Deep learning for classification of islanding and grid disturbance based on multi-resolution singular spectrum entropy

Tie Li1, Xiaoyang He1, Junci Tang1, Hui Zeng2, Chunying Zhou1, Nan Zhang1, Hui Liu3, Zhuoxin Lu4,5, Xiangrui Kong4 and Zheng Yan4

1 Liaoning Electric Power Co. Ltd, State Grid Corporation of China, Shenyang 110006, China
2 Research Institute of Liaoning Power Co., Ltd, State Grid Corporation of China, Shenyang 110006, China
3 State Grid Corporation of China, Beijing 100017, China
4 Shanghai Jiao Tong University, Shanghai 200240, China
5 Zhuoxin_Lu@sjtu.edu.cn

Abstract. Forasmuch as the distinguishment of islanding is easy to be interfered by grid disturbance, island detection device may make misjudgment thus causing the consequence of photovoltaic out of service. The detection device must provide with the ability to differ islanding from grid disturbance. In this paper, the concept of deep learning is introduced into classification of islanding and grid disturbance for the first time. A novel deep learning framework is proposed to detect and classify islanding or grid disturbance. The framework is a hybrid of wavelet transformation, multi-resolution singular spectrum entropy, and deep learning architecture. As a signal processing method after wavelet transformation, multi-resolution singular spectrum entropy combines multi-resolution analysis and spectrum analysis with entropy as output, from which we can extract the intrinsic different features between islanding and grid disturbance. With the features extracted, deep learning is utilized to classify islanding and grid disturbance. Simulation results indicate that the method can achieve its goal while being highly accurate, so the photovoltaic system mistakenly withdrawing from power grids can be avoided.

1. Introduction
The increasing penetration of distributed generation is forcing system operators to account for its influences. Unexpected islanding conditions resulted from line fault and other reasons might cause serious hazards. Islanding detection is a technical problem that must be solved for grid-connected distributed generation [1]-[3]. Nevertheless, unexpected grid disturbance may interfere the reliability of islanding detection and cause misjudgment, from which grid disturbance might be recognized as islanding, thus causing DG out of service [4]-[6]. Therefore, islanding detection is supposed to be capable of differentiation between islanding and grid disturbance. This problem has received some attention in recent years. Islanding detection methods can be divided into communication methods, active methods, and passive methods [7]. The high cost of the communication method is barrier to its application. The active method has an adverse effect on grid operation due to injection signals. The mainstream method is the passive method. It extracts voltage and frequency signals at the point of
common coupling (PCC) and compares the signals with a given threshold value. It is quite convenient, but the threshold values are usually set empirically, which might be misleading and unreliable. Meanwhile, using conventional wavelet energy coefficients as eigenvectors is susceptible to the noise caused by an increasing amount of power electronics equipment. In [4]-[6], voltage value and frequency at the point of common coupling (PCC) are extracted for wavelet transformation, then the absolute values of the coefficients are acquainted for the comparison with the set threshold values of voltage and frequency. It is recognized as islanding only if the two numerical values exceed the threshold values simultaneously, else as other grid disturbance. Nonetheless, the threshold value is set by experiments and experience, and the two values of real islanding does not necessarily exceed the threshold value at the same time.

It is the crux of the matter that how to utilize islanding and disturbance signals more effectively so that the distinct feature vectors can be acquainted. Merging multi-resolution analysis and information entropy, wavelet entropy has been introduced into power system analysis resulted from the excellent ability for signal analysis. Based on singular spectrum entropy, multi-resolution singular spectrum entropy has been applied successfully in fault diagnosis. Multi-resolution singular spectrum entropy focuses on digging essential feature of signals with no effects from wavelet coefficients. It is an effective postprocessing method of wavelet to eliminate noise interference. In [8], a multi-resolution singular spectrum entropy-based method is proposed to process the signals. However, the accuracy of the SVM-based method with tiny sample size in that paper is not high enough. This paper combines deep learning with multi-resolution singular spectrum entropy to classify islanding and grid disturbance. Deep learning is making significant advances in solving classification problems that have resisted the best attempts of the artificial intelligence community for many years. It has turned out to be very good at discovering intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government. Deep learning involves a class of models which try to hierarchically learn deep features of input data with very deep neural networks. The simulation results indicate that this approach is able to distinguish islanding from grid disturbance accurately so that the maloperation of grid-connected distributed generation can be avoided.

2. Feature extraction based on multi-resolution singular spectrum entropy

2.1. Multi-resolution singular spectrum entropy calculation

(1) By selecting appropriate wavelet basis function and decomposition layer, discrete signal \( f(k) (k=1,2,\ldots,N) \) is processed using Mallat algorithm. The discrete dyadic wavelet transform of discrete signal can be determined by (1) [9]-[10].

\[
\begin{align*}
  c_{j+1}(k) &= Hc_j(k) \\
  d_{j+1}(k) &= Gc_j(k)
\end{align*}
\]

As in (1), \( H \) and \( G \) are low pass filter and high pass filter, respectively. \( c_j \) and \( d_j \) indicate the approximate part and the detailed part of signal scale. After the decomposition of scale 1, 2, \ldots, \( j \) (\( j \) is the decomposition layer), \( f(k) \) is decomposed into \( d_1, d_2, \ldots, d_j, c_j \), which indicate information of different bands from high frequency to low frequency, respectively.

(2) For the decomposed signal of each layer, wavelet transform coefficient reconstruction is conducted by (2).

\[
c_{j-1} = H^*c_j + G^*d_j
\]

As in (2), \( H^* \) and \( G^* \) are the dual operators of \( H \) and \( G \), respectively.

(3) Reconstruct the reconstruction signal of each layer in phase space.

Assumption to reconstruct a \( n \)-dimensional phase space. Let the reconstruction signal of layer \( j \) be \( D_j=[d_j(k)] \), from which \( d_j(1), d_j(2), \ldots, d_j(n) \) is supposed to be the first vector of the \( n \)-dimensional phase space. Then, take \( d_j(2), d_j(3), \ldots, d_j(n+1) \) as the second vector. By this analogy, a \( (N-n+1)\times n \) dimensional matrix \( A \) is constructed.

(4) For the matrix \( A \) of each layer, singular value decomposition (SVD) is conducted to calculate
the singular spectrum entropy. The matrix $A_{(N,n+1)\times l}$ is decomposed using SVD and the result is $A=U_{(N,n+1)}V_{l\times l}^T$. The nonzero diagonal elements $\lambda_i (i=1,2,3,\cdots,l)$ are singular values of the matrix $A$ from layer $j$. According to the informational entropy theory, definition of the signal singular spectrum entropy is as follows. As in (4) and (5), $H_j$ is the information entropy of level $j$; the $p_{ji}$ is the uncertain probability distribution of $\lambda$.

$$A=\begin{bmatrix} d_j(1) & d_j(2) & \cdots & d_j(n) \\ d_j(2) & d_j(3) & \cdots & d_j(n+1) \\ \vdots & \vdots & \ddots & \vdots \\ d_j(N-n+1) & d_j(N-n) & \cdots & d_j(N) \end{bmatrix}$$  \quad (3)$$

$$H_j=-\sum_{i=1}^{l} p_{ji}(2p_{ji}) \quad (4)$$

$$p_{ji}=\frac{\lambda_j}{\sum_{i=1}^{l} \lambda_i} \quad (5)$$

2.2. The Application mechanism analysis of multi-resolution singular spectrum entropy

Multi-resolution singular spectrum entropy is defined as a kind of wavelet entropy based on different principles and approach [11]. Resulted from the certain measurement given by entropy, multi-resolution singular spectrum entropy can represent the essential character of the signal, which is exceedingly suitable for the feature extraction of islanding and grid disturbance.

2.3. Feature extraction

Specific processes of the feature vector extraction are as follows:

1) Decompose the voltage signals to analyzed using wavelet transformation, of which the decomposition layer is $j$. Then reconstruct the reconstruction signal of each layer in phase space. In this paper, the number of sampling points is 6000. 600 dimensional phase space is reconstructed to get $j$ matrices $A_j$ of 5401×600 dimensions.

2) Using SVD for matrix $A_j$ of each layer, thus obtaining 600 singular values of each layer.

3) Entropy calculation is conducted for the singular values of each layer and combine the entropy values to extract the feature vector:

$$T=[h_1,h_2,h_3,\cdots,h_l]$$  \quad (6)$$

As in (6), $h_1$, $h_2$, $h_3,\cdots, h_l$ are the entropy values of each layer. Vector $T$ is the feature vector of islanding and grid disturbance.

Thus it can be seen that the method is simple and feasible. The feature vector extraction method not only can correctly reflect the 2 different situations, but also has the relative stability for the variation of each kind of situation, namely whether how the signal changes under the 2 cases, the strong similarity of the feature vector is conducive to the recognition of deep learning.

3. Deep learning based islanding detection framework

3.1. Deep learning

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics [12]-[13].

Conventional neural network (NN) adopts back-propagation (BP) as one of core of training principles, which may readily fall into a local optimum. This drawback becomes apparent when the NN architecture goes deep because there is a large number of parameters to be optimized in this situation.

Due to the lack of proper training algorithms in early years, people could not harness this powerful model until Hinton proposed his deep learning idea in 2006 [14]. They proposed a following two-step procedure to train the deep networks:

1) Initialization of weights using unsupervised techniques like Boltzmann machine or Auto-
encoders.

2) Fine-tuning of the previously initialized weights using supervised data to provide better classification.

3.2. Stacked auto-encoder

An auto-encoder, also known as auto-associator, is a three-layer neural network that tries to reconstruct the input at the output layer after being passed through an intermediate hidden layer. A sample auto-encoder is illustrated in Figure 1 where it tries to learn a function \( h_{\theta, \phi}(x) \approx x \), and \( W, b \) correspond to the weight matrix and bias of the input respectively. The aim of the auto-encoder is to learn a latent or compressed representation of the input, by minimizing the reconstruction error between the input and the reconstructed one from the learned representation.

Let \( N_i \) and \( N_h \) denote, respectively, the number of input and hidden units in an auto-encoder. \( X = \{ x_i \in \mathbb{R}^{N_i} \}_{i=1}^M \) is a given set of training samples from \( M \) subjects, from which an auto-encoder maps \( x_i \) to a hidden representation \( y_i \in \mathbb{R}^{N_h} \) through a linear determination mapping and a non-linear activation function \( f \) as follows:

\[
y_i = f(W_i x_i + b_i)
\]

As in (7), \( W_i \in \mathbb{R}^{N_i \times N_h} \) denotes an encoding weight matrix and \( b_i \in \mathbb{R}^{N_h} \) is a bias vector. In this paper, we consider to take a logistic sigmoid function as the activation function which is the most widely used in the field of pattern recognition or machine learning [10]:

\[
f(a) = 1/(1+\exp(-a))
\]

The representation \( y_i \) of the hidden layer is then mapped to a vector \( \hat{x}_i \in \mathbb{R}^{N_i} \), which approximately reconstructs the input vector \( x_i \) by another linear mapping as follows:

\[
\hat{x}_i = W_2 y_i + b_2 \approx x_i
\]

where \( W_2 \in \mathbb{R}^{N_h \times N_i} \) and \( b_2 \in \mathbb{R}^{N_i} \) denote a decoding weight matrix and a bias vector, respectively.

From a learning perspective, we aim to minimize the reconstruction error between the input \( x_i \) and the output \( \hat{x}_i \) with respect to the parameters. Let \( \hat{X} = \{ \hat{x}_i \}_{i=1}^M \) and \( l(X, \hat{X}) \) denote a reconstruction error.

\[
l(X, \hat{X}) = \frac{1}{2} \sum_{i=1}^{M} \| x_i - \hat{x}_i \|^2
\]

In order for the sparseness of the hidden units, we further consider a Kullback-Leibler (KL) divergence between the average activation \( \hat{\rho}_j \) of the \( j \)th hidden unit over the training samples and the target average activation \( \rho \) defined as follows:

\[
\text{KL}(\rho \| \hat{\rho}) = \rho \log \frac{\rho}{\hat{\rho}} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}}
\]

where \( \hat{\rho}_j \) and \( \rho \) are Bernoulli random variables. Then the objective function can be written as follows:

\[
\min l(X, \hat{X}) + \gamma \sum_{j=1}^{N_h} \text{KL}(\rho \| \hat{\rho})
\]

With the introduction of the KL divergence weighted by a sparsity control parameter \( \gamma \) to the target objective function, we penalize a large average activation of a hidden unit over the training samples by setting \( \rho \) small [7]. This penalization drives many of the hidden units’ activation to be equal or close to zero, resulting in sparse connections between layers.

3.3. Supervised fine-tuning process

All of the parameters in deep neural network (DNN) are appropriately initialized based on layer-wise pre-training method. These parameters are required to be slightly adjusted in a supervised manner
until the loss function of DNN reaches its minimum. In this paper, BP is adopted for this type of task due to its effectiveness and efficiency. During the fine-tuning process, BP periodically works in a top-down manner. One period means that all of the parameters are updated one time, resulting in smaller classification errors. The errors are then back-propagated through the training set to re-correct the parameters of the DNN towards their optimal states. Therefore, after certain BP periods, the optimal states of all of the parameters can be found, and thus, the training process of DNN has been completed.

3.4. **Deep learning based classification of islanding and grid disturbance**

This paper focuses on binary classification of islanding and grid disturbance. Let the training set be \( \{ (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \} \), where \( y_i = 1 \) denotes grid disturbance and \( y_i = -1 \) denotes islanding. The specific application procedure is shown as figure 2:

1) Take two types of samples, from which the islanding situation include the voltage drop and rise after the circuit breaker is disconnected, and the grid disturbance include the voltage drop and rise. The above samples are processed uniformly by applying the multi-resolution singular spectrum entropy. And the feature vector can be extracted according to the method of Section 1.3.
2) Construct the DNN framework based on practical problems. Deep neural network weights are initialized using weights obtained from training a stacked auto-encoder.
3) Training. Taking the eigenvectors of the 2 types of sample as input, which are extracted by the method based on multi-resolution singular spectrum entropy, and the recognition model is obtained by pre-training and fine-tuning.
4) After training, using the identified model to test the testing samples. The value of the equation (7) can be determined by the value of equation (12): the function value \( f(x) \) is greater than 0 means that it
is an islanding condition; \( f(x) \) is less than 0 indicates the disturbance of the power grid. Thus the islanding can be separated from the disturbance.

4. Simulation analysis and classification results

4.1. PV system model

Using Matlab/Simulink to simulate 3 kW grid-connected PV single-phase power generation system, the effective voltage of the grid is 220 V. The inverter output is fed to the load and the grid via a filter inductor, and the load is tied with RLC in parallel. The PV system model is shown in Figure 3.

Select the 2 classes (4 types) of samples: in the case of islanding the voltage of the common point rises and falls; non-island grid disturbance caused by the common point voltage rises and falls. In the case of islanding, take the common point voltage rises and falls as island samples. The grid disturbance signal is based on the formula (without noise).

(1) Voltage swells.

Where \( a = 0.1 \sim 0.9 \), indicates the voltage swell amplitude. \( t_1 \) indicates the beginning time of voltage swell and \( t_2 \) is the ending time.

(2) Voltage sags.

Where \( a = 0.1 \sim 0.9 \), indicates the voltage sag amplitude. \( t_1 \) indicates the beginning time of voltage sag and \( t_2 \) is the ending time.

In this paper, the opening time of the grid connected circuit breaker is 0.08s, and the disturbance time of the power grid is 0.08\sim 0.14 s, \( t_1 = 0.08 \) s, \( t_2 = 0.14 \) s, which lasted for a total of three cycles.

4.2. Wavelet Selection and DNN Parameters Setting

In order to divide the above four kinds of signals meticulously, and not make the input vector dimension of the classifier too large, the signal is decomposed by 6 layers, and the \( \text{Db4} \) wavelet [8] is used to analyze the time domain signal.

4.3. Simulation results analysis

![Figure 3 DG system](image)

![Figure 4 Time-domain waveforms of four cases](image)

![Figure 5 Frequency-domain waveforms of four cases](image)
Simulation sampling frequency selected 30000 Hz, sampling time of 0.2 s, a total of 6000 sampling points.

The time domain waveform of common point voltage rise and fall in the four cases of islanding and the non-islanding grid disturbance are shown in Figure 4.

Meanwhile, the frequency domain waveforms of point of common coupling voltage rise and fall in the four cases of islanding and the non-islanding grid disturbance under different scales of transform coefficients are shown in Figure 5.

Analysis of the reconstructed time domain signals frequency domain signals in Figure 4 and Figure 5 can lead to the conclusions:

1) The voltage rises (or descents) at the PCC in the case of islanding cannot be distinguished from the point of common coupling voltage swells (or sags) caused by the grid disturbance by the amplitude threshold in the time domain.

2) Although there are some differences between the voltage rises (or descents) at the PCC in the case of islanding and the point of common coupling voltage swells (or sags) caused by the grid disturbance in the frequency domain, the two cases cannot be distinguished by comparing the reconstruction values, since both have the possibility of reaching the same value. So it is necessary to extract the eigenvector of the multi-resolution singular spectral entropy of the frequency domain signal.

DNN test results are shown in Table 1. The number of the training set is 1000, in which the four cases of islanding and the non-islanding grid disturbance are of 250 samples respectively.

| Samples       | Voltage at PCC | Sample size | Number of misclassification | Accuracy (%) |
|---------------|----------------|-------------|------------------------------|--------------|
| Islanding     | Rises          | 120         | 2                            | 98.3         |
|               | Drops          | 120         | 2                            | 98.3         |
| Disturbance   | Swells         | 120         | 1                            | 99.2         |
|               | Sags           | 120         | 4                            | 96.7         |

Compared with decision tree (DT) and support vector machine (SVM), the method proposed behaves better performance in both accuracy and detection time. The comparison results are shown in Table 2. The results indicate that the deep learning based method proposed is much better than SVM.

| Method       | Accuracy (%) | Detection time (s) |
|--------------|--------------|--------------------|
| DT           | 92.7         | 0.32               |
| SVM          | 95.6         | 0.26               |
| Proposed     | 98.3         | 0.18               |

5. Conclusions
In this paper, multi-resolution singular spectrum entropy is combined with deep neutral network and applied to the classification of islanding and grid interference. Firstly, the signals of the four kinds of cases are transformed by wavelet transform. Then, the phase space reconstruction matrix of each reconstruction factor is decomposed by singular value. The entropy of each layer is further calculated, from which the eigenvectors are structured as input to the DNN framework.

(1) The deep learning based method proposed in this paper can distinguish the islanding and the grid disturbance accurately and rapidly, of which the accuracy can reach 98.3% and the detection time is 0.18s, so that the safe operation of microgrid with DGs can be utterly ensured.

(2) Compared with other classification method, the method proposed behaves better performance in both accuracy and detection time.

References
[1] Trujillo, C. L., et al. 2010 Analysis of active islanding detection methods for grid-connected microinverters for renewable energy processing. Applied Energy 87.11:3591-3605.

[2] Han, Hua, et al. 2016 Review of power sharing control strategies for islanding operation of AC microgrids. IEEE Transactions on Smart Grid 7.1:200-215.
[3] Fares, Robert L., and M. E. Webber. 2015 Combining a dynamic battery model with high-resolution smart grid data to assess microgrid islanding lifetime. Applied Energy 137:482–489.

[4] Pigazo, A., et al. 2009 Wavelet-based islanding detection in grid-connected PV systems. Industrial Electronics IEEE Transactions on 56.11:4445-4455.

[5] Liu, Yung Hsiang, et al. 2008 Method and system for detecting stand-alone operation of a distributed generating system. US, US 7342758 B2.

[6] Gupta, P., R. S. Bhatia, and D. K. Jain 2015 Average absolute frequency deviation value based active islanding detection technique. IEEE Transactions on Smart Grid 6.6:26-35.

[7] Pigazo, A., et al. 2009 Wavelet-based islanding detection in grid-connected PV systems. Industrial Electronics IEEE Transactions on 56.11:4445-4455.

[8] Zhu, Yanwei, Xinchun Shi, and Peng Li 2011 Classification of islanding and grid disturbance based on multi-resolution singular spectrum entropy and svm. Proceedings of the Csee, 31(7), 64-70.

[9] Walden, and T. Andrew 2000 Wavelet methods for time series analysis. Cambridge University Press.

[10] Gomes, Jonas, and L. Velho 2015 The wavelet transform from fourier analysis to wavelets. Springer International Publishing:61-73.

[11] Ping, X., Liu, B., Lin, H. B., and Xiao, W 2004 Study on multi-resolution singular-spectrum entropy and its usage in vibration signal monitoring. Journal of Translucution Technology:949-950.

[12] Hecht-Nielsen 1989 Theory of the backpropagation neural network. International Joint Conference on Neural Networks IEEE:593-605 vol.1.

[13] Wang, Qinglong, Guo Wenbo, et al. 2016 Learning adversary-resistant deep neural networks.

[14] Hinton, G. E., and R. R. Salakhutdinov 2006 Reducing the dimensionality of data with neural networks. Science 313.5786:504-7.