Applied Energy 210 (2018) 776–785

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Deep learning hybrid method for islanding detection in distributed generation

Xiangrui Kong, Xiaoyuan Xu, Zheng Yan*, Sijie Chen, Huoming Yang, Dong Han

School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

HIGHLIGHTS

- For the first time, deep learning is designed for islanding detection.
- The noises can be effectively eliminated by the feature extraction method proposed.
- The competitive performance of the islanding detection method were proved.

ARTICLE INFO

Keywords:
Distributed energy
Microgrid
Islanding
Deep learning
Multi-resolution singular spectrum entropy

ABSTRACT

The increasing penetration of distributed energy brings significant uncertainty and noises to microgrid operation, which enlarge the difficulty of microgrid monitoring. For as much as the detection of islanding is prone to be interfered by grid disturbance, island detection device may make misjudgment thus causing the consequence of distributed generations (DGs) 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 the 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, a deep learning based algorithm is proposed to classify islanding and grid disturbance. Simulation results indicate that the method can achieve its goal while being highly accurate, so the DGs mistakenly withdrawing from power grids can be avoided.

1. Introduction

1.1. Motivations

In a microgrid with increasing distributed energy, the islanding condition resulted from line fault and other reasons might cause serious hazards. Islanding conditions must be detected in order for protection devices to function. Minor disturbances, such as minor voltage or frequency changes, are sometimes misidentified as islanding or fault conditions and trigger protection devices to malfunction [1–3]. This malfunction is harmful in that it severely disrupts the normal operation of a distribution system. Hence it is vital to precisely detect islanding and fault scenarios and distinguish them from minor disturbances.

With the increasing penetration of distributed generations (DGs), however, islanding detection faces unprecedented challenges. On the one hand, renewable generation brings significant uncertainties and noises to a distribution system [4,5]. This increases the intensity of disturbances and creates a barrier to the accuracy of islanding detection. On the other hand, if minor disturbances are misidentified as islanding conditions, DGs would be forced out of service, which can be extremely dreadful. Hence an accurate islanding detection method becomes decisive to the security of a distribution system with high penetration of DGs.

Islanding detection methods can be divided into communication methods, active methods, and passive methods [6,7]. 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

* Corresponding author.
E-mail addresses: xr_kong@sjtu.edu.cn (X. Kong), xuxiaoyuan@sjtu.edu.cn (X. Xu), yanz@sjtu.edu.cn (Z. Yan), sijie.chen@sjtu.edu.cn (S. Chen), yanghuoming@sjtu.edu.cn (H. Yang), D.Han1984@hotmail.com (D. Han).

http://dx.doi.org/10.1016/j.apenergy.2017.08.014
Received 7 April 2017; Received in revised form 31 July 2017; Accepted 6 August 2017
Available online 10 August 2017
0306-2619/

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.

The voltage value and frequency at the point of common coupling 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 in [7,8]. 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 do not necessarily exceed the threshold value at the same time. With the same flaws, an average absolute frequency deviation value based active islanding detection technique is utilized in [9]. In [10], wavelet energy coefficients in different frequency bands of the transient phase current signal are extracted as eigenvectors by wavelet transformation, with which the islanding is detected by means of decision tree, neural network, support vector machine and other pattern recognition technology. This method is a novel and quick way to recognize islanding. However, using wavelet energy coefficients as eigenvectors is susceptible to noise. Moreover, contrastive analysis and classification of islanding and similar disturbance are not conducted in that paper.

This paper aims to propose a method that can detect islanding and grid disturbances accurately with high penetration of distributed energy.

1.2. Contributions

The contributions of this paper include the following:

(1) We propose a novel feature extraction method for islanding detection. The method combines multi-resolution singular spectrum entropy with wavelet decomposition. Multi-resolution singular spectrum entropy is good at digging the essential feature of signals from islanding and disturbance conditions irrespective of how wavelet coefficients are set. It is found to be effective in eliminating noise interference. The method is proved to outperform the popular wavelet energy coefficient method in feature extraction.
(2) For the first time, a deep-learning-based islanding detection method is proposed for binary-classification of islanding and disturbance conditions in a microgrid with high penetration of distributed energy. Deep learning uses computational models composed of multiple processing layers to learn representations of data with multiple levels of abstraction. The deep learning method proposed is found to dramatically improve the accuracy of islanding detection.

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 the discrete signal can be determined by Eq. (1) [11–15].

$$\begin{cases} c_{j+1}(k) = Hc_j(k) \\ d_{j+1}(k) = Gc_j(k) \end{cases} \tag{1}$$

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 the 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 Eq. (2).

$$c_{j-1} = H^*c_j + G^*d_j \tag{2}$$

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

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

The assumption to reconstruct an 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, an $(N - n + 1) \times n$ dimensional matrix $\boldsymbol{A}$ is constructed.

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

(4) For the matrix $\boldsymbol{A}$ of each layer, singular value decomposition (SVD) is conducted to calculate the singular spectrum entropy.

The matrix $\boldsymbol{A}_{(N-n+1)\times n}$ is decomposed using SVD and the result is $\boldsymbol{A} = \boldsymbol{U}_{(N-n+1)}\boldsymbol{\Lambda}_{l\times l}\boldsymbol{V}^{\mathrm{T}}{}_{n\times l}$. The nonzero diagonal elements $\lambda_{ji}(i = 1, 2, 3, \ldots, l)(l=)$ from $\boldsymbol{\Lambda}_{l\times l}$ are singular values of the matrix $\boldsymbol{A}$ from layer j. According to the informational entropy theory, definition of the signal singular spectrum entropy is as follows.

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

$$p_{ji} = \frac{\lambda_{ji}}{\sum_{i=1}^{l}\lambda_{ji}} \tag{5}$$

As in Eqs. (4) and (5), H_j is the information entropy of level j; the p_{ji} is the uncertain probability distribution of λ.

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 approaches [16,17]. The wavelet reconstruction coefficients matrix uniquely represents the information of the corresponding layer of the signal, and the singular value vector of the corresponding layer uniquely represents the coefficients matrix as well [18]. Therefore, the singular value vector of the corresponding layer uniquely represents the information of the corresponding layer as well. Moreover, 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 eigenvector extraction are as follows:

(1) Decompose the voltage signals to analyze 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 matrixes $\boldsymbol{A}_j$ of 5401 × 600 dimensions.
(2) Using SVD for matrix $\boldsymbol{A}_j$ of each layer, thus obtaining 600 singular values of each layer.
(3) Entropy calculation is conducted for the singular values of each layer and combines the entropy values to extract the eigenvector:

$$\boldsymbol{T} = [h_1,h_2,h_3,...,h_j] \tag{6}$$

As in Eq. (6), $h_1, h_2, h_3, \ldots, h_j$ are the entropy values of each layer. Vector $\boldsymbol{T}$ is the eigenvector of islanding and grid disturbance.

Thus it can be seen that the method is simple and feasible. More meticulous frequency bands division of signals is obtained by wavelet decomposition. Its information is quantitatively described in the form of information entropy, and the complexity of the original signal is given a certain measure. The eigenvector extraction method not only can correctly reflect the two different situations, but also has the relative stability for the variation of each kind of situation, namely whether how the signal changes under the two cases, the strong similarity of the eigenvector is conducive to the recognition of deep learning.

3. Deep learning based islanding detection framework

3.1. Deep learning

Deep learning (DL) enables computational models that consist 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 object detection, visual object recognition and many other domains such as medical and genomics science.

As early as 1989, the universal expressive power of three-layer nets was proved [19]. The proof showed surprisingly that given sufficient number of hidden units and proper nonlinearities in activation function and weights, any continuous function from input to output can be implemented in a three-layer net.

Conventional neural network (NN) adopts back-propagation (BP) as one of the core of training principles, which may readily fall into a local optimum [20,21]. This drawback becomes apparent when the NN architecture goes deep because there are 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 [22]. A two-step procedure is proposed to train the deep architectures as follows:

(1) Initialization of weights using greedy layer-wise unsupervised learning algorithm like auto-encoders.
(2) Using supervised data, fine-tuning of the previously initialized weights to provide better classification.

Regarding deep neural network (DNN), the training process consists of a layer-wise pre-training process and a fine-tuning process. The former is used to provide good initial values for all parameters based on stacked auto-encoder, while the latter is used to search the optimum based on the given initial states of the network. The rest of this section discusses stacked auto-encoders, supervised fine-tuning and softmax regression to form DNN framework.

3.2. Auto-encoder

The auto-encoder is a three-layer neural network that tries to reconstruct the input at the output layer after being passed through a hidden layer [23,24]. A sample auto-encoder is illustrated in Fig. 1 where it tries to learn a function $h_{\boldsymbol{W},\boldsymbol{b}}(x) \approx x$, and $\boldsymbol{W}$, $\boldsymbol{b}$ correspond to the weight matrix and bias of the input respectively. An auto-encoder minimize the reconstruction error between the input and the reconstructed one from the learned representation, aiming to learn a latent representation of the input.

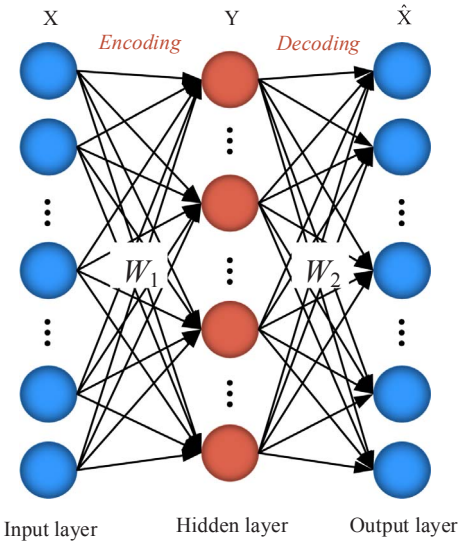


Fig. 1. Illustration of the auto-encoder.

Let N_I and N_H denote, respectively, the number of input and hidden units in an auto-encoder. $\boldsymbol{X} = \{x_i \in \Re^{N_\mathrm{I}}\}_{i=1}^{M}$ is a given set of training samples from M subjects, from which an auto-encoder maps $\boldsymbol{x}_i$ to a hidden representation $\boldsymbol{y}_i \in \Re^{N_\mathrm{H}}$ through a linear determination mapping and a non-linear activation function f as follows:

$$\boldsymbol{y}_i = f(\boldsymbol{W}_1\boldsymbol{x}_i + \boldsymbol{b}_1) \tag{7}$$

As in Eq. (7), $\boldsymbol{W}_1 \in \Re^{N_\mathrm{H}\times N_\mathrm{I}}$ denotes an encoding weight matrix and $\boldsymbol{b}_1 \in \Re^{N_\mathrm{H}}$ is a bias vector. In this paper, a logistic sigmoid function is considered as the activation function, which is the most widely used in pattern recognition and machine learning [23,24]:

$$f(a) = 1/(1 + \exp(-a)) \tag{8}$$

The representation $\boldsymbol{y}_i$ of the hidden layer is then mapped to a vector $\hat{\boldsymbol{x}}_1 \in \Re^{N_\mathrm{I}}$, which approximately reconstructs the input vector $\boldsymbol{x}_i$ by another linear mapping. As an unsupervised learning algorithm, auto-encoders use the back-propagation mechanism to make the target output value approach the input value. Since the encoder tries to learn a constant function:

$$\hat{\boldsymbol{x}}_i = \boldsymbol{W}_2\boldsymbol{x}_i + \boldsymbol{b}_2 \approx \boldsymbol{x}_i \tag{9}$$

where $\boldsymbol{W}_2 \in \Re^{N_\mathrm{I}\times N_\mathrm{H}}$ and $\boldsymbol{b}_2 \in \Re^{N_\mathrm{I}}$ denote a decoding weight matrix and a bias vector, respectively.

For a learning perspective, the reconstruction error is supposed to be minimized between the input $\boldsymbol{x}_i$ and the output $\hat{\boldsymbol{x}}_i$ on the parameters. Let $\hat{\boldsymbol{X}} = \{\hat{\boldsymbol{x}}_i\}_{i=1}^{M}$ and $l(\boldsymbol{X},\hat{\boldsymbol{X}})$ denote a reconstruction error.

$$l(\boldsymbol{X},\hat{\boldsymbol{X}}) = \frac{1}{2}\sum_{i=1}^{M} ||\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i||_2^2 \tag{10}$$

To achieve the sparseness of hidden units, we further introduce a Kullback-Leibler (KL) divergence in equation (11), which is between the average activation $\hat{\rho}_j$ of the jth hidden unit over the training set and the target average activation ρ.

$$\mathrm{KL}(\rho||\hat{\rho}) = \rho\log\frac{\rho}{\hat{\rho}} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}} \tag{11}$$

where $\hat{\rho}_j$ and ρ are Bernoulli random variables. Thus we can draw the objective function as follows:

$$\min\ l(\boldsymbol{X},\hat{\boldsymbol{X}}) + \gamma\sum_{j=1}^{N_\mathrm{H}} \mathrm{KL}(\rho||\hat{\rho}) \tag{12}$$

With the introduction of the Kullback-Leibler divergence weighted

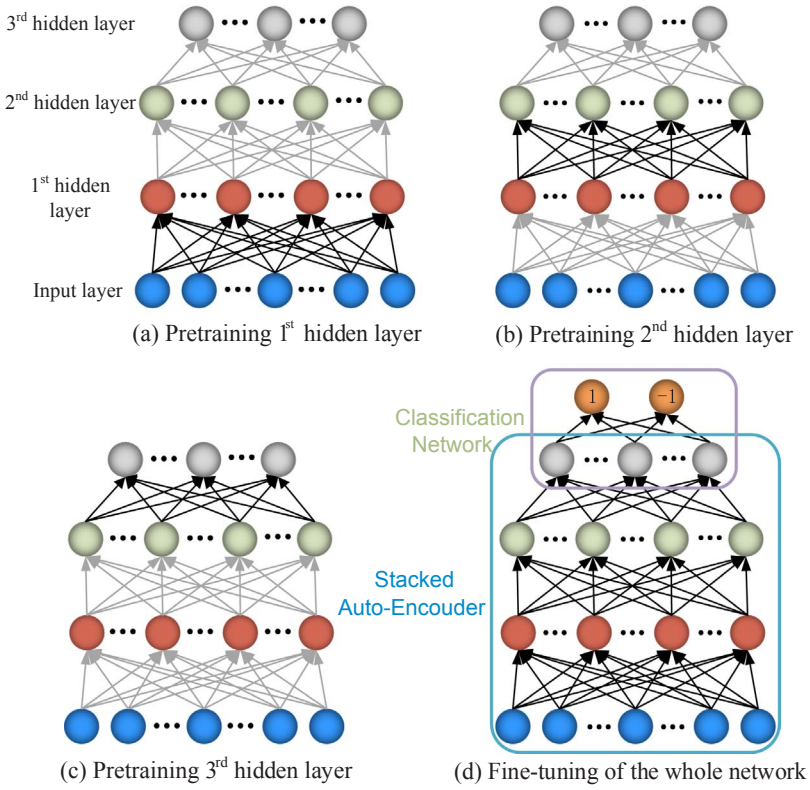


Fig. 2. A deep architecture of our stacked auto-encoder and the two-step (unsupervised greedy layer-wise pretraining and supervised fine-tuning) parameter optimization scheme. (The black arrows denote the parameters to be optimized in the current stage). (a) Pre-training of the first hidden layer with the training samples as inputs, (b) pre-training of the second hidden layer with the outputs from the first hidden layer as inputs, (c) pre-training of the third hidden layer with the output from the second hidden layer as inputs, (d) fine-tuning of the whole network with an additional label-output layer, taking the pre-trained parameters as the starting point in optimization.

by a sparsity control parameter γ to the target objective function, a large average activation of a hidden unit over the training set is penalized by setting ρ small. Many of the hidden units' activation are driven to be equal or close to 0 by the penalization, from which sparse connections between layers can be achieved.

Note that the latent representation of the input vector is determined by the output of the hidden layer. However, the representational power of a single-layer auto-encoder is quite limited for the reason of its simple shallow structural characteristic.

3.3. Stacked auto-encoder

Inspired by the biological model of the human visual cortex, recent studies in machine learning have shown that a hierarchical or deep architecture is useful to discover complicate and highly non-linear patterns in data [25,26]. Motivated by these studies, in this paper we consider a stacked auto-encoder, in which an auto-encoder turns into a building block. Specifically, auto-encoders are stacked one after another taking the outputs from the lower hidden layer as the input to the upper layer's input units, and so on. A SAE model example with three auto-encoders stacked hierarchically is shown in Fig. 2. Note that in the lower input layer the number of units is equal to the dimension of the input eigenvector. But in the upper layers the number of hidden units could be determined by the nature of the input, which means it might be even larger than the input dimension.

According to the nature of the hierarchical architecture, one of the most significant characteristics of the SAE is to find or learn complex and highly non-linear patterns such as the latent relations between input features. Another significant characteristic of the deep learning is that the latent representation can be learned directly from the data. With its self-taught and representational learning properties, a latent representation of the original voltage features can be found directly extracted from the input signal data. When an input sample is presented to a SAE model, the different layers of the network represent different levels of information. That is, the lower the layer in the network, the more linear relations of features; the higher the layer, the more complex and more non-linear relations of features inherent in the input eigenvector.

With regard to training parameters in the deep architecture of the SAE model, a straightforward method is to apply BP method with the gradient-based optimization technique starting from random initialization. However, it is generally known that deep neural networks trained in that manner perform worse than networks with a shallow structure, suffering from gradient dispersion and falling into a poor local optimum [24,27]. However, in 2009, Hinton introduced an unsupervised greedy layer-wise learning algorithm and showed its success to learn a deep belief network [22]. The main idea of the greedy layer-wise learning algorithm is to train one layer at a time by maximizing the variational lower bound. That is, the first hidden layer is trained firstly with the training data as input, and then the second hidden layer is trained with the outputs from the first hidden layer as input, and so on.

3.4. Supervised fine-tuning process

All of the parameters in 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. More details on the BP algorithm in DNN can be found in [27–29]. 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. The deep learning framework is shown in Fig. 2.

3.5. Softmax classification

After the training, the softmax regression is introduced to classify

the test set. In the softmax regression, the training set consists of $\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_n,y_n)\}$, where $\hat{\boldsymbol{x}}_1 \in \Re^{N_1}$. The hypothesis function is shown in Eq. (13).

$$h_\theta(x) = \frac{1}{1+\exp(-\theta^T x)} \tag{13}$$

where θ is the trained parameters that minimizes the loss function in Eq. (12).

For the given test input x, we use the hypothesis function to estimate the probability value $p(y=j|x)$ of each class j. For k-classes test set, the hypothesis function will output a k-dimensional vector to represent the probability of this k-classes test set. Specially, the hypothesis function is as follows:

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)}=1)|x^{(i)};\theta \\ p(y^{(i)}=2)|x^{(i)};\theta \\ \vdots \\ p(y^{(i)}=k)|x^{(i)};\theta \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \tag{14}$$

Notice that $\frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}}$ normalizes the probability distribution so that the sum of all probabilities is 1.

As for the binary classification of islanding and grid disturbance in this paper, the softmax regression degenerate into the logistic regression. When $k=2$, the hypothesis function of softmax regression is as follows:

$$h_\theta(x^{(i)}) = = \frac{1}{e^{\theta_1^T x^{(i)}} + e^{\theta_2^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \end{bmatrix} \tag{15}$$

3.6. Deep learning based classification algorithm of islanding and grid disturbance

This paper focuses on the binary classification of islanding and grid disturbance. Let the training set be $\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\ldots,(\boldsymbol{x}_n,y_n)\}$, where $y_i = 1$ denotes grid disturbance and $y_i = -1$ denotes islanding. The specific application procedure is shown in Fig. 3.

(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 eigenvector can be extracted according to the method of Section 2.3.
(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. Take the eigenvectors of the 2 types of the sample as input, 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 label value y_i can be determined by the value of Eq. (15): the function value $h(x)$ is greater than 0.5 means that it is an islanding condition; $h(x)$ is less than 0.5 indicates the disturbance of the power grid. Thus the islanding can be separated from the disturbance.

In this paper, the required samples are obtained by function and simulation. Eigenvectors of these samples are extracted by Matlab, which are taken as input of the DNN classifier framework for training and testing.

4. Simulation analysis and classification results

4.1. DG 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 Fig. 4.

Select the 2 classes (4 types) of samples: in the case of islanding the voltage of the point of common coupling rises and falls; non-island grid disturbance caused by voltage rises and falls at the PCC. In the case of islanding, take voltage rises and falls at the PCC as island samples. The grid disturbance signal is based on the formula [30] (without noise).

(1) Voltage swells.

$$u(t) = [1 + a(u(t_2) - u(t_1))]\sin(\omega t) \tag{16}$$

where $a = 0.1$–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.

$$u(t) = [1 - b(u(t_2) - u(t_1))]\sin(\omega t) \tag{17}$$

where $b = 0.1$–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 0.08 s, and the disturbance time of the power grid is 0.08–0.14 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 Db4 wavelet [11] is used to analyze the time domain signal.

4.3. Simulation results analysis

4.3.1. Time domain and frequency domain waveforms

Simulation sampling frequency selected 30,000 Hz, sampling time of 0.2 s, and a total of 6000 sampling points.

The time domain waveforms of voltage rise and fall at PCC in the four cases of islanding and the non-islanding grid disturbance are shown in Fig. 5.

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 Fig. 6.

Analysis of the reconstructed time domain signals frequency domain signals in Figs. 5 and 6 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.

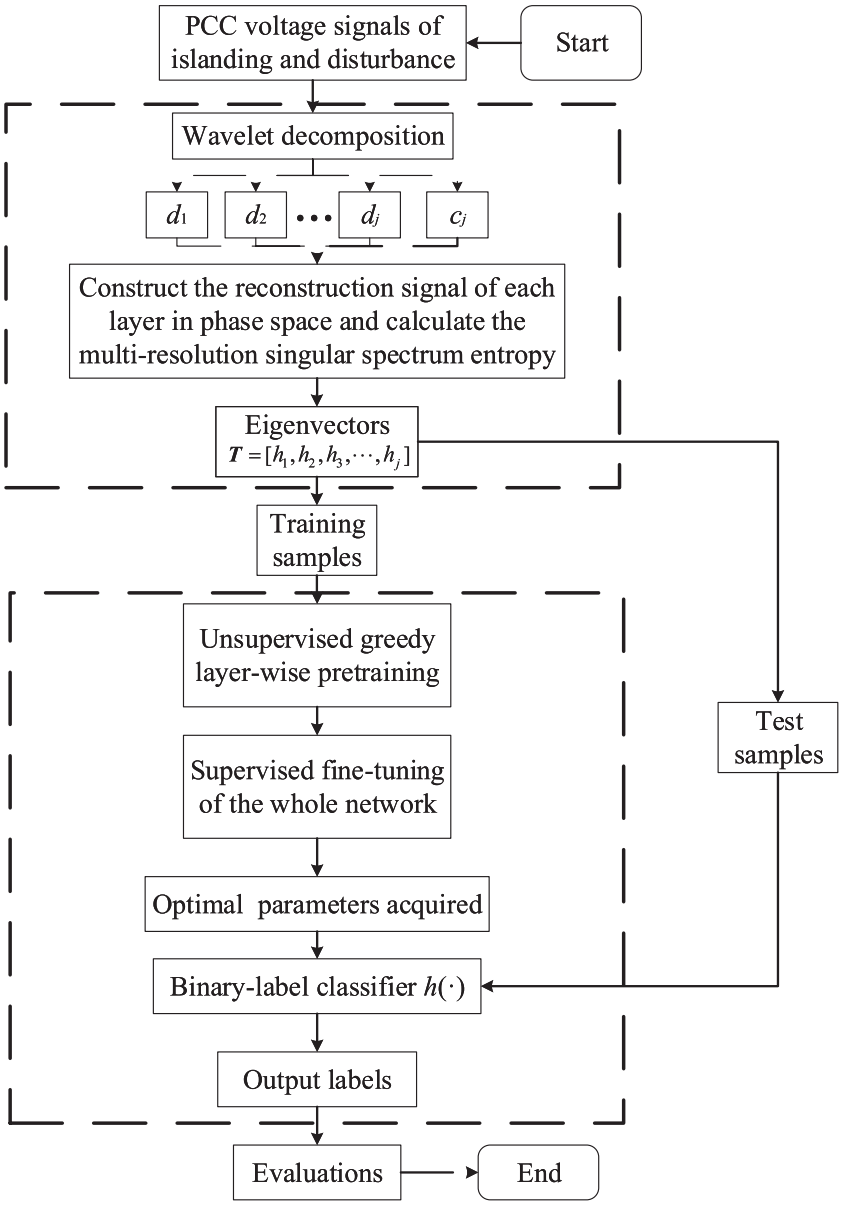


Fig. 3. Method flow chart with multi-resolution singular spectrum entropy and DNN.

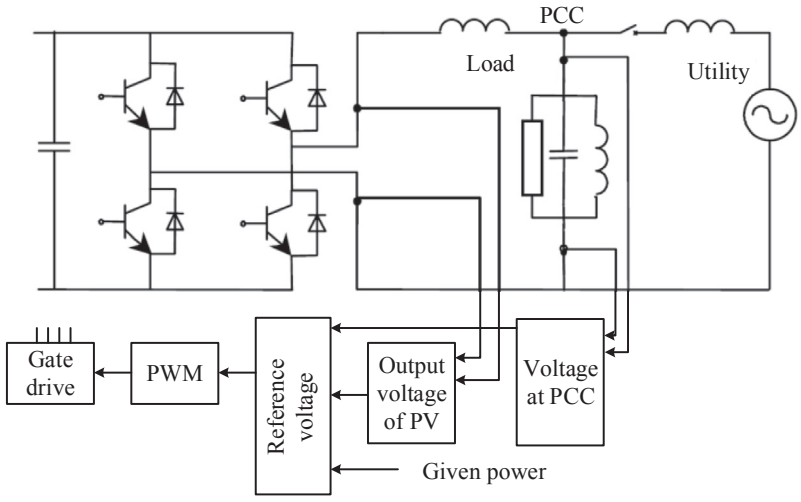


Fig. 4. DG system.

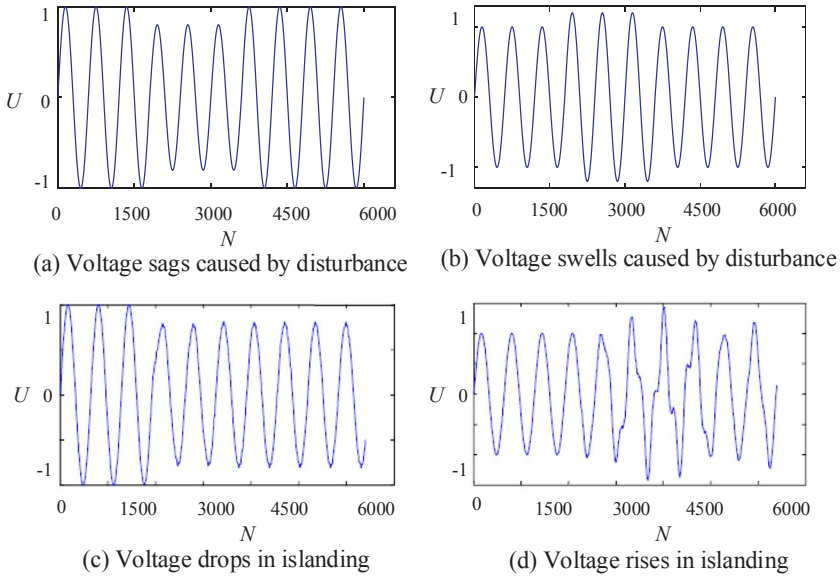


Fig. 5. Time-domain waveforms of four cases.

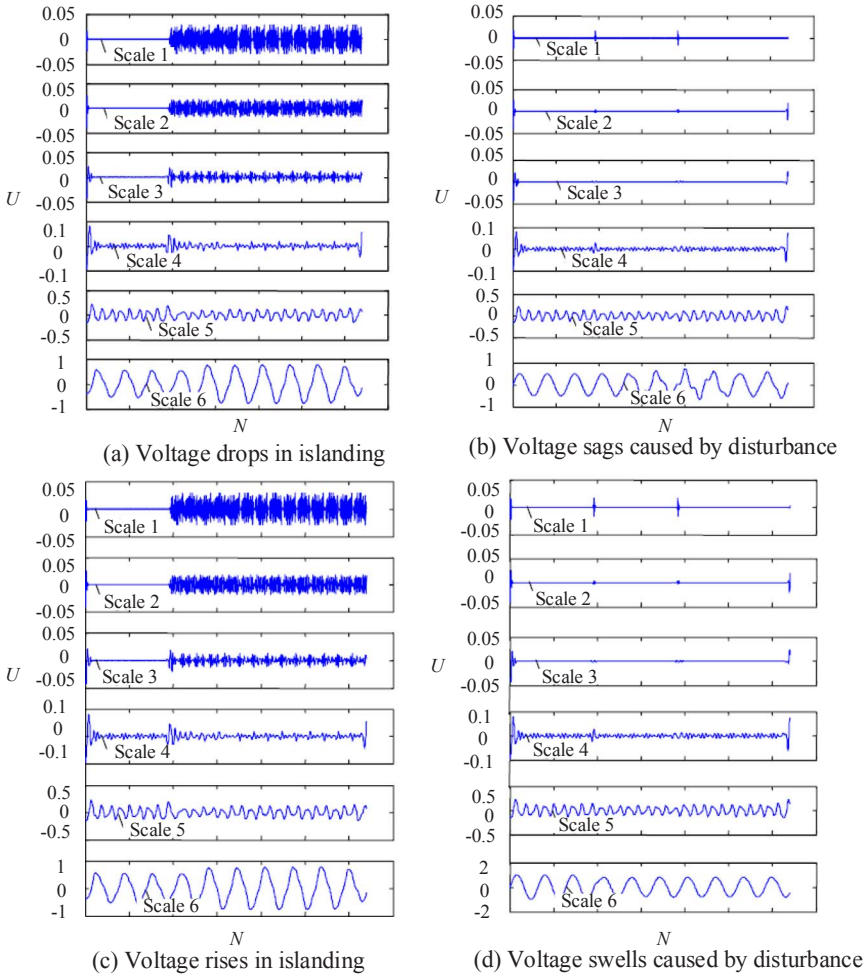


Fig. 6. Frequency-domain waveforms of four cases.

4.3.2. Eigenvector and recognition

The entropy values of the four cases in the different frequency bands (frequency from high to low) are shown in Table 1, where only three of each sample are listed.

Table 1 analyses are as follows:

(1) The comparison is conducted between the islanding and the non-islanding. Comparing the entropy of the voltage drop at the PCC of the islanding with the entropy of the voltage sag at the point of common coupling voltage caused by grid disturbance, the difference is obvious and the entropy of each layer is different. That is, the eigenvectors of the two are different and can be separated. The

Table 1
Entropy values of 4 cases.

Voltage change rate at PCC (%)		h_1	h_2	h_3	h_4	h_5	h_6
Islanding: voltage rises	20	2.81842	3.09455	3.56171	3.82242	3.97955	2.58613
	15	2.70281	2.85924	3.38809	3.73992	3.90576	2.54287
	10	2.73982	2.91289	3.44488	3.76912	3.96499	2.58765
Islanding: voltage drops	20	2.70221	2.82298	3.17797	3.65695	3.84239	2.53437
	15	2.65564	2.79567	3.34416	3.67881	3.84078	2.47399
	10	2.65256	2.79663	3.34664	3.67516	3.83880	2.47198
Disturbance: voltage swells	20	2.63001	2.63366	2.71456	2.80255	2.42800	2.43420
	15	2.61804	2.65210	2.82069	2.87340	2.62099	2.43305
	10	2.62278	2.65373	2.78965	2.86227	2.47567	2.43280
Disturbance: voltage sags	20	2.63135	2.67032	2.83591	2.89681	2.63022	2.43631
	15	2.62818	2.67001	2.83967	2.88601	2.81368	2.43634
	10	2.63123	2.66603	2.82577	2.89714	2.54853	2.43658

Table 2
Classification results of DNN.

Samples	Voltage at PCC	Number of samples	Number of misclassification	Classification rate (%)
Islanding	Rises	120	2	98.3
	Drops	120	2	98.3
Disturbance	Swells	120	1	99.2
	Sags	120	4	96.7

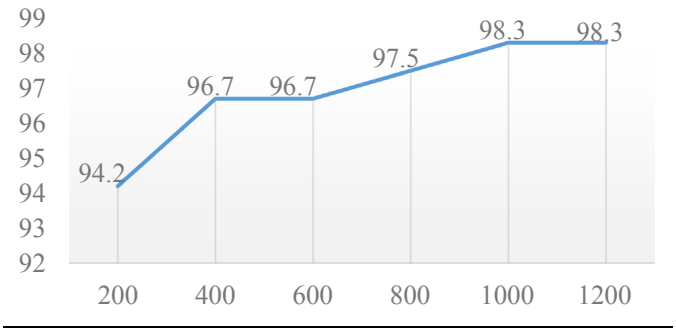
entropy of the voltage rise at the PCC of the island and the entropy of the point of common coupling voltage swell also have the same characteristics.

(2) Comparison of the same kind of samples. Although the state is different, the entropy numerical values are relatively stable and the change is relatively small.

(3) Comparison between similar samples, that is, the comparison of voltage rises and falls at the PCC in the case of islanding, the comparison of voltage swells and sags in the case of grid disturbance. When the voltage at PCC in the case of islanding rises and falls, the entropy and the change trend of the two kinds of samples are very similar. The voltage swell and sag in the case of grid disturbance are also the same, which is, the entropy is also relatively stable for the different signals of the same kind of samples.

From the above analysis we can see that the eigenvectors based on multi-resolution singular spectral entropy are not only stable for the same kind of signals, but also stable for the same kind of sample, which is used to facilitate the recognition and classification as the input of DNN and so as to determine the time of the islanding in line with the relevant provisions. DNN test results are shown in Table 2. The number of the training set is 1000.

Table 4
Accuracy under different sample size (%).



4.3.3. The influence of noise

In the engineering test process, due to various factors, the measured signal is inevitably subject to different levels of noise pollution. Therefore it is necessary to add a certain amount of noise into the above model. In this paper, the Gaussian white noise is added to the four types of signals, where the SNR is 30 dB, and the simulation is carried out. The new entropy results are as follows (see Table 3).

After adding noise, the entropy of each layer of the signal is improved, but still maintain the characteristics described in Table 1, which does not affect the final classification results. These results show that multi-resolution singular spectral entropy is not sensitive to noise, and the method proposed has good practical value.

Table 3
The new entropy value with noise.

Voltage change rate at PCC (%)		h_1	h_2	h_3	h_4	h_5	h_6
Islanding: voltage rises	20	2.99389	3.12553	3.70076	3.83159	4.12433	2.71659
	15	2.70715	2.91203	3.58479	3.80803	3.95842	2.55473
	10	2.76918	2.95419	3.49708	3.81346	4.06546	2.68617
Islanding: voltage drops	20	2.88642	2.92432	3.18599	3.81874	3.85744	2.71530
	15	2.79532	2.98652	3.34976	3.76183	3.84663	2.52805
	10	2.85247	2.94027	3.46741	3.87025	3.94645	2.61283
Disturbance: voltage swells	20	2.75847	2.68256	2.85421	2.98072	2.43718	2.57052
	15	2.75382	2.76276	2.89027	2.97406	2.72497	2.46027
	10	2.70626	2.83644	2.80957	3.00468	2.57746	2.62598
Disturbance: voltage sags	20	2.79232	2.79467	2.87664	3.00383	2.68797	2.58745
	15	2.79406	2.67093	2.91040	2.90080	3.00044	2.61120
	10	2.81218	2.71953	2.93551	2.89811	2.63478	2.55405

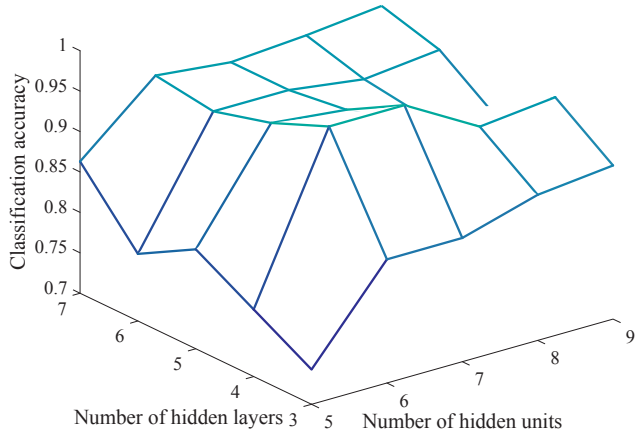


Fig. 7. Accuracy under different number of hidden layers/hidden units (%).

Table 5
Comparison results of different classifiers.

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

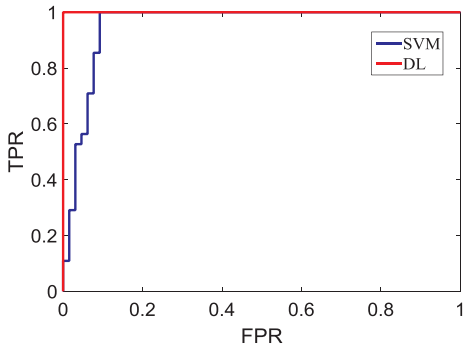


Fig. 8. ROC curve of DL and SVM.

4.3.4. Influence of the number of samples

It is empirically considered that the deep architecture needs relatively massive sample number [31–33]. Thus, the effects of different sample sizes on the results were also examined (see Table 4).

As it shown in the table, classification results of the method proposed are more accurate as the sample size increases. The lowest accuracy is 94.2% when the sample size is 200. The accuracy reaches 98.3% when the sample size is more than 1000.

4.3.5. Influence of the number of hidden layers and hidden units

The effects of different number of hidden layers and hidden units on the results were also examined in this paper. It is also empirically considered that as the number of hidden layers/hidden units increase, the network error will decrease.

In Fig. 7, we can find that the most accurate result is obtained at the 4-layer and 7-unit DNN model. The accuracy is not quite obviously effected when the numbers of hidden layers/hidden units are more than 4-layer and 7-unit.

5. Comparison of different classifiers

Comparing 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 5.

The ROC curve in Fig. 8 also indicates that the deep learning based method proposed is much better than SVM.

6. 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. The voltage at the point of common coupling is selected as the analysis object. 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. The conclusions can be drawn as follows:

(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 0.18 s, so that the safe operation of microgrid with DGs can be utterly ensured.
(2) As the feature extraction tool, multi-resolution singular spectrum entropy is not sensitive to noise, and the method proposed has outstanding practical value in engineering. The entropy value is stable for different samples in each case and is favorable for DNN classification.
(3) The method proposed needs relatively massive sample number. And the classification accuracy increases with bigger sample size. The accuracy reaches 98.3% when the sample size is more than 1000.
(4) Compared with other classification method, the method proposed behaves better performance in both accuracy and detection time.
(5) The next step in the research work will take into account more complex conditions including more disturbance, faults, islanding and apply the method for meticulous microgrid monitoring.

Acknowledgement

This research was supported by National Key R & D Program of China (No. 2017YFB0902803), Science and Technology Project of State Grid Corporation of China (No. 2017GW-01).

References

[1] Trujillo CL, et al. Analysis of active islanding detection methods for grid-connected microinverters for renewable energy processing. Appl Energy 2010;87(11):3591–605.
[2] Han Hua, et al. Review of power sharing control strategies for islanding operation of AC microgrids. IEEE Trans Smart Grid 2016;7.1:200–15.
[3] Fares Robert L, Webber ME. Combining a dynamic battery model with high-resolution smart grid data to assess microgrid islanding lifetime. Appl Energy 2015;137:482–9.
[4] Dugan RC, Mcdermott TE. Distributed generation. IEEE Ind Appl Mag 2002;8(2):19–25.
[5] Hung Duong Quoc, Mithulananthan N, Bansal RC. Analytical strategies for renewable distributed generation integration considering energy loss minimization. Appl Energy 2013;105(2):75–85.
[6] Pepermans G, et al. Distributed generation: definition, benefits and issues ☆. Energy Policy 2005;33(6):787–98.
[7] Pigazo A, et al. Wavelet-based islanding detection in grid-connected PV systems. Ind Electron IEEE Trans 2009;56(11):4445–55.
[8] Liu, Yung Hsiang, et al. Method and system for detecting stand-alone operation of a distributed generating system. US, US 7342758 B2; 2008.
[9] Gupta P, Bhatia RS, Jain DK. Average absolute frequency deviation value based active islanding detection technique. IEEE Trans Smart Grid 2015;6(6):26–35.
[10] Lidula NWA, Perera N, Rajapakse AD. Investigation of a fast islanding detection methodology using transient signals. Power & Energy Society General Meeting, 2009. PES '09. IEEE IEEE Xplore; 2009. p. 1–6.
[11] Mallat SG. A wavelet tour of signal processing: the sparse way. 31.3; 2013: p. 83–5.
[12] Walden, Andrew T. Wavelet methods for time series analysis. Cambridge University Press; 2000.
[13] Gomes Jonas, Velho L. The Wavelet Transform. From Fourier Analysis to Wavelets.

Springer International Publishing; 2015. p. 61–73.
[14] Chernick Michael R. Wavelet methods for time series analysis. Cambridge University Press Cambridge 43; 2016: p. 491–91.
[15] Deo Ravinesh C, Wen X, Qi F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Appl Energy 2016;168:568–93.
[16] Xie Ping, et al. Study on multi-resolution singular-spectrum entropy and its usage in vibration signal monitoring. J Trans Technol 2004: p. 949–50.
[17] Chen Shilong, et al. Distinguish faults located inside/outside protection zone of UHVDC transmission line by multi-resolution singular spectrum entropy and support vector machine. Power System Technology; 2015.
[18] Wang Shengjie, et al. Study on machinery fault diagnosis of high voltage circuit breaker based on multi-resolution singular spectrum entropy and SVM. Electric Power Science & Engineering; 2012.
[19] Hecht-Nielsen. Theory of the backpropagation neural network. International joint conference on neural networks IEEE, vol. 1; 1989. p. 593–605.
[20] Wang HZ, et al. Deep belief network based deterministic and probabilistic wind speed forecasting approach. Appl Energy 2016;182:80–93.
[21] Wang Qinglong, Guo Wenbo, et al. Learning adversary-resistant deep neural networks; 2016.
[22] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504–7.
[23] Sainath Tara N, Kingsbury B, Ramabhadran B. Auto-encoder bottleneck features using deep belief networks. IEEE international conference on acoustics, speech and signal processing IEEE; 2012. p. 4153–56.
[24] Palm Rasmus Berg. Prediction as a candidate for learning deep hierarchical models of data. Technical University of Denmark; 2012.
[25] Hinton Geoffrey, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Sig Process Mag 29.6; 2012: p. 82–97.
[26] Thirukovalluru Raghuveer, et al. Generating feature sets for fault diagnosis using denoising stacked auto-encoder. IEEE international conference on prognostics and health management IEEE; 2016. p. 1–7.
[27] Schmidhuber J. Deep learning in neural networks: an overview. Neural Networks Off J Int Neural Network Soc 2014;61:85–117.
[28] Haykin Simon S. Neural networks and learning machines. 3rd ed. Englewood Cliffs (NJ, USA): Prentice-Hall; 2008.
[29] Wang Qinglong, Guo Wenbo, et al. Random feature nullification for adversary resistant deep architecture; 2016.
[30] García-Gracia Miguel, et al. Modelling wind farms for grid disturbance studies. Renew Energy 33.9; 2008: p. 2109–21.
[31] Hinton Geoffrey, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Sig Process Mag 29.6; 2012: p. 82–7.
[32] Lecun Yann, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.
[33] Guo Wenbo et al. Towards Interrogating Discriminative Machine Learning Models. arXiv preprint arXiv:1705.08564; 2017.