<table>
<thead>
<tr>
<th>Title</th>
<th>Comparative study in determining features extraction for islanding detection using data mining technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Aziah Khamis; Xu, Yan; Azah Mohamed</td>
</tr>
<tr>
<td>Date</td>
<td>2017</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/50251">http://hdl.handle.net/10220/50251</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2017 Institute of Advanced Engineering and Science. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.</td>
</tr>
</tbody>
</table>
Comparative Study in Determining Features Extraction for Islanding Detection using Data Mining Technique: Correlation and Coefficient Analysis

Aziah Khamis*, Yan Xu, and Azah Mohamed

1Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Malaysia
2School of Electrical and Information Engineering, The University of Sydney, NSW, Australia
3School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
4Department of Electrical, Electronic and System Engineering, Universiti Kebangsaan Malaysia, Malaysia

ABSTRACT

A comprehensive comparison study on the data mining based approaches for detecting islanding events in a power distribution system with inverter-based distributed generations is presented. The important features for each phase in the island detection scheme are investigated in detail. These features are extracted from the time-varying measurements of voltage, frequency and total harmonic distortion (THD) of current and voltage at the point of common coupling. Numerical studies were conducted on the IEEE 34-bus system considering various scenarios of islanding and non-islanding conditions. The features obtained are then used to train several data mining techniques such as decision tree, support vector machine, neural network, bagging and random forest (RF). The simulation results showed that the important feature parameters can be evaluated based on the correlation between the extracted features. From the results, the four important features that give accurate islanding detection are the fundamental voltage THD, fundamental current THD, rate of change of voltage magnitude and voltage deviation. Comparison studies demonstrated the effectiveness of the RF method in achieving high accuracy for islanding detection.

Copyright © 2017 Institute of Advanced Engineering and Science. All rights reserved.

Corresponding Author:
Aziah Khamis
Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Malaysia.
Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.
aziah83@gmail.com

1. INTRODUCTION

A small localized power source called as distributed generation (DG) becomes an alternative to bulk electric generation due to yearly demand growth. These DGs can be in the form of wind farm, micro hydro turbine and photovoltaic (PV) generator. Generally, these DGs are in the range of kW up to MW with several advantages such as environmental benefits, improved reliability, increased efficiency, improved power quality and reduced transmission and distribution line losses [1–3]. However, one of the major drawbacks of DGs is when subjected to islanding mode of operation. Islanding is referred as disconnection of the main source in which it can be operated either intentional or unintentional. When disconnection occurs, the active part of the distribution system should sense the disconnection from the main grid and shut down the DGs, where island operation is prohibited or control action must be activated to stabilize the islanded part of system [4, 5]. Islanding operation has some benefits but several drawbacks are still observed, especially in unintentional islanding events which may cause problems related to power quality, safety, voltage and frequency stabilities, and interference [6, 7].

Various techniques have been developed to detect islanding. Islanding techniques can generally be classified into remote and local methods. Remote methods are based on communication between the power utility and the DGs. Remote methods are highly reliable, but the practical implementation of these schemes can be inflexible, complex and
expensive. For instance, the cost of implementing a remote method can be extremely expensive especially when it is implemented in networks that do not initially have any communication infrastructure with the power utility. Therefore, local methods are favourable for detecting islanding condition. These local methods can be categorized as passive, active and hybrid techniques [8–10]. The passive islanding detection technique monitors the system parameters such as voltage, current, frequency and harmonic distortion at the point of common coupling (PCC) with the utility grid for detecting events [3,11–13]. In the active islanding detection technique, disturbances are intentionally injected into the network and the island is detected based on the system responses to the disturbances [6, 14–16]. Meanwhile, the hybrid technique is a combination of the active and passive techniques, in which active technique is applied only if islanding is not detected by the passive technique [3, 17–20].

Data mining is widely used in numerous area including islanding detection [21–24]. For instance, an intelligent islanding detection technique was developed in [25] using decision tree (DT) classifier to identify and classify islanding operations at specific target locations. However, the DT classifier is not capable in capturing all possible islanding events. To improve the accuracy of the DT classifier, fuzzy rule-based incorporated with DT was utilized in detecting the islanding events [26]. In [13], a statistical signal processing algorithm is applied by using features from voltage and frequency waveforms. The accuracy of this technique is acceptable, but the delay in statistical processing makes this technique slower than other islanding detection techniques. Realizing the potential of data mining techniques for islanding detection, new techniques have been developed by combining the discrete wavelet transform with various classifiers, namely, DT, probabilistic neural-network (PNN) and support vector machines (SVM) [27]. The test results showed that the best accuracy can be achieved by the DT classifier model [27]. In [28] a pattern recognition approach based on the DT classifier was employed for islanding detection. However, DT classifier have limitations, such as possibility of spurious relationships, possibility of duplication with the same sub-tree on different paths and limited to one output per attribute, and inability to represent test that refer to two or more different objects, which requires an exploration of others intelligent technique. On the basis of the comprehensive literature review, the data mining using correlation and coefficient analysis had rarely been reported. Therefore, the main objective of this study is to propose a new islanding scheme using the correlation and coefficient analysis for features extraction and data mining techniques. Initially, features are extracted using the correlation and coefficient analysis in which seven parameter indices at the target DG location have been identified as important features for identifying the islanding events. Then five different data mining techniques, namely, DT, SVM, neural network (NN), bagging and random forest (RF) have been developed as classifiers in islanding detection. The proposed islanding detection scheme is tested on the IEEE 34 bus system with inverter based DGs.

2. BUILDING THE DATA SET

2.1. Test System

Fig. 1 shows the single-line diagram of the IEEE 34-bus distribution system model in MATLAB/SIMULINK software. The DG and the load are connected to distribution system by a 100-kVA 24.9-kV/480-V transformer. Meanwhile, the PCC is connected with R load with 100-kW. The DG is an inverter-based DG with current controlled interface using the same control units in the previous study [29].

![Figure 1. System under test: IEEE 34-bus system.](image)

2.2. Database Generation

Various islanding and non-islanding events should be generated with a wide range of dataset for training the classifier. The possible situations that may create islanding and non-islanding conditions are given as follows:

i. Load and capacitor switching at different buses,

ii. Several types of fault at different busses, and

iii. Event that can trip breaker and reclosers, and island the DG.
The above situations are simulated under possible variation in operating condition which are considered as:
i. Normal DG loading,
ii. Different operating points that cause power mismatch at the local load connected at bus 848.

2.3. Features Selection

The main idea of features selection is to choose the most significant input variables by eliminating features with non/less-predictive information. The use of significant features can greatly improve the classifier model performance and thus, increase the prediction accuracy as well as the computational speed. In this paper, the combination of various features parameters has been chosen from previous islanding detection methods focusing on inverter-based-DG. The extracted features include $X_a$ frequency deviation ($\Delta f$), $X_b$ voltage deviation ($\Delta V$), $X_c$ rate of change of voltage magnitude ($\Delta V/\Delta t$), $X_d$ fundamental current total harmonic distortion ($THD_C f$), $X_e$ current total harmonic distortion ($THD_C$), $X_f$ fundamental voltage total harmonic distortion ($THD_V f$) and $X_g$ voltage total harmonic distortion ($THD_V$).

The features are extracted by per phase basis in order to identify the most essential feature parameters for islanding detection. Figs. 2 and 3 show examples of features signals obtained from islanding event for phase-A at DG terminal in distribution system. The signals in Figs. 2a-c and d-f represents the voltage and frequency of phase-A during islanding condition case, respectively. The signals in Figs. 2b and c are the voltage deviation ($\Delta V$) and rate of change of voltage magnitude ($\Delta V/\Delta t$), respectively, obtained from the voltage signal of Fig. 2a. The frequency signals of Fig. 2d are evaluated to get the frequency deviation ($\Delta f$) as illustrated in Fig. 2f. Meanwhile, the information of THD for voltage and current are selected as shown in Figs. 3a and b. The entire features information is then utilized as the input for the classifier. The features are then rearranged and expressed as,

$$\text{Input} = \begin{bmatrix} X_{a1} & X_{b1} & X_{c1} & X_{d1} & X_{e1} & X_{f1} & X_{g1} \\ X_{a2} & X_{b2} & X_{c2} & X_{d2} & X_{e2} & X_{f2} & X_{g2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{ay} & X_{by} & X_{cy} & X_{dy} & X_{ey} & X_{fy} & X_{gy} \end{bmatrix}$$

where $X_a$ is referred to the frequency deviation ($\Delta f$), $X_b$ is referred to the voltage deviation ($\Delta V$), $X_c$ is referred to rate of change of voltage magnitude ($\Delta V/\Delta t$), $X_d$ is referred current total harmonic distortion ($THD_C$), $X_e$ is referred to fundamental current total harmonic distortion ($THD_C f$), $X_f$ is referred voltage total harmonic distortion ($THD_V$), $X_g$ is referred to fundamental voltage total harmonic distortion ($THD_V f$) and $y$ is referred to the number of points taken after the disturbance detected.
Figure 2. Example features extraction for islanding case: (a) Phase A voltage signal, (b) Voltage deviation (ΔV), (c) Rate of change of voltage (ΔV)/(Δt), (d) Phase A frequency signal, (e) Zoom in frequency after disturbance, (f) Frequency deviation (Δf).

Figure 3. Example features extraction for islanding case: (a) Voltage total harmonic distortion ($THD_V$), (b) Current total harmonic distortion ($THD_C$).
3. FEATURE EXTRACTION USING CORRELATION AND COEFFICIENT ANALYSIS

The inclusion of irrelevant and redundant features extraction in the classifier model may result in poor performance in classification accuracy and increases the computation time. To obtain high classification accuracy, high quality of features need to be extracted in describing the islanding events using the correlation and coefficient analysis. Fig. 4 shows the correlation between 28 features variable. The colours and shape element in the figure are used to show the degree of correlation [30]. The variables are said to have perfect correlation with itself, which is in the diagonal lines on the diagonal of the graphic (see Fig. 4). The blue colours shows the positive value, whereas the red for negative value that used to encode the sign of correlation. Meanwhile, filled circled means positive value, while anti-clockwise is for negative values. In this analysis, the Pearson correlation coefficient is utilized to measure the strength between 28 variable features. Mathematically, the coefficient is expressed as follows:

\[
r = \frac{N \sum kl - (\sum k)(\sum l)}{\sqrt{[N \sum k^2 - (\sum k)^2][N \sum l^2 - (\sum l)^2]}}
\]  

where \( N \) is referred to number of pairs of scores, \( \sum kl \) is referred to sum of the products of paired scores, \( \sum k \) is referred to sum of \( k \) scores, \( \sum l \) is referred to sum of \( l \) scores, \( \sum k^2 \) is referred to sum of squared \( k \) scores, and \( \sum l^2 \) is referred to sum of squared \( l \) scores.

Figure 4. Visual summary of correlation between the 28 candidate attributes for phase A.

For instance, Fig. 4 shows that the most positive correlation variable is \( X_g \), where most of the relationship between the variables are in positive value. The relation correlation between \( X_{b,1} \) and \( X_{c,1} \), \( X_{b,1} \) and \( X_{c,4} \), and \( X_{b,3} \) and \( X_{c,1} \) are evaluated as -0.6746369, -0.6300237 and -0.3214842, respectively. Therefore, the circle with red colours in Fig. 4, show the negative correlated between \( X_b \) and \( X_c \). This finding proves that \( X_b \) is the most negative correlation between the features as shown in Fig. 4.

The significant of the variables is again highlighted by the importance analysis report from the RF learning as illustrated in Fig. 5. The figure shows that the top four variable are listed as \([X_g, X_e, X_b, X_c]\). The out-of bag accuracy-based ranking results in approximately same with the top four, even though the \( X_b \) should be substituted to the lower correlated with \( X_g \).

Similar to the islanding detection classifier procedure adopted to phase-A, classifier training and testing data set procedures are applied on the other two phases namely, phase B and C. Figs. 6(a) and 6(b) shows the correlation between 28 features variable for phase-B and C, respectively. The figure shows the correlation relationship with the 28 features variable by depicting the pattern of relations among the variables. Meanwhile, Fig. 7 shows that the important behaviour report from the RF model classifier for phase-B and C. Phase-B and C show an equal important variable. The observation likewise reveals that the top four variable are listed as \([X_g, X_e, X_b, X_c]\).
Figure 5. Top-down importance of variable according to accuracy loss or misclassification rate reduction (gini) for phase A.
Figure 6. Visual summary of correlation between the 28 candidate attributes: (a) phase B, (b) phase C.
Figure 7. Top-down importance of variable according to accuracy loss or misclassification rate reduction (gini): (a) phase B, (b) phase C.
4. IMPLEMENTATION OF DECISION TREE AND RANDOM FOREST AS CLASSIFIERS

Fig. 8 illustrates the DT structure for the islanding classification model for inverter-based DG consists of 8 nodes. At the top of the tree, the value of $X_e$ is first compared with the threshold value 0.632898 and it will split into two descendent subsets. This subset is then split into several leaf nodes which are designated by a class label. There are two class label in this study, namely, islanding and non-islanding cases. From the figure, all the cases having $X_e$ within 0.63 and 0.65 are predicted as non-islanding state. However, for cases with $X_e$ less than 0.63, the classification depends on the value of $X_b$ and $X_c$.

Fig. 9 shows the multidimensional scaling (MDS) plot for islanding and non-islanding events utilizing the RF classifier. This MDS is used to discover the underlying structure of distance measured between objects. The MDS assign the observations to specific locations in a conceptual space (commonly 2 or 3 dimensional space used), thus the distance between points in space match the given dissimilarities as closely as possible.

5. TEST RESULTS

The simulation data were obtained using MALTAB/SIMULIK software and the data were randomly divided into training and testing data set as summarized in Table 1. The features are extracted from the information given in (1). The open-source software, Rattle is used to implements the conventional DT, bagging and RF classifier. For easy comparison, all the classifier use the same training and testing data sets which gives two predictors of class label called as islanding and non-islanding events. Table 2 shows classification results for testing data set of phase A with three different classifiers, namely DT, bagging and RF classifiers. This result reveals that the highest accuracy can be achieved with the RF classifier with percentage classification of 98.9% and 100% for the non-islanding and islanding events.
cases, respectively.

Table 1. NUMBER OF SAMPLE

<table>
<thead>
<tr>
<th></th>
<th>Non-islanding</th>
<th>Islanding</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training data set</td>
<td>95</td>
<td>91</td>
<td>186</td>
</tr>
<tr>
<td>Testing data set</td>
<td>91</td>
<td>94</td>
<td>185</td>
</tr>
</tbody>
</table>

Table 2. CLASSIFICATION RESULTS ON TESTING DATA SETS FOR PHASE A

<table>
<thead>
<tr>
<th>Classifier Model</th>
<th>No of Cases</th>
<th>Actual Class</th>
<th>Non-islanding</th>
<th>Islanding</th>
<th>Classification Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DecisionTree</td>
<td>91</td>
<td>Non – islanding</td>
<td>78</td>
<td>4</td>
<td>85.71</td>
</tr>
<tr>
<td></td>
<td>94</td>
<td>Islanding</td>
<td>13</td>
<td>90</td>
<td>95.74</td>
</tr>
<tr>
<td>Bagging</td>
<td>91</td>
<td>Non – islanding</td>
<td>89</td>
<td>1</td>
<td>97.80</td>
</tr>
<tr>
<td></td>
<td>94</td>
<td>Islanding</td>
<td>2</td>
<td>93</td>
<td>98.94</td>
</tr>
<tr>
<td>RandomForest</td>
<td>91</td>
<td>Non – islanding</td>
<td>90</td>
<td>0</td>
<td>98.90</td>
</tr>
<tr>
<td></td>
<td>94</td>
<td>Islanding</td>
<td>1</td>
<td>94</td>
<td>100</td>
</tr>
</tbody>
</table>

Further comparison is then made for islanding detection using SVM, NN, DT, bagging, and RF classifiers considering all the three phases. The performances of accuracy of these classifiers is evaluated as shown in Fig. 10 and Table 3. Table 3 show the accuracy of the five classifiers for islanding detection at each phase, i.e, phase-A, B and C. For all the phases, the RF classifier gives the highest accuracy compared to the other classifiers in detecting islanding events as indicated in bold. This result proves that the best classifier model to predict the islanding condition based on per phase feature extraction can be obtained using the RF classifier.

![Figure 10. Accuracies of various model](image-url)
Table 3.

PERFORMANCE OF VARIOUS MODEL CLASSIFIER IN TERM OF ACCURACIES

<table>
<thead>
<tr>
<th>Classification Model</th>
<th>Accuracies</th>
<th>Average Accuracies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phase A</td>
<td>Phase B</td>
</tr>
<tr>
<td>SupportVectorMachine(SVM)</td>
<td>0.9892</td>
<td>0.973</td>
</tr>
<tr>
<td>NeuralNetwork(NN)</td>
<td>0.8726</td>
<td>0.9323</td>
</tr>
<tr>
<td>DecisionTree(DT)</td>
<td>0.9081</td>
<td>0.9622</td>
</tr>
<tr>
<td>Bagging</td>
<td>0.9838</td>
<td>0.946</td>
</tr>
<tr>
<td>RandomForest(RF)</td>
<td>0.9946</td>
<td>0.9946</td>
</tr>
</tbody>
</table>

6. CONCLUSION

A new islanding detection scheme for a power distribution system with inverter-based DG has been developed. The proposed scheme implements feature extraction using correlation and coefficient analysis for each phase and data mining techniques for classifying the islanding events. Unlike previous research works which ignore the per phases analysis, the proposed islanding scheme work progressively at each phase for a comprehensive identification of islanding cases. The simulation results on the IEEE 34 bus system with inverter-based DG showed that the proposed islanding detection classifier using per phase feature extraction can accurately detect the islanding events. A comparison between various classifiers, namely, SVM, NN, DT, bagging and RF has been made and the results showed that the highest accuracy in detecting islanding events can be achieved from the RF classifier.

ACKNOWLEDGEMENT

The authors wish to acknowledge the Ministry of Education, Malaysia and Universiti Teknikal Malaysia Melaka, for supporting this research financially under Grants RAGS/1/2015/TK0/FKE/02/B00092.

REFERENCES


Aziah Khamis received her B.Eng. from Universiti Putra Malaysia (UPM), Malaysia in 2006, M.Sc from Newcastle University, United Kingdom (UK) in the year 2009 and the Ph.D. degree from Universiti Kebangsaan Malaysia(UKM), Malaysia in the year 2014. She was a Post-doctoral researcher at School of Electrical and Information Engineering, The University of Sydney, Australia. She is now a senior lecturer in Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. Her main research interests include intelligent system application of power system study, distributed generation, and microgrid.

Further info on her homepage: https://sites.google.com/site/aziahkhamis/

Yan Xu (S10-M13) received the B.E. and M.E degrees from South China University of Technology, Guangzhou, China in 2008 and 2011, respectively, and the Ph.D. degree from The University of Newcastle, Australia, in 2013. He is now an Assistant Professor with Nanyang Assistant Professorship at School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include power system stability, control, and optimization, microgrid, and data-analytics for power engineering.

Azah Mohamed received her B.Sc. from University of London in 1978 and M.Sc. and Ph.D. from Universiti Malaya in 1988 and 1995, respectively. She is a professor in the Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan Malaysia. Her main research interests are power system security, power quality, artificial intelligence and renewable energy systems. She is a senior member of IEEE. She can be reached at azah@eng.ukm.my.