the field of power system protection. For more complex problems, this method can be continuously improved and will have a broader application prospect in the future.

2 Distributed generation model established

Logistic regression method is a typical data mining classification algorithm. To train the logic regression classifier, a series of data of power system operation, including the data of the island state and the data of the non-island state, need to be obtained. These two types of data in combination with the state tags are used for classifier training.

Shown in Fig. 1 is a typical distributed generation system. Using Simulink tools in Matlab simulation software to establish the power system simulation model, as shown in Fig. 1, the distributed power DR1 is the target power supply. The model equates EPS to infinity power. System island operation state is divided into, single island and global island, two kinds of island forms: target DG feeder on single or multiple nodes off, then the target DG into the island running state, for example, the circuit breaker cb-DR1 is disconnected, the DR1 and period load brought into island operation; circuit breaker cb1 open end, then the entire distributed generation system into the global island operation.

To be able to train the classifier as comprehensively as possible, you need to pick as many running events as possible to get the data. At the same time, the selected events can be used to characterise the various operating states that may occur on the islanding system. During the island event setup, set the event generator to generate 72 events. The 50% (36) of these incidents are island incidents and the other 50% are non-island incidents. These 72 events were generated by a combination of eight possible events under nine network operating loads. The eight events are: (1) cb1 off; (2) cb3 off; (3) cb-DR2 off; (4) cb-DR3 off; (5) three phase faults on the PCC-LV bus consisting of cb1, cb3 and cb4 isolation clear; (6) TL1 ground fault occurs and the three phase fault on line TL1 is cleared by disconnecting TL1 from both sides; (7) ground fault of load circuit L3 cleared by cb-L3; (8) cb4 off. The nine operating loads are: Normal EPS (Power System) Load, Minimum EPS Load, Maximum EPS Load, Normal DG Load, Minimum DG Load, Maximum DG Load, Nominal DR1 Generation (85%), Minimum DR1 Generation (50%). The largest DR1 generation (100%). Incidents 1, 2, 5, and 6 are island events, while incidents 3, 4, 7, 8 are non-island incidents. These events are used to generate islanding feature information on the cb-DR1 (DR1 circuit breaker).

At the same time, it is assumed in the simulation event that the protective devices at the circuit breakers cb1, cb2, cb3, cb4 and cb-L3 are ideal instantaneous components and that the proposed method assumes that the protective devices have worked well together. Therefore, the order of disconnection between protective devices is not part of the methodological study here.

3 Method selection

3.1 Classification of island features

System parameters can include all sensitive system metrics that are affected by the islanding operation and can be measured locally. According to the characteristics of data sensitive to the island condition proposed in [5], the following four system parameters are selected to model the proposed method. The selected system parameters, the corresponding pattern vectors, and the data model in the database are given by the following mathematical expressions:

$$X_i = \left\{ f_i, V_i, \left(\frac{\Delta f}{\Delta t}\right)_i, \left(\frac{\Delta P}{\Delta t}\right)_i \right\} \tag{1}$$

$$\{(X_i, y_i), i = 1, 2, \dots, N\} \tag{2}$$

among them, i – event number, N – The total number of incidents, X_i – the ith pattern of the incident vector, f_i – The frequency of the ith event, V_i – The voltage change of the ith event, $(\Delta f/\Delta t)_i$ – The rate of change in frequency of the ith event, $(\Delta P/\Delta t)_i$ – The rate of change of the voltage of the ith event, $y_1, y_2, y_3, \dots y_i, \dots y_n$ is a variable corresponding to each type of event. The possible output can be defined as a set $\{c_0, c_1\}$, such as a $y_i = c_0 = 0$ non-islanding state and $y_i = c_1 = 1$ an islanding state.

In its general form, it is a schema database that contains the entire data model for all events and gets the values from the set of dimension real numbers. Its system parameters (or features) can be represented by the following expressions:

$$X = \{X1, X2, X3, X4\} \tag{3}$$

among them,

$$X1 = \{X_i(1), i = 1, 2, \dots, N\} \tag{4}$$

$$X2 = \{X_i(2), i = 1, 2, \dots, N\} \tag{5}$$

$$X3 = \{X_i(3), i = 1, 2, \dots, N\} \tag{6}$$

$$X4 = \{X_i(4), i = 1, 2, \dots, N\} \tag{7}$$

3.2 Logistic regression classification algorithm

There are only two of our outputs, 0 or 1, that correspond to the non-island state 0 and the island state 1 in the islanding detection system.

$$y = \begin{cases} 0 & g(z) \le 0.5 \\ 1 & g(z) \ge 0.5 \end{cases} \tag{8}$$

Therefore, the prediction function $g(z)$, we use Sigmoid function, the function is:

$$g(z) = \frac{1}{1 + e^{-z}} \tag{9}$$

The corresponding function image is an S-curve with a value between 0 and 1, as shown in Fig. 2.

For the case of a linear boundary, the boundary function can be constructed as follows:

$$\theta_0 + \theta_1 x_1 + \cdots \theta_n x_n = \sum_{i=0}^{n} \theta_i x_i = \theta^T x \quad (10)$$

n represents the number of selected features.

Construct the prediction function as:

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \quad (11)$$

$h_\theta(x)$ The value of the function represents the probability that the result is 1, and the probability of the system is in islanding operation. Therefore, the result of classification for x can be expressed by the following formula.

$$P(y|x;\theta) = (h_\theta(x))^y (1 - h_\theta(x))^{1-y} \quad (12)$$

Take the likelihood function as:

$$\begin{aligned} L(\theta) &= \prod_i^m P(y^{(i)} | x^{(i)};\theta) \\ &= \prod_i^m (h_\theta(x^{(i)}))^{y^{(i)}} (1 - h_\theta(x^{(i)}))^{1-y^{(i)}} \end{aligned} \quad (13)$$

Log likelihood function is:

$$\begin{aligned} l(\theta) &= \log L(\theta) \\ &= \sum_{i=1}^{m} (y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))) \end{aligned} \quad (14)$$

The classifier parameters make the log-likelihood function reach the maximum, then θ is the best parameter.

θ is to obtain that this paper uses the gradient descent method, the construction criteria function is as follows.

$$J(\theta) = -\frac{1}{m} l(\theta) \quad (15)$$

According to the gradient descent method available θ update process:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta), \quad (j = 0 \cdots n) \quad (16)$$

α for the learning step, seeking partial derivative for the learning step, seeking partial derivative

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \quad (17)$$

Therefore, (16) can be written as

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \quad (18)$$

Bringing the obtained θ back into the predictive function $g(z)$, the classification model is trained.

4 Example analysis

According to the islanding events set in Chapter 2, the corresponding settings are made on MATLAB. The distributed power DR1 is simulated for 72 times and the corresponding set of 72 characteristic data at cb-DR1 are recorded. According to [9], islanding target, DG trip time should not exceed 0.2 seconds. Therefore, it is considered that the islanding condition detection event should be completed within 0.2 s, so the data recording time

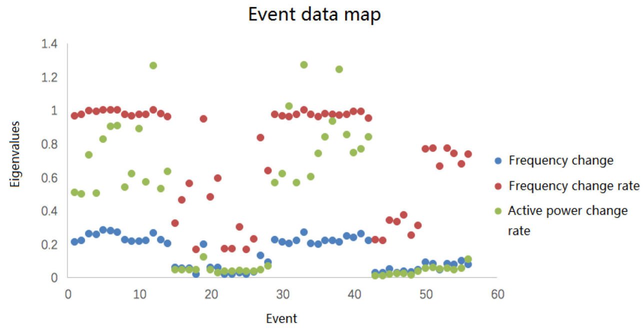

Fig. 3 *Graph of events*

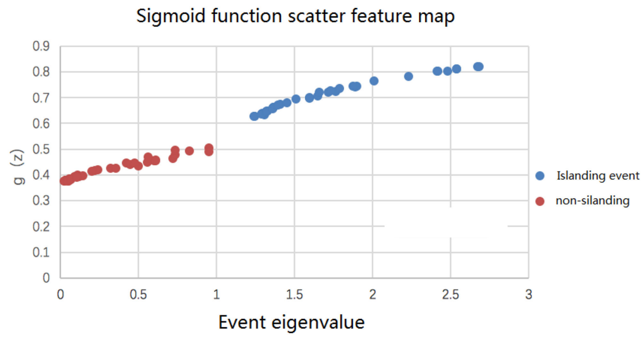

Fig. 4 *Graph of feature*

of 0.15 s is selected. The sampling rate of change rate data is set to eight cycles, that is, 0.8 ms.

72 sets of data, see the schedule, of which the first 56 sets of data as training data see Schedule 1, the latter 16 sets of data as validation data in Schedule 2.

The 56 groups of data are classified and aggregated, the horizontal axis is the event number, the vertical axis is the eigenvalue of each data, and the data are processed by per unitisation, wherein 1–14 and 29–42 groups of events are island events and others are non-island events shows in Fig. 3.

It can be seen that the frequency change and frequency change rate of the non-islanding events in group 19 are very close to those of islanding events, respectively, and the islanding thresholds are not well handled and easily misjudged.

In the TF environment, 56 sets of training data are used to train the logistic regression classifier. It can be seen from the simulation results that the voltage does not change much in the isolated island and non-isolated island. Therefore,

$$X_i = \left\{ f_i, V_i, \left(\frac{\Delta f}{\Delta t}\right)_i, \left(\frac{\Delta P}{\Delta t}\right)_i \right\}$$

is chosen as the modelling parameter, $i = 56$; $n = 3$; the selected training step $\alpha = 0.001$, the number of iterations $m = 10{,}000$ times, $\theta_0 = -0.6291306$, $\theta_1 = 0.23079669$, $\theta_2 = 0.47317266$, $\theta_3 = 1.27615726$. Incident event data characteristic value is $x = x_1^2 + x_2^2 + x_3^2$, after data processing, the predicted function scatter characteristic diagram is shown in Fig. 4:

As can be seen from Fig. 4, some non-island operation data $g(z)$ output > 0.5, it cannot be effectively classified. Increase the number of iterations $m = 20{,}000$, have $\theta_0 = -1.23738182$, $\theta_1 = 0.43590146$, $\theta_2 = 0.86694384$, $\theta_3 = 2.03876638$, after the data are processed to predict the function scatter characteristics shown in Fig. 5:

It can be seen from Fig. 5 that all of $g(z)$ output results <0.5 are all non-islanding operation events and $g(z)$ output results>0.5 are all island operation events.

Fourteen sets of validation data were categorised using trained prediction functions (seven non-island cases and seven island cases).

The predicted results of the 14 sets of test data are shown in Fig. 6.

The results of the test are shown in the following Tables 1–3. The results are as follows: The accuracy of the proposed method

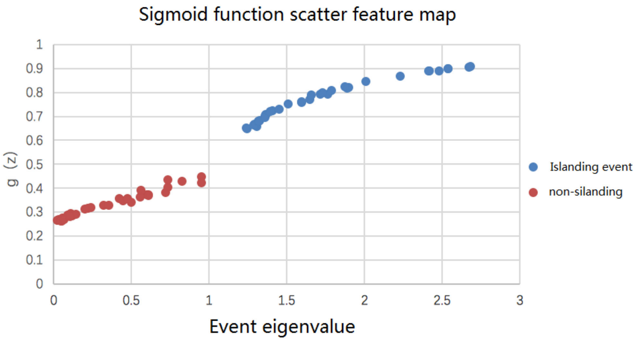

Fig. 5 *Graph of train data feature*

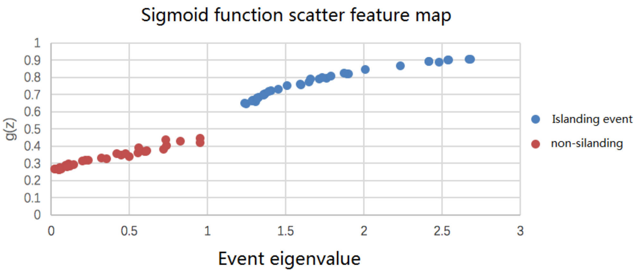

Fig. 6 *Graph of test data feature*

Table 1 Performance of logistic regression

Classifier	The total number of incidents	The correct number of categories	Correct rate
logistic regression	16	16	100%

for classification of existing data is 100% for both non-island case and island case, and has good generalisation ability.

However, this result does not prove that the proposed method has no shortcomings. First of all, although this paper tries to simulate a variety of islanding and non-islanding events possibly occurring in the system as far as possible, the total amount of islanding events set is limited after all. At the same time, distributed power only considers the synchronisation. The power supply of the generator, without regard to the grid-connected DG via the inverter, poses a certain challenge to the validity of the method if the grid structure is more complicated and the operation status is more diversified.

To deal with the more complex operating environment, we can improve from two aspects. One is that we can choose more kinds of feature data, allocate the weight of each feature data reasonably, improve the prediction ability and increase the threshold redundancy. Second, when more data features based on the introduction of the theory of deep learning into the prediction model to improve the model's ability to regress complex non-linear data characteristics.

5 Conclusion

Based on the idea of machine-learning, this paper introduces the logic regression algorithm into the islanding detection, and trains a prediction function using the feature data under the TF environment. The prediction function is used to detect whether the islanding occurs or not, and good results have been obtained. The advantage of using the proposed method is that based on the inherent professional knowledge, the idea of machine-learning can automatically mine the logic behind the events and reasonably distribute the weight indexes of different characteristic data so as to improve the redundancy of system island detection thresholds redundancy, and as the amount of data increases, the accuracy of monitoring can be predictably enhanced. At the same time, this paper introduces the TF environment into the field of power system processing, and uses its built-in data processing module to improve the modelling and training speed and provide the operating platform for the application of more sophisticated depth learning algorithms.

Table 2 Training data for logistic regression

Serial number	f/Hz	(df/dt)/(Hz/s)	(dP/dt)/(MW/s)	Island status
1	0.210	0.00967	5097	1
2	0.2192	0.00975	4965	1
3	0.2593	0.00996	7332	1
4	0.2577	0.00994	5030	1
5	0.282	0.010	8252	1
6	0.280	0.010	9006	1
7	0.271	0.010	9085	1
8	0.224	0.009755	5374	1
9	0.216	0.009645	6204	1
10	0.216	0.009737	8909	1
11	0.222	0.009745	5696	1
12	0.264	0.01	12662	1
13	0.225	0.00977	5296	1
14	0.201	0.009625	6319	1
15	0.0573	0.003233	463	0
16	0.0563	0.004611	478	0
17	0.0549	0.005608	445	0
18	0.0195	0.001681	437	0
19	0.1993	0.009466	1215	0
20	0.0605	0.004829	466	0
21	0.0591	0.005930	283	0
22	0.0202	0.001697	380	0
23	0.0184	0.001728	380	0
24	0.0285	0.002992	415	0
25	0.0208	0.001648	371	0
26	0.0322	0.002302	366	0
27	0.1324	0.008367	441	0
28	0.0909	0.006389	699	0
29	0.2230	0.009741	5642	1
30	0.2109	0.009649	6206	1
31	0.2018	0.009605	10221	1
32	0.2219	0.009745	5668	1
33	0.2685	0.01	12,687	1
34	0.2039	0.009755	6040	1
35	0.1984	0.009587	7417	1
36	0.2196	0.009795	8400	1
37	0.2202	0.009759	9329	1
38	0.2131	0.009715	12,445	1
39	0.2466	0.009728	8511	1
40	0.2385	0.009902	7461	1
41	0.2594	0.009940	7697	1
42	0.2195	0.009495	8400	1
43	0.0289	0.00226	86	0
44	0.0291	0.002186	93	0
45	0.0516	0.003396	202	0
46	0.0279	0.003319	232	0
47	0.0356	0.003747	240	0
48	0.0341	0.002508	147	0
49	0.0461	0.003090	362	0
50	0.0909	0.007657	550	0
51	0.0828	0.007735	598	0
52	0.0446	0.006643	490	0
53	0.0826	0.007731	534	0
54	0.0755	0.007397	478	0
55	0.1004	0.006796	532	0
56	0.0753	0.007366	1108	0

Table 3 Testing data for logistic regression

Serial number	f/Hz	(df/dt)/(Hz/s)	(dP/dt)/(MW/s)	Island status
1	0.277	0.0101	11,753	1
2	0.282	0.01	10,732	1
3	0.198	0.009545	12,104	1
4	0.200	0.009559	6730	1
5	0.0049	0.007063	285	0
6	0.015	0.001399	458	0
7	0.1255	0.008435	869	0
8	0.1886	0.009545	732	0
9	0.1917	0.009515	12,130	1
10	0.1969	0.009531	6663	1
11	0.2132	0.009715	12,445	1
12	0.2137	0.009712	6807	1
13	0.0699	0.004385	485	0
14	0.0245	0.003187	206	0
15	0.1400	0.008314	1508	0
16	0.1546	0.008869	1179	0

6 Acknowledgments

The work described in this paper was fully supported by a grant from the Project Supported by State Grid Corporation of China (No. GYW17801700097).

7 References

[1] 'IEEE standard for interconnecting distributed resources with electric power systems'. IEEE Std. 1547-2003, June 2003

[2] The Electricity Council: 'Engineering Recommendation G.59/1, Recommendations for the Connection of Embedded Generating Plant to the Regional Electricity Companies', *Distribution Systems*, (The Electricity Council, UK, 1991)

[3] Funabashi, T., Koyanagi, K., Yokoyama, R.: 'A review of islanding detection methods for distributed resources'. IEEE Bologna Power Tech Conf., Bologna, Italy, June 23–26, 2003, vol. 2, pp. 23–26

[4] Xu, W., Mauch, K., Martel, S.: 'An Assessment of Distributed Generation Islanding Detection Methods and Issues for Canada'. CANMET Energy Technology Centre-Varennes, National Resources Canada, Rep. 2004-074(TR), July 2004

[5] Jang, S.I., Kim, K.H.: 'An islanding detection method for distributed generations using voltage unbalance and total harmonic distortion of current', *IEEE Trans on Power Delivery*, 2004, **19**, (2), pp. 745–752

[6] El-Arroudi, K., Joós, G.: 'Data mining approach to threshold settings of islanding relays in distributed generation', *IEEE Trans. Power Syst.*, 2007, **22**, (3), pp. 1112–1119

[7] Xia, L., Xiaofeng, D., Yuping, L., *et al.*: 'A method of distributed system island detection based on data mining', *Power System & Automation*, 2010, **23**, (2), pp. 38–44

[8] Cox, R.: 'The regression analysis of binary sequences', *J. R. Stat. Soc. Ser. B (Methodological)*, 1958, **20**, (2), pp. 215–242

[9] State Grid Corporation.: '*Q/GDW 480-2010 technical specification for access to distributed power Systems*' (China Electric Power Press, Beijing, 2010)