Applications of Stacked Autoencoder Network on Communication Transmitter Individual Feature Extraction

Lei Yingke¹,* and Liao Xichang¹
¹National University of Defense Technology, Hefei, China

*Corresponding author’s e-mail: leiyingke@163.com

Abstract. Aiming at the problem that square integral bispectrum cannot accurately represent communication transmitter individuals with the same type under small sample, a communication transmitter individual deep square integral bispectra feature extraction method based on stacked autoencoder network is proposed. The square integral bispectrum (SIB) features of real communication transmitters signals were extracted firstly from the instantaneous frequency. Then the deep SIB features were represented through the layer-wise semi-supervised learning by a stacked autoencoder network. Experiments on real FM communication transmitter signals showed that the identification accuracy is around 80% and the results of multiple experimental variables setting indicated the robustness of the algorithm proposed.

1. Introduction
Due to the random discreteness of component performance, production process, installation and commissioning, and working process, the radiation source of communication emitter must have different characteristics[1], of which are different from other similar devices. In the civil field, it can be used to supervise the electromagnetic spectrum, find the illegal communication radio in time, and monitor the illegal occupation of the electromagnetic spectrum resources[2]. In 2007, the square integral bispectrum method(SIB) Xu[3] put forward is now widely developed to a series of signals’ steady state methods, which can completely reflect the nonstationarity, nonlinearity and non-Gaussian of signal source. In recent years, many researches have been devoted to reprocessing SIB features, extracting the fine features of low dimensional communication source signals, and making them suitable for communication transmitter individual identification. In 2013, Tao[4] designed the local contour bispectrum, which improved the representative ability of subtle characteristics to transmitter individual. In 2016, Tang[2] used the maximum correlation entropy to represent the individual differences of communication radios on the basis of extracting the SIB characteristics of signals, and achieved robust recognition performance in multiple radios recognition.

However, the above research has not been able to solve an important problem in the practical application, that is, the small sample problem. Small sample problem is that when there is not enough labeled samples, the feature extraction algorithm can not effectively represent the radio individuals, resulting in a significant decrease in the accuracy of individual identification. Label information in labeled samples need to be tagged manually. It is not possible to label enough signals when receiving new signals. Therefore, how to extract features accurately is very important for small sample size. Since Hinton[5] puts forward the encoder algorithm in 2006, the concept of deep learning has been developing rapidly. At present, autoencoder has related researches and applications in signal...
processing, which can explain the potential of its application to SIB feature processing. The continuous autoencoder network designed by Hu[6] in 2010 can reduce the dimension of high-dimensional image signals, which shows that the autoencoder has the ability to deal with high-dimensional features. In 2014, a method based on sparse neural network proposed by Wang[7] can accurately distinguish the forest fire images with similar patterns, showing that the autoencoder feature has good individual discrimination. In 2015, Ma[8] extracted feature information of vocal signals of different speakers with sparse autoencoder, proving that the encoder has a good ability to segment different speakers, and the great representative ability of speech signal. In 2017, Zhang[9] realized the single sample face recognition by autoencoder. It provides a new idea for solving the small sample problem of communication transmitter individual identification, that is, using a large number of unlabeled samples for unsupervised learning, and also obtaining accurate category label information. On the basis of the above analysis, in this paper, a stack autoencoder network method for communication transmitter individual SIB feature extraction is proposed. Based on the SIB feature, the model is firstly trained by labeled samples, and the training result is used for unsupervised learning and finally the deep SIB feature (DSIB) is extracted. The experimental results show that the DSIB can effectively distinguish different communication transmitter individuals, and the method proposed show great robustness under multiple variant experimental conditions.

2. PROPOSED METHOD

2.1. SIB extraction

SIB calculate the integral calculation according to a group of square that centered at the origin as the integral path (shown in Figure 1). It can fully express transmitter signals without missing or repeated bispectrum information, and with time shift invariance, scale variety and maintenance of phase.

Assuming the type of received communication signal is zero intermediate frequency I/Q orthogonal, and the number of signal in total is n. The first step is Fourier transformation, transferring the continuous signal to discrete signal:

\[ F_k(\omega) = \sum_{q=0}^{Q-1} f_k(q) e^{-j\omega q} \]  (1)

Where , Q is the length of signal data. Then the bispectrum can be expressed as:

\[ B_1(\omega_1, \omega_2) = F_1(\omega_1)F_1(\omega_2)F_1(-\omega_1 - \omega_2) \]  (2)

If we assume the length of SIB vector to be G, the calculation of SIB of communication signals can be expressed as:

\[ S(g) = \sum_{g_0} B(\omega_1, \omega_2), \quad g = 0, 1, 2, \ldots, G-1 \]  (3)

Where Pg means the g-th integral path and g=0, 1, \ldots, G-1. Finally the SIB to be input to the stacked autoencoder network (SAE) is \( \{S_i(g)\}_{g=0}^{G-1} \) (S for short).
2.2. signal fine feature extraction by SAE
SAE is a kind of unsupervised learning algorithm which piles up multiple autoencoder units to be a stack, having the ability of reducing the dimension of features in every layer. In the Figure 2, the first layer is input layer, and the last layer is output layer, the rest are hidden layers. And \( x \) means the input feature vector, \( W \) represent the weight of path linking the neurons in neighbor layers, \( b \) is the bias unit in each hidden layer, \( z \) is the representation of input vector in each layer.

![Figure 2. Structure of SAE](image)

The core idea of SAE is to nonlinearly mapping the input communication signal by low dimensional feature, to represent the high dimensional data. Since the output reconstruction has to be as close as possible to the input, the low dimensional features in intermediate layers probably contain the class information of high dimensional data, and may contribute to identifying different communication transmitters. Generally speaking, SAE training consists of two parts: feedforward network and back propagation. The purpose is to minimize cost function \( J \) and achieve accurate reconstruction. Feedforward network model is operated by nonlinear activation functions in multiple hidden layers.

Assuming the discrete input signal to be \( x \), the representation of input in the SAE to be \( z \), the number of hidden layers to be \( L \), and the parameter matrix \([W,b]\) to be initialized according to Gaussian distribution , as for the \( l \)-th hidden layer of the SAE, the input feature can be:

\[
\mathbf{z}^{(l)} = \mathbf{W}^{(l)} \mathbf{z}^{(l-1)} + \mathbf{b}^{(l)}
\]

Where the input signals are defined as \( x = z^{(0)} \), \( W^{(l)} \) is the weight coefficient vector of the \( l \)-th neurons, and \( b^{(l)} \) is the bias unit vector of the \( l \)-th neurons. So when it comes to \( l = L \), the input feature would be expressed in the deepest hidden layer by:

\[
\mathbf{z}^{(L)} = \mathbf{W}^{(L)} \mathbf{z}^{(L-1)} + \mathbf{b}^{(L)}
\]

If we take a nonlinear function \( f(z) = \text{sigmoid}(z) = \frac{1}{1 + e^{-z}} \) to represent the nonlinear transformation of input feature in this paper, and thus the nonlinear activations of neurons in hidden layers are:

\[
a^{(l)} = \text{sigmoid}\left(\mathbf{z}^{(l)}\right)
\]

The decoder implements the feature mapping from back towards front by the \( L \)-th layer neurons. After the nonlinear transformation of \( L \) layers, the reconstruction of input is output to be \( \hat{x} \) at the \( n_i \)-layer \( (n_i = 2L) \). So in the hidden layers of decoder \( l = (L+1): (n_i - 1) \), we will have:

\[
\mathbf{z}^{(l)} = \mathbf{W}^{(n_i-l)} \mathbf{z}^{(n_i-l-1)} + \mathbf{b}^{(n_i-l)} \quad \text{(input represented in decoder)}
\]

\[
a^{(l)} = f\left(\mathbf{z}^{(l)}\right) \quad \text{(nonlinear activation in decoder)}
\]

\[
\hat{x} = \mathbf{z}^{(n_i-1)} = \mathbf{W}^{(L)} \mathbf{z}^{(n-2)} + \mathbf{b}^{(L)} \quad \text{(reconstruction output by decoder)}
\]

The learning performance of SAE is estimated by the difference between reconstruction and input. The reconstruction error is defined as:
ReconError = \hat{x} - x^2 \tag{10}

The decoder performs the feature mapping from the back to the front through the L-th layer, and through the nonlinear transformation of the L hidden layer, the reconstruction of the feature \( \hat{x} \) is output at the output layer \( n \)-layer \( (n = 2L) \). In order to make the definition of reconstruction error in the next step, it can be seen from the following that the reconstruction error has important significance to the definition of the cost function. The role of the decoder is back propagation of the neural network to ensure that the cost function is minimized. According to the reconstruction error, the cost function of SAE can be defined. And when it becomes a minimum, the reconstruction will come to least, too. The cost function is:

\[ J(W, b; x) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} \hat{x}_k - x_k^2 + \frac{\lambda}{2} \sum_{l=1}^{L} \sum_{i=1}^{S_l} \sum_{j=1}^{S_{l+1}} (W_{lji})^2 \]
\[ + \beta KL(\hat{\rho}||\rho) \tag{11} \]

Where the first term is reconstruction error, \( m \) is the number of input features, \( \hat{x}_i \) and \( x_i \) represent the k-th element of reconstruction and input, respectively. The second term is weight decay, \( \lambda \) is the weight decay coefficient, \( S_{l+1} \) and \( S_l \) represent the number of neurons of the l+1-l-th and the l-th layer respectively, and \( W_{lji} \) is the weight of propagation path between the j-th neuron in the l-th layer and the i-th neuron in the l+1-th layer. The third term is sparsity constrain, \( \beta \) is the sparsity constrain coefficient, \( KL(\hat{\rho}||\rho) \) means the relative entropy of activation of all neurons and the average activation preset, and can calculated by:

\[ KL(\hat{\rho}||\rho) = \rho \ln \frac{\rho}{\hat{\rho}} + (1 - \rho) \ln \frac{1 - \rho}{1 - \hat{\rho}} \tag{12} \]

Where \( \rho \) is the set average incentive value, a value close to 0. \( \hat{\rho} \) is the average of all the neurons activated. When \( \rho \) is a constant \( \rho_0 \), relative entropy \( KL(\hat{\rho}||\rho) \) has a minimum value at \( \rho = \rho_0 \). The first two items in the formula (11) are non-convex functions, which can be solved by iterative method. The number of iterations is \( I \), and the gradient descent direction of the cost function is solved by reverse propagation, the parameter \( [W, b] \) is updated, thus the cost function is minimized.

The activation \( a^{(n)} \) of the output layer is calculated according to the formula (8), thus the residual vector \( \delta^{(n)} \) of the output layer can be calculated:

\[ \delta^{(n)} = - (x - a^{(n)}) f'(z^{(n)}) \tag{13} \]

As for the l-th \((l = n_1 - 1, n_2 - 2, \ldots, 2, 1)\) hidden layers, the residual vector of each layer is calculated from the back forward:

\[ \delta^{(l)} = \left( (W^{(l)})^T \delta^{(l+1)} \right) f'(z^{(l)}) \tag{14} \]

Therefore, the gradient descent vector of the weight coefficient and the bias coefficient of the L-th layer are:

\[ \nabla_{w^{(L)}} J(W, b; x) = \delta^{(L+1)} (a^{(L)})^T \tag{15} \]
\[ \nabla_{b^{(L)}} J(W, b; x) = \delta^{(L+1)} \tag{16} \]

It is known that the updating method of the weight coefficient and bias coefficient of layer l is:

\[ W^{(l)} = W^{(l)} - \alpha \nabla_{w^{(l)}} J(W, b; x) \tag{17} \]
\[ b^{(l)} = b^{(l)} - \alpha \nabla_{b^{(l)}} J(W, b; x) \tag{18} \]

In the upper form, \( \alpha \) is the learning rate. When the number of iterations reaches I, the input feature distribution \( z^{(l)} \) of the L layer represents the fine feature of the signal, DSIB.
2.3. DSIB feature extraction process

The process of extracting the characteristics of the signal DSIB based on the stack autoencoder network is shown in Figure 3, and the specific steps are described below.

![Figure 3. Flowchart of DSIB feature extraction method](image)

The zero intermediate frequency I/Q dual channel orthogonal signal $r(t)$ is input to the communication source, which is carried out as follows:

**Step 1** A rectangular integral bispectrum transformation is performed on the input signal $r(t)$ based on the formula (1) ~ (3), and the SIB characteristic $S$ of the input signal is obtained.

**Step 2** In order to facilitate the learning of stack self-coding network, the $S$ is normalized and one dimension expansion is converted into vector $x$. Where

$$S_{\text{Norm}} = \frac{S - \text{min}(S)}{\text{max}(S) - \text{min}(S)}$$  \hspace{1cm} (19)

In the formula, the maximum and minimum eigenvalues in the matrix $S$ are expressed by $\text{max}(S)$ and $\text{min}(S)$ respectively. Then $x_{(i-1)\times N+j}^{N \times 1}$ represents the $j$-th characteristic value of the matrix $S$'s $i$-th line, and $x_{(i-1)\times N+j}^{N \times 1}$ represents the eigenvalue of the $(i-1)\times N + j$ of the vector $x$.

**Step 3** The labeled sample $x_c$ is input to the supervised stack autoencoder network training, and its cost function is different from the formula (11). The average negative logarithm probability of the reconstructed cost function is the output reconstruction.

$$J_c(W_b; x) = -\frac{1}{m} \sum_{n=1}^{m} \log P(\hat{x}_n | x)$$  \hspace{1cm} (20)

Where $m$ is the dimension of the one dimensional vector that is input to the SIB feature. Calculate the supervised cost function $J_c$. When the current price function is the minimum, the supervised parameter $[W_c, b_c]$ can be obtained. This parameter has important significance for the training of the next test set. It will be used to initialize the unsupervised self-encoding network, in order to get DSIB features, the specific content as shown in step (4).

**Step 4** $[W_c, b_c]$ to initialize unsupervised self-encoding stack network, and no label SIB $x_o$ according to the characteristics of type (4) ~ (18) training, when the unsupervised cost function by gradient descent iterative calculation to obtain the minimum value when the input output distribution of the L layer of $z^L$, namely DSIB feature.

If we need to use DSIB feature to identify communication source, we only need to initialize softmax classifier [9] with DSIB feature, and then input test samples to classify and recognize.

3. Experimental Results

3.1. The composition of the experimental system

The data set used in the experiment is composed of the speech signal of 10 Kenwood handsets which are collected in the actual environment. The signal acquisition process is like Figure 4.
The signal style is zero intermediate frequency I/Q double path orthogonal signal, the center frequency is 160MHz, and the bandwidth is 25kHz. The receiver channel bandwidth is 100kHz, the RF receiver output signal frequency is 12.8MHz, the sampling frequency is 204.8kHz, the sampling time is 15s, each handset collects 3 speakers' voice segments. The total number of samples obtained by sampling $3.072 \times 10^7$, which serves as the data set of the sample is $2.048 \times 10^7$.

The data set is divided into three parts: the label training sample set, the unlabeled training sample set and the test sample set. The sample size of subdivided subsets is shown in Table 1.

| Serial number | Labelled training sample | Unlabeled training sample | Test sample |
|---------------|--------------------------|----------------------------|-------------|
| E1            | 2.048×106                |                            | 1.024×106   |
| E2            |                          |                            | 2.048×106   |
| E3            |                          |                            | 3.072×106   |
| E4            |                          |                            | 1.024×106   |
| E5            | 4.096×106                | 1.024×107                  | 2.048×106   |
| E6            |                          |                            | 3.072×106   |
| E7            |                          |                            | 1.024×106   |
| E8            | 6.144×106                |                            | 2.048×106   |
| E9            |                          |                            | 3.072×106   |

Due to the use of stack self coding network in the DSIB feature extraction method, we need to design the appropriate network hidden layer to ensure good performance. The main parameters of the network performance are the depth of the hidden layer and the width of each layer [10], and the other parameters are set as shown in Table 2.

| Weight attenuation coefficient $\lambda$ | Sparsity constraint coefficient $\beta$ | Average excitation value $\rho$ | Learning rate $\alpha$ | Iteration $I$ |
|------------------------------------------|----------------------------------------|---------------------------------|------------------------|----------------|
| $3 \times 10^{-3}$                       | 3                                      | 0.2                             | 0.01                   | 400            |

Considering the recognition performance verification experiments using SIB features at least 64 dimension, it will be the bottleneck of network width is set to 64, and according to the principle of increasing layer, design 64, 256-64 and 1024-256-64 three kinds of network structure. Figure 5 shows the reconstruction error of three kinds of structures. It can be seen that the self encoded network of 1024-256-64 structure has the smallest reconstruction error. Therefore, in the recognition performance verification experiment, the supervised stack self coding networkand the unsupervised stack self coding network adopt this structure.
Figure 5. Reconstruction error of stacked autoencoder network with different structures

In the recognition performance verification experiment, the control algorithm is the maximum correlation entropy method[2] (MCER) of the traditional rectangular integral bispectrum method[3] (SIB/PCA) and the rectangular integral bispectrum. The SIB features are processed by principal component dimension reduction to get the fine features of the signal.

The recognition performance verification experiments were conducted according to the following procedure: divided the labeled training set input control method model, extracting the corresponding communication emitter signal subtle features in classifying softmax classifiers, comparing the classification results and label categories are consistent. There will be no label label training set and the training set were used to train the algorithm supervised and unsupervised self encoding network stack stack from the encoding network, and unsupervised feature extraction DSIB encoding stack from the network, the input of softmax classifier, used to classify the test sample. The recognition experiment was carried out 50 times, and the average recognition accuracy was used as the index of evaluation recognition performance, and the performance of the algorithm and the control method were compared. The experimental platform is a host of CPU Intel E5-2620v4 2.6GHz.

3.2. Identification performance verification experiment

From the experimental characterization ability and robustness of the three aspects of testing DSIB characteristics of the emitter, the experimental variables are SIB feature dimensions, test sample number and the number of individuals to be identified, the identification number corresponding to D, T, C, a specific set of experimental variables as shown in Table 3.

Table 3. Variable settings of identification performance experiments

| Serial number | SIB dimension | Selected sample set | Number of individuals to be identified |
|---------------|---------------|---------------------|----------------------------------------|
| D1            | 64            | E1 E4 E7            | 5                                      |
| D2            | 128           | E1 E4 E7            | 5                                      |
| D3            | 256           | E1 E4 E7            | 5                                      |
| T1            | 64            | E1 E4 E7            | 5                                      |
| T2            | 64            | E2 E5 E8            | 5                                      |
| T3            | 64            | E3 E6 E9            | 5                                      |
| C1            | 64            | E1 E4 E7            | 5                                      |
| C2            | 64            | E1 E4 E7            | 10                                     |

The three SIB features used in experiment D, as shown in Figure 6, test the performance of fine features extracted by algorithms under different SIB dimensions. Obviously, with the increase of the SIB feature dimension, the information complexity is increasing, providing a lot of learning information and increasing the complexity of the learning process. The experimental results of the recognition performance of the DSIB features extracted from the different SIB features of the signal are shown in Figure 7.
The results can be obtained by the graph, this algorithm in D1, D2, D3 recognition rate in SIB was higher than that of the control algorithm, the characteristic dimensions of higher learning increases the available information, but no obvious influence for the characterization of DSIB features, the recognition rate stable at around 80%; the control algorithm is not, with the increase of SIB the feature dimension, a slight increase in recognition performance, illustrate the subtle feature extraction SIB/PCA and MCER methods are dependent on the amount of training samples.

To speed up the training, the experimental T is carried out on the condition that the SIB characteristic dimension is 64 and the number of individuals to be identified is 5. It is verified that whether the DSIB characteristics can still distinguish the radio individuals accurately when the number of test samples increases. The results of the experiment are shown in Figure 8. It can be seen that with the increase of the number of test samples, the ability of DSIB to distinguish the individual of radio has a downward trend, but it still ensures effective performance of individual identification. The recognition performance of the control method is more obvious, especially in the E1–E6 with less label sample. Experiment C with 10 radio data set for training, respectively, 5 and 10 recognition radio stations, in order to check the performance of DSIB features to identify more individual and individual recognition, is whether there is a large deviation. The results of inspection are as shown in Figure 9.

It can be seen that when the identification number of individuals increased from 5 times to 10, individual identification performance characteristics of DSIB remained unchanged, while the control method was decreased with varying degrees of performance, and in a label training set E1 and E4 lower performance by more than E7.

To sum up the above experimental phenomena, Table 4 lists the features of the DSIB and the control algorithms in each experiment.
Figure 9. Results of Experiment C

Table 4. Summarization of identification performance verification experiment results

| Experimental variable identification number | DSIB Performance                                                                 | Performance of control group                                                                 |
|---------------------------------------------|----------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| D                                           | Basically invariant, better than the control method                              | As the SIB dimension increases, performance improves                                        |
| T                                           | The more the test sample is, the lower the recognition rate, and the lowest recognition rate when the test sample is more than the label training sample. | The more the test samples are, the lower the recognition rate is, the decrease is more than DSIB |
| C                                           | The number of individuals to be identified is increasing, and the recognition performance is basically unchanged | The performance is significantly decreased, and when the label sample is less, the decrease is greater. |
| E                                           | The recognition rate is on the rise with the increase of the label training sample, and the recognition rate is lowest in less than the test sample. | The more label samples, the better the recognition performance.                              |

A comprehensive analysis of Figures 7, 8, 9, and table 4 can sum up the features of the DSIB feature extraction algorithm.

1. we can use the SIB characteristics of the radiated source signals in the label free communication, and use it to represent the subtle information of the individuals of the communication emitter, so that we can get a fairly individual discrimination with the fine features extracted from the labeled samples.
2. stack self coding network can accurately characterize the potential fine features of communication source signals. The greedy learning method can effectively decompose the information contained in high-dimensional input samples, and finally get low dimensional fine features.
3. first, we use label samples to initialize the stack self coding network through supervised training, which sets an accurate direction for subsequent learning of unlabeled samples, and helps to improve the representation ability of the subtle characteristics of communication emitter signals.
4. under the condition of small samples, DSIB features can maintain robust recognition performance for multiple communication emitter individuals, and the recognition rate is higher than that of SIB based control methods, which shows that the algorithm has good robust performance.

4. Conclusion and Future Work
In this paper, under the condition of small sample stack based self encoding network extraction communication radiation source signal in deep rectangular integral bispectrum feature, feature based
on rectangular integral bispectrum and extraction, improved the rectangular integral bispectrum characterization of emitter accuracy defects. In the feature extraction phase, with the idea of semi supervised learning, we first use labeled samples to initialize the stack self encoded network, and extract the accurate and effective communication source signal fine features from a large number of unlabeled samples. Experiments were carried out in different SIB feature dimensions, number of training samples, and the number of samples to identify the number of individuals in variable conditions, experimental results show that the DSIB signal feature extraction algorithm in the condition of small samples with individual strong representation capability, and should be used for the low sensitivity of the scene changes, robustness. The next step is not limited to the study of rectangle integral bispectrum, but also to deal with all kinds of characteristics including the individual information of the communication emitter signal, and improve the accuracy of the subtle characteristics to the individual of the radiation source.

Acknowledgments
This work is supported by the grants of the National Science Foundation of China, Nos. 61272333, the National Defense Foundation of China under Grants Nos. 9140C130502140C13068 & 9140A33030114JB39470. The authors would like to thank all the guest editors and anonymous reviewers for their constructive advices.

References
[1] Chen Zhiwei, Xu Zhijun, Wang Jinming, et al. Emitter identification method based on cyclic spectrum density slice[J]. Journal of Data Acquisition and Processing, 2013, 28(3):284-288.
[2] Tang Zhe, Lei Yingke. Method of individual communication transmitter identification based on maximum correntropy[J]. Journal of Communications, 2016, 37(12):171-175.
[3] Xu S H. Radio transmitter classification using a new method of stray features analysis combined with PCA[C]// Military Communication Conference, 2007:1-5.
[4] Tao Wanglin, Lu Xuanmin, Liu Lijuan. Transmitter individual identification based on local surrounding-line integral bispectrum. Computer Engineering and Applications, 2013, 49(1):131-133.
[5] Hinton G. Reducing the dimensionality of data with neural network[J]. Science, 2006, 313:504-507.
[6] Hu Zhaohua, Song Yaoliang. Dimensionality reduction and image reconstruction based on continuous autoencoder network[J]. Journal of Data Acquisition and Processing, 2010, 25(3):318-323.
[7] Wang Yong, Zhao Jianhui, Zhang Dengyi, et al. Forest fire image classification based on deep neural network of sparse autoencoder[J]. Computer Engineering and Applications, 2014, 50(24):173-177.
[8] Ma Yong, Bao Changchun. Speaker segmentation based on sparse neural network[J]. Journal of Beijing University of Technology, 2015, 41(5):662-667.
[9] Arribus J I, Cid-Sueiro J, Adali T, et al. Estimates of constrained multi-class a posteriori-probabilities in time series problems with neural networks[C]// Proc of the International Joint Conference on Neural Networks. Washington, USA: IEEE, 1999:1560-1561.
[10] Coates A, Ng A Y, Lee H. An analysis of single-layer networks in unsupervised feature learning[C]// International Conference on Artificial Intelligence and Statistics, 2011:215-223.