International Journal of Robotics and Automation, Vol. 36, 2021

Special Issue: Artificial Intelligence in Environmental Automation Systems

FEATURE EXTRACTION OF MOTOR IMAGERY EEG SIGNALS BASED ON MULTI-SCALE RECURRENCE PLOT AND SDA

Wenbo Wang,*,**,***,**** Lin Sun,** and Guici Chen*

Abstract

When recognizing multi-class motor imagery electoencephalogram (EEG) signals directly using stacked denoising autoencoders (SDA), it is difficult to fully train the weights due to the small sample size, which results in poor classification effect. To overcome this problem, the multi-scale recurrence plot and SDA method are combined to extract features of multi-class motor imagery EEG signals for recognition. Firstly, multi-class motor imagery EEG signals are decomposed into a series of intrinsic mode functions (IMFs) with different scale by synchrosqueezed wavelet transform, and the recurrence plot of each IMF is constructed to form one-level feature data as input samples of SDA. Then, high-level abstract features which can better express category attributes are extracted from multi-scale recurrence plot by SDA. Finally, EEG signals are classified by using Softmax classifier. Four types of motor imagery EEG data of Datasets 2a in BCI Competition IV are used to verify the proposed method. The average classification accuracy is 0.89, which shows that the proposed method has good effectiveness and robustness.

Key Words

Motor imagery EEG, feature extraction, synchrosqueezed wavelet transform, recurrence plot, stacked denoising autoencoder

1. Introduction

Brain–computer interface (BCI) allows direct communication of information between the brain and the outside world without passing through human nerves and muscle tissues. Serious neuromuscular disorder patients, such as cerebral palsy and spinal cord injury, can use BCI system to control external auxiliary equipment such as wheelchair and service robot through brain nerve activity, so as to restore the ability of movement and communication to a certain extent [1].

When people perform different limb movements or imagines, specific electoencephalogram (EEG) signals are produced in the sensory areas of brain neurons [2]. BCI system captures EEG signals of motor imagery task and recognizes them, so as to realize information exchange and control between human brain and external equipment. However, the motor imagery EEG signal is a highly complex non-linear and non-stationary signal. How to extract features from EEG and recognize the motion imagination task effectively is very important for the operation of BCI system.

At present, many methods have been applied to feature extraction of EEG signals in motion imagery, such as wavelet transform [3], [4], [7]–[9], empirical mode decomposition [5], common spatial mode method [6], [7], Bayesian network method [8] and phase space recurrence plot method [9], [10]. Among them, the phase space recurrence plot method is widely used in feature extraction of EEG signals and has achieved good results. Niknazar *et al.* [9] used recurrence plot to classify EEG signals of epilepsy patients. In the EEG classification tests of healthy and sick patients, the classification accuracy reached 98.67%. Shabani *et al.* [10] analysed EEG signals under drowsiness and alertness by quantitative analysis of recurrence plot. The classification experiments were carried out by using permutation entropy and recursion rate. The experimental results show that the classification accuracy can reach 90%. Bian *et al.* [11] analysed and studied the EEG signals of acupuncture and moxibustion with phase space reconstruction recursive graph. The results showed that the deterministic recursive vector feature and correlation dimension could effectively distinguish the EEG signals before and during acupuncture.

However, these methods have not achieved ideal accuracy in multi-class motor imagery task recognition. The main reason is that for many kinds of motor imagery EEG signals, although the phase space reconstruction recurrence plot can reveal the internal structure of EEG signal sequence to the greatest extent and provide the recursive

* School of Science, Wuhan University of Science and Technology, Wuhan, China; e-mail: wangwenbo@wust.edu.cn, 11937185@qq.com

** College of General Education, Wuchang University of Technology, Wuhan, China; email: 19404425@qq.com

*** Hubei Key Laboratory of Transportation Internet of Things, Wuhan University of Technology, Wuhan, China

**** National Engineering Research Center for Water Transport Safety, Wuhan, China

Corresponding author: Wenbo Wang

Recommended by Dr. Dong Ren

(DOI: 10.2316/J.2021.206-0613)

state of EEG system, the recurrence plot still has complex spatial distribution and low discrimination, which cannot provide effective features for subsequent recognition. To improve the classification accuracy, firstly, multi-scale filtering is needed for EEG signals, and multi-scale recurrence plots are used to provide more concise and effective distinguishing features. Secondly, it is necessary to extract more distinguishing features from recurrence plot. Recently, the synchrosqueezed wavelet transform (SWT) and deep learning theory provide a good way to construct multi-scale recurrence plots and extract the quadratic features of recurrence plots.

SWT is a non-linear time–frequency reallocation algorithm developed on the basis of continuous wavelet transform [12]. By squeezing the coefficients of continuous wavelet transform in the direction of frequency domain, SWT can obtain more accurate time–frequency curve. So, SWT method can not only improve the phenomenon of mode aliasing, but also have good robustness to noise. SWT has been widely used in the analysis of electrocardiogram and EEG signals [13], [14] and has achieved better results than wavelet analysis and empirical mode decomposition. Stacked denoising autoencoders (SDA) [15] is a typical deep learning network model. It imitates human brain mechanism to interpret data and forms more abstract and high-level feature expression by combining low-level features. However, SDA learns the features of data automatically in semi-supervised way. If the recurrence plots of motor imagery EEG signal are directly used as the original data, because the texture of the recurrence plots is too complex and the discrimination between recurrence plots is low, the features that SDA learns may not be closely related to the category attributes, which ultimately leads to the unsatisfactory classification results [16], [17] After SWT processing, the modal components of different frequencies are extracted separately. The recurrence plot of each modal component eliminates a lot of interference information, which provides conditions for SDA to further extract category features.

To fully extract the effective features of motion imagery EEG data and obtain better classification results, this paper combines SWT recursive graph and SDA method. For each motion imagination task EEG, firstly, SWT is used to decompose it into a series of intrinsic modal functions and construct the recurrence plot of each intrinsic modal function. Then, the SWT recurrence plots are used as the input signals of SDA to extract high-level features, and a two-level multi-class EEG feature extraction method for multi-class motor imagery is constructed. Finally, in the experiments, the data of Datasets 2a in BCI Competition IV are used to verify the classification effect.

2. Multi-Scale Phase Space Reconstruction Recurrence Plot based on Synchrosqueezed Wavelet Transform

2.1 Synchrosqueezed Wavelet Transform Multi-Scale Decomposition

SWT is a new multi-scale time–frequency analysis method on the basis of continuous wavelet transform. Supposing the multi-component non-linear signal is

$$f(t) = \sum_{k=1}^{K} f_k(t) = \sum_{k=1}^{K} A_k(t)\cos[2\pi\phi_k'(t)]$$

where $A_k(t) > 0$, $\phi_k'(t) > 0$. SWT can accurately analyse the frequency of each component in $f(t)$, and successfully extract the harmonic component $f_k(t)$. The basic theory of SWT is as follows [12]:

Definition 1 (Intrinsic mode type function, IMTF). *If the function $f(t) = A(t)\cos(2\pi\phi(t))$ satisfies the following conditions*:

$$A(t) \in C^1(R) \cap L^\infty(R), \quad \phi \in C^2(R) \quad \inf_{t\in R}\phi'(t) > 0, \quad \sup_{t\in R}\phi'(t) < \infty$$

$$\sup_{t\in R}\phi''(t) < \infty, \quad |A'(t)|, \quad |\phi''(t)| \le \varepsilon|\phi'(t)|$$

Then the function $f(t)$ is called the intrinsic mode type function with precision ε (ε-IMTF). In SWT transform, the signal is decomposed into the sum of a finite number of IMTFs

Definition 2. *If a multi-component harmonic function $f(t)$ is said to be a superposition of well-separated ε -IMTFs with separation d, the set of which is denoted by $A_{\varepsilon,d}$ in the sequel, if there exists a finite K such that*

$$f(t) = \sum_{k=1}^{K} f_k(t) = \sum_{k=1}^{K} A_k(t)\cos(2\pi\phi_k'(t))$$

where all the $f_k(t)$ are ε -IMTFs satisfying:

$$\phi_k'(t) > \phi_{k-1}'(t) \text{ and } |\phi_k'(t) - \phi_{k-1}'(t)| \ge d(\phi_{k-1}'(t) + \phi_k'(t))$$

In the following, the condition involving d will be called the *separation condition.*

Assuming that the coefficients of the continuous wavelet transform of $f(t)$ are $W_f(a,b)$ (a and b represent the scaling and translation parameters of wavelet function, respectively), the instantaneous frequency calculated by $W_f(a,b)$ is $\omega_f(a,b)$, the main conclusions of SWT are as follows [12]:

Theorem 1. *Let the function $f(t) \in A_{\varepsilon,d}$ and set $\varepsilon_1 = \varepsilon^{1/3}$. Select a function $h \in C_0^\infty$ and $\int h(t)dt = 1$, select a wavelet function ψ such that the Fourier transform $\hat{\psi}$ is supported in $[\xi_\Psi - \Delta, \xi_\Psi + \Delta]$ with $\Delta < \frac{d\xi_\Psi}{1+d}$. The result obtained by synchrosqueezing $W_f(a,b)$, with threshold ε_1 and accuracy δ, is as follows*:

$$S_{f,\varepsilon_1}^{\delta}(b,\omega) = \int_{A_{\varepsilon_1,f}(b)} W_f(a,b)\frac{1}{\delta}h\left(\frac{\omega - \omega_f(a,b)}{\delta}\right)a^{-3/2}da \tag{1}$$

where $A_{\varepsilon_1,f}(b) = \{a \in R^+; |W_f(a,b)| > \varepsilon_1\}$.

Theorem 2. *When ε_2 is sufficiently small, the perfect reconstruction of component $f_k(t)$ can be realized*, i.e., *for each $k \in \{1, \cdots, K\}$, let*

$$\tilde{f}_k(b) = \lim_{\delta \to 0} \left(R_{\psi}^{-1} \int_{|\omega(a,b) - \phi_k'(b)| < \varepsilon_2} S_{f,\varepsilon_1}^{\delta}(b,\omega) d\omega \right) \quad (2)$$

then there exists a constant C so that for $\forall b \in R$, $\tilde{f}_k(b)$ satisfies

$$|\tilde{f}_k(b) - A_k(b)\cos[2\pi\phi_k(b)]| \leq C\varepsilon_2$$

From (2), *it can be seen that after the EEG signal $f(t)$ is decomposed by SWT, a set of IMTF functions $\{\tilde{f}_k(t), k = 1, \cdots, K\}$ can be obtained.*

2.2 Phase Space Reconstruction Recurrence Plot

Using the famous embedding theorem proposed by Takens [18], the phase space reconstruction of non-linear signals can be carried out. The reconstructed phase space is equivalent to the original dynamic system which generates signals and has the same topological structure as the original dynamic system. After choosing the appropriate delay time τ and embedding dimension m, an m dimensional phase space can be obtained by construction of one-dimensional non-linear signal $\{x(n)|n = 1, 2, \ldots, N\}$, *i.e.*

$$\begin{aligned} \mathbf{x}_i &= [x(i), x(i+\tau), \cdots, x(i+(m-1)\tau)], \\ & i = 1, 2, \cdots, L, \quad L = N - (m-1)\tau \end{aligned}$$

After reconstruction, the phase space trajectory matrix can be expressed as:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_L \end{bmatrix} = \begin{bmatrix} x(1) & x(1+\tau) & \ldots & x_{1+(m-1)\tau} \\ x(2) & x(2+\tau) & \cdots & x_{2+(m-1)\tau} \\ \vdots & \vdots & & \vdots \\ x(L) & x(L+\tau) & \cdots & x_{L+(m-1)\tau} \end{bmatrix}$$

where the row vector $\mathbf{x}_i$ constitutes the phase points of multi-dimensional phase space, and L phase points together constitute the reconstructed phase space trajectory. In phase space reconstruction, the choice of delay time τ and embedding dimension m is very important. In this paper, mutual information analysis method [19] is used to calculate the optimal delay time τ of IMTF functions, and Cao-Liangyue method [20] is used to calculate the minimum embedding dimension m.

After reconstructing the phase space of non-linear signal $x(n)$, calculating the distance r_{ij} between any two phase points $\mathbf{x}_i$ and $\mathbf{x}_j$, then the recurrence plot of $x(n)$ can be defined by r_{ij} as follows:

$$R_{ij} = \theta(\xi - r_{ij}), \quad i, j = 1, 2, \cdots, N - (m-1)\tau$$

where $r_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||$, ξ is distance threshold, $\theta(x)$ is Heavside function, means $\theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$. So the recurrence plot can be represented as:

$$R_{ij} = \begin{cases} 1, & \xi \geq r_{ij} \\ 0, & \xi < r_{ij} \end{cases} \quad (3)$$

Formula (3) shows that the recurrence matrix R_{ij} consists of 0, 1. If 0 in the recurrence matrix is represented by white dots and 1 by black dots, then the recurrence matrix can be transformed into a black-and-white recurrence plot. From the recurrence plot, the change characteristics and trends of the non-linear time series can be clearly and intuitively observed.

2.3 Multi-Scale Recurrence Plot of Motion Imagination Electoencephalogram Signals

After SWT decomposition of motor imagery EEG signals, a set of intrinsic modal type functions $\{\tilde{f}_k(t), k = 1, \cdots, K\}$ with different scales is obtained. By reconstructing the phase space of each $\tilde{f}_k(t)$ and calculating its recurrence plot, the multi-scale recurrence plot of EEG signal can be obtained. Taking the Datasets 2A data set [21] of the 2008 International BCI Competition IV as an example, the EEG signals of the left hand, right hand, tongue and foot motor imagery tasks of the A01 tester were selected. Five-level decomposition is performed by using SWT for the selected EEG signals. The delay time and embedding dimension of each IMTF are calculated, respectively, by mutual information analysis and Cao-Liangyue method, then the multi-scale recurrence plot of each IMTF can be obtained. The results are shown in Fig. 1.

Figure 1 intuitively presents the running state of different motor imagery EEG signals. However, the recurrence plot has a large amount of information, which is not easy to be directly used for classification. So, the recurrence plots are used as the input of SDA, and the secondary features of motor imagery EEG signal are extracted through supervised training.

3. Stacked Denoising Autoencoders

3.1 Autoencoders

Autoencoders (AE) is a feature expression network on the basis of the hierarchical structure of artificial neural networks, which is a classical model of deep learning [22]. For an AE network, if the output is equal to the input, then the network can be trained to reconstruct the input, and the weight of each hidden layer can be obtained. Naturally, we get the centralized expression of input because the sample label is not used in the learning process, so it is an unsupervised method. AE consists of input layer, hidden layer and output layer, the network structure is shown in Fig. 2.

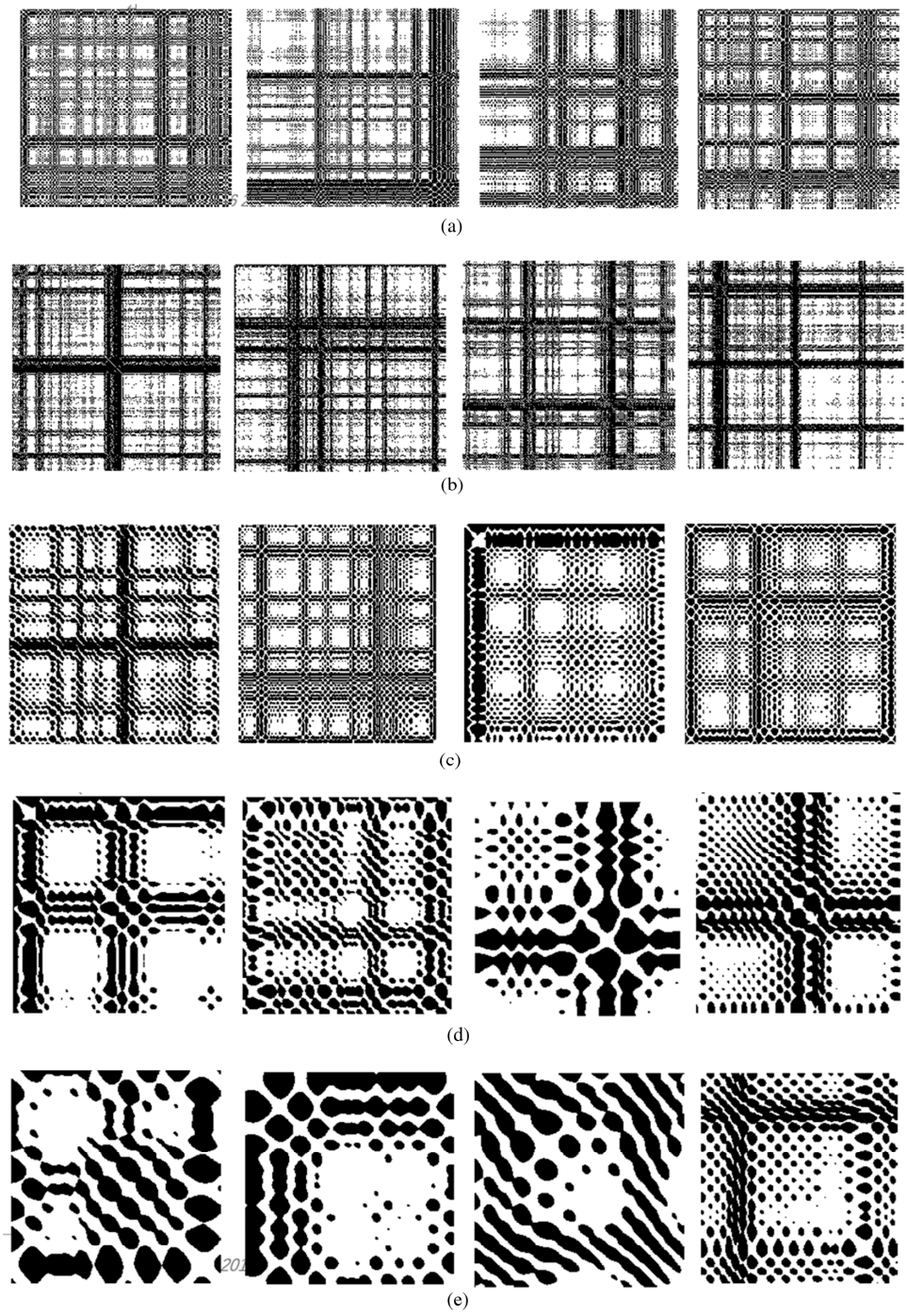

Figure 1. The recurrence plots of IMTF1 (a), IMTF2 (b), IMTF3 (c), IMTF4 (d) and IMTF5 (e), from left to right, are left hand, right hand, tongue and foot of motor imagery EEG signals of the A01 tester.

The work of AE is divided into two processes, encoding and decoding, in which the encoding process is a mapping of input to the hidden layer, expressed as:

$$h = f_\theta(x) = S_f(Wx + b) \tag{4}$$

where S_f is a non-linear activation function, commonly used Sigmoid function, *i.e.* $S_f = 1/(1 + \exp(-x))$, that set of parameters is recorded as $\theta = \{W, b\}$. The decoding process is that the function maps the hidden layer data h back to reconstruct y, which is expressed as:

$$y = g_{\tilde{\theta}}(h) = S_g(\tilde{W}h + \tilde{b}) \tag{5}$$

where S_g is a non-linear activation function using Sigmoid function. The set of parameters is denoted as $\tilde{\theta} = \{\tilde{W}, \tilde{b}\}$. In two sets of parameters θ and $\tilde{\theta}$, the weight matrix

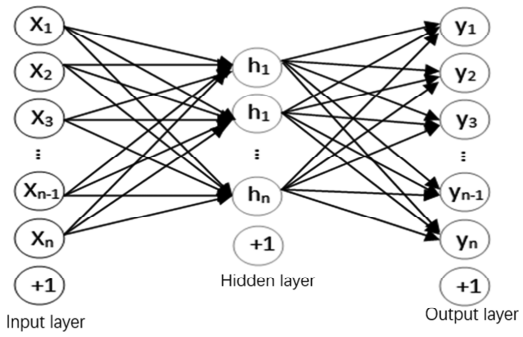

Figure 2. The autoencoder architecture.

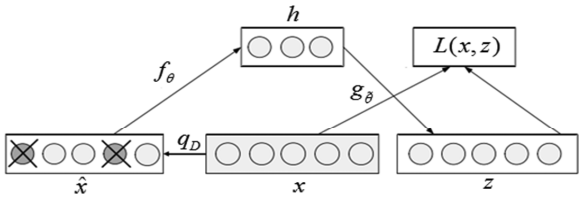

Figure 3. The procedure of corrupting and reconstruction of DAE.

W and $\tilde{W}$ are restricted to satisfy $\tilde{W} = W^T$. Then the optimal parameters are calculated by minimizing the reconstruction error $L(x, y) = ||x - y||^2$ of the network using training sample data.

$$\theta^*, \tilde{\theta}^* = \arg\min_{\theta^*, \tilde{\theta}^*} L(x, g_{\tilde{\theta}}(f_\theta(x))) \tag{6}$$

3.2 Denoising Autoencoders

Denoising autoencoder (DAE) is a distortion of AE model. In DAE, adding a certain proportion of noise to the input data x of AE, then, the polluted data x' is trained as input to the encoder, and the parameters are adjusted to reconstruct the original input data x [23]. The noise addition and reconstruction process of DAE is shown in Fig. 3. Firstly, according to a certain probability P, the random function is used to randomly select the input data x to be zero and the remaining data remain unchanged. Then, for the data$\hat{x}$ added with noise, according to (6)–(8), the network parameters are trained for reconstructing the original data x to the maximum extent. In Fig. 3, x is the original data, z is the output, $\hat{x}$ is the input data with noise, h is the hidden layer data, $L(x, z)$ is the error of supervisory training and q_D indicates adding noise.

Because DAE network needs to eliminate the contamination caused by noise and reconstruct the original data which is not polluted, it makes the DAE network learn the robust expression of input and also shows that DAE has stronger generalization ability than the general AE network

3.3 Stacked Denoising Autoencoders

By stacking several DAEs together, SDA with deep network structure can be formed [15], [23], as shown in Fig. 4. Each layer of SDA uses the output of the previous layer

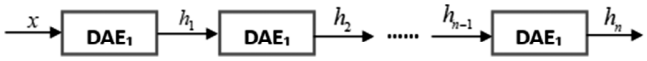

Figure 4. Stack noise reduction autoencoders structure.

as clean input data. The polluted data are obtained by adding noise to the original input data, then the polluted data will be used for network training, so that each layer is a feature representation of the input data.

SDA is a feature extractor, which can compress high-dimensional data into low-dimensional data by training network to get different expressions of original input, so as to achieve the purpose of feature extraction.

4. Two-Level Feature Extraction Method by Multi-Scale Recurrence Plot and Stacked Denoising Autoencoders

Combining the characteristics of multi-scale recurrence plot and SDA in feature extraction, a two-level feature extraction method is proposed for multi-class motion imagery EEG. Firstly, the pretreated data are decomposed by SWT and the recurrence plot of each IMTF is calculated. Then, the recurrence plot of IMTF is used as the input of SDA network for extracting the second-level feature. Because SDA does not have the ability to classify, after second-level feature extraction of SDA network, we select Softmax classifier to classify motion imagery tasks. The steps of EEG feature extraction and classification of multi-class motion imagery tasks are as follows:

Step 1. Because the EEG characteristics of motor imagery are mainly embodied in α rhythm and β rhythm, the 8–30 Hz Chebyshev bandpass filter is selected to preprocess EEG signal.

Step 2. For the filtered training data set, the recurrence plot is used to extract the first-level features. Firstly, the EEG signals of four kinds motor imagery tasks are decomposed by SWT, and the IMTFs are obtained. Then, the recurrence plot of each IMTF is calculated to obtain the first-level feature.

Step 3 The first-level recurrence plot feature of training data set is used as the input of SDA network, and the greedy layer-by-layer training algorithm [17] is used to train the network. That is to say, after adding noise to the output of the former layer, as the input of the latter layer, the layer-by-layer network is trained and the set $\theta = \{W, b\}$ of parameters of the network is calculated.

Step 4 Softmax classifier is added to the n-layer network as the n+1 layer of the network, and the whole network is fine-tuned by supervised learning. In the process of fine tuning, the backpropagation algorithm is used to transfer errors from the last layer to the front layer-by-layer, and the parameter set of the network is updated by minimizing errors.

Step 5 The trained network is used to classify the test data. Firstly, the first-level features of the test data set are extracted in Step 2, and then the second-level features are extracted in the SDA network obtained in Step 4. Finally, the classification results of the test data are calculated.

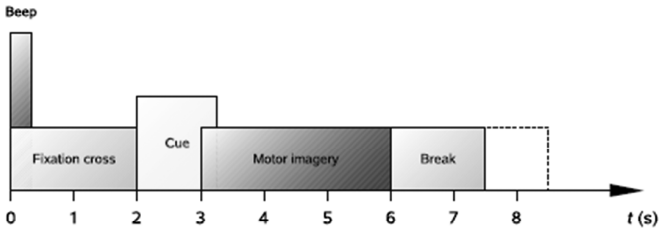

Figure 5. Timing scheme of the paradigm.

5. Analysis of Experimental Results

5.1 Description of Experimental Data

In this paper, two groups of experimental data are used for the experiment. The first set of experimental data is from the Datasets 2A data set of the 2008 International BCI Competition IV, which was provided by Graz University of Science and Technology, Austria. The data included four types of motor imagery tasks: left hand, right hand, tongue and foot movements of nine subjects (Nos. A01–A09). The experimental method of data acquisition is shown in Fig. 5. During the first 2 s after the start of experiment, the subjects sat relaxedly and comfortably in front of the screen, and the screen displayed the "+"' symbol. At the end of 2 s, the screen begins to provide arrows corresponding to the top, bottom, left and right directions of the four tasks. Subjects need to do corresponding motor imagery tasks according to the direction of the arrows that appear. Task imagination lasts for 4 s, *i.e.*, until 6 s. Then there was a period of rest, the subjects relaxed to prepare for the next group of experiments. Each participant's experiment was completed in 2 days. Six groups were collected every day. Each group had 48 motor imagery data. All data were divided into 288 training samples and 288 test samples. In the experiment, 25 channels were collected, 22 of which were EEG, and the other 3 channels were EEG.

The second set of experimental data is from K3b data set of the 2005 International BCI Competition IIIa, which is also provided by Graz University. The data set recorded the subjects' four types of motor imagery tasks: left hand, right hand, foot and tongue, as shown in Fig. 5. Each experiment lasts 7 s and consists of four stages. The experimental process is the same as that of the first experimental data. In the second group of experimental data, 90 experiments were conducted for each motor imagery and 360 experiments were conducted in total. The number of EEG acquisition channels is 60. Hold-out cross-validation method was used to calculate the classification accuracy of 360 groups in data sets (90 times each for left hand, right hand, foot and tongue). The 90 sets of data in each group are divided into two subsets. One group contains 45 sets of data as training set and the other group contains 45 sets of data as test set, which are classified and exchanged with each other. The average value of the two classification results is taken as the final classification accuracy.

In both sets of experimental data, the sampling frequency is 250 Hz, and the filter is carried out by using 0.05–100 Hz bandpass filter and 50 Hz power frequency notch filter. Kappa coefficient is used as the criterion to measure the classification accuracy in the competition. The calculation method of Kappa coefficient is as follows:

$$kappa = \frac{D - 1/C}{1 - 1/C}$$

where C is the correct rate of classification and D is the number of categories.

5.2 Parameters Selection

In extracting the features of the first-level multi-scale recurrence plot, it is necessary to determine the decomposition level number m of EEG signals when they are decomposed by SWT. In the extracting of the second-level feature, the structure of SDA network and noise level needs to be determined. These parameters will directly affect the accuracy of feature extraction and the final classification accuracy.

In this paper, we use the method of 10-fold cross-validation of training sample data to select these parameters. The trend of the average classification accuracy of each subject's data and that of all subjects' data varying with the parameter m is shown in Fig. 6. As can be seen

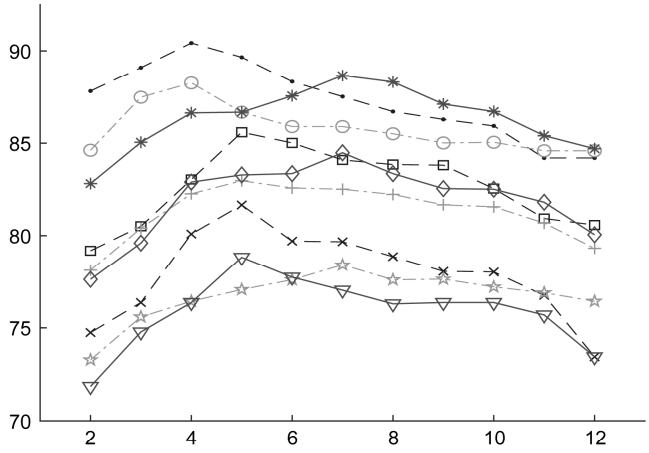

Figure 6. Classification accuracies with various value of m.

Table 1
Mean Kappa Coefficient Variation with the Number of Hidden Layers

Number of layer	2	4	6	8	10
Accuracy	0.7175	0.7532	0.7250	0.7033	0.6852

Table 2
Mean Accuracy Variation with the Number of Units in the Hidden Layer

Combination	24–24–24–24	24–28–24–8	24–32–32–20
Accuracy	0.7504	0.7532	0.7521

from Fig. 6, with the increase in parameter m from 2 to 11, the classification accuracy of each subject's data increases gradually at first, and then decreases after reaching the maximum. This result is consistent with the discussion in [24] that selecting too many features does not improve the classification accuracy. Although the classification accuracy of each subject's data is quite different, the corresponding m values are between 4 and 7 when the accuracy is the highest. When $m = 5$ in SWT decomposition, the average correct rate of classification is the highest for all subjects, considering the adaptability of the algorithm to different subjects, this paper chooses the optimal decomposition level $m = 5$ to extract multi-scale recurrence plot features.

The number of hidden layers and the neurons unit number of each hidden layer are selected separately to carry out experiments. To compare the effects of different layers on classification accuracy, the number of neurons units in each hidden layer is taken as 30, and the average Kappa value varies with the number of layers as shown in Table 1. As can be seen from Table 1, the average Kappa coefficient is optimal for all subjects when the hidden layer is four layers. As the number of layers increases, the network becomes more complex; however, the classification performance decreases significantly. After the number of hidden layers is determined to be 4, the influence of the number of each hidden layers on classification accuracy is analysed. Considering reducing the feature dimension and avoiding over-fitting, besides setting the same number neurons unit of each hidden layer, the network structures with different number neurons unit of each hidden layer are selected for classification experiments. The average Kappa coefficients of three typical composite structures are shown in Table 2. In comparison to the number of hidden layer, the combination of different neurons units of each hidden layer has no obvious effect on the average Kappa coefficients.

Therefore, in SDA network, the number of network layers is set to $n = 4$, the number of neurons unit in each hidden layer is 30–28–24–8, and the noise intensity is $P = 0.12$.

5.3 Analysis of Experimental Results

Four classifications of motor imagery tasks were tested on Datasets 2A data set by using two-level feature extraction method on the basis of multi-scale recurrence plot and SDA. Table 3 presents the classification results of each subject's data. For comparison, both the classification results of the top three in BCI Competition IV on Datasets 2A data set [25] and the classification results of [26] on the same data set are given in Table 3. In the first place of the competition, the improved filter bank CSP is extended to many classes by OVR, and the classifier used in experiments is Naive Bayesian Parzen window classifier [25], [27]. In the second place of the competition, OVO–CSP was used to extract features, then LDA was used to further reduce dimensionality and Bayesian classifier was used to classify EEG signals [25], [28], [29]. CSP is used for feature extraction in the third place of the competition, and SVM is used as a classifier to construct three groups of two-level binary tree multi-class classifiers for classification [25]. In [26], after wavelet packet decomposition, OVR–CSP is used to extract features, and then SVM and BP neural networks are combined for classification.

As can be seen from Table 3, the average Kappa coefficient of all subjects obtained by the proposed method is 0.6325, which is 10.77% higher than the first place in the competition, and 8.73% higher than that of the literature [26]. Because the proposed method carries out the second-level SDA feature extraction on the basis of the multi-scale recurrence plot, the classification performance

Table 3
Comparison of Kappa Coefficient Obtained from Proposed Method, First Three Teams of the Competition and Other Reference Method

	A01	A02	A03	A04	A05	A06	A07	A08	A09	Average
First place	0.6805	0.4214	0.7516	0.4807	0.4008	0.2703	0.7717	0.7522	0.6102	0.5710
Second place	0.6923	0.3451	0.7107	0.4425	0.1614	0.2125	0.6623	0.7318	0.6902	0.5165
Third place	0.3855	0.1890	0.4854	0.3319	0.0749	0.1474	0.2926	0.4975	0.4462	0.3167
Reference [26]	0.7322	0.4525	0.7737	0.4729	0.2290	0.3276	0.7632	0.7731	0.7114	0.5817
Our method	0.7520	0.4807	0.8176	0.5266	0.4128	0.3782	0.7874	0.8010	0.7356	0.6337

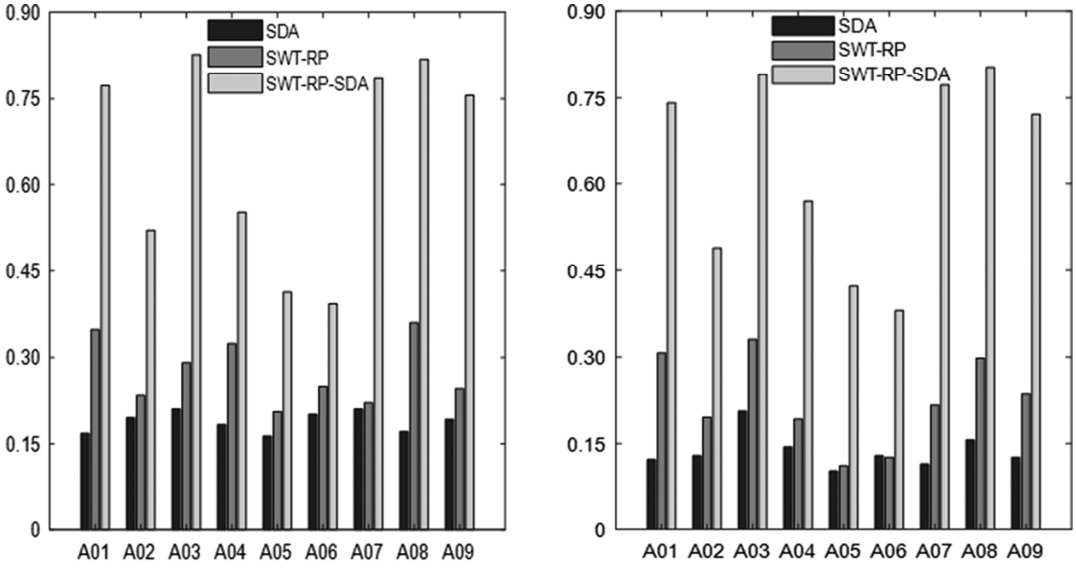

Figure 7. Comparison of classification performance of SDA, SWT-RP and SWT-RP-SDA. Evaluation data comprised of training samples (a) and test samples (b).

is significantly improved than that of the first-level feature extraction method [30], [31], and the overall Kappa coefficient average increases about 0.136 than the first-level feature extraction methods. Although reference [18] points out that the classification performance on the basis of contest data will be influenced by random fluctuation factors, which cannot fully show the performance of an algorithm, the comprehensive comparison results in Table 3 still show that the proposed method has good performance in feature extraction of motion imagery EEG.

To further analyse the advantages of two-level feature extraction methods, the classification performances of multi-scale recurrence plot feature extraction method, SDA feature extraction method and the proposed method are compared [32]. In the experiment, 198 training samples were randomly selected from 288 training samples of the data set to train SDA and classifier. The classification test used two test sets, one consisted of the remaining 90 training samples, and the test results are shown in Fig. 7(a); the other consisted of 90 randomly selected samples from 288 test samples of the data set, and the test results are shown in Fig. 7(b).

As can be seen from Fig. 7(a) and (b), the results of using SDA alone are significantly lower than those of the other two methods. The average correct rate is slightly higher than 0.25, which is close to the performance of random classification. This result shows that when only SDA is used to extract EEG features, it is difficult to learn the features related to category attributes because of complexity of EEG signal. When the test set is training sample, the classification accuracy of two-level feature extraction method proposed in this paper is higher than the single multi-scale recurrence plot method, the advantage is particularly evident in subjects A03, A07 and A09. When the test set comes from the test sample, the accuracy of the two methods decreases, but the two-level feature extraction method is still better than the single multi-scale recurrence plot method. In conclusion, the two-level feature extraction method is superior to the single multi-scale recurrence plot and single SDA feature extraction. In the two-level feature extraction method, the first-level recurrence plot extracts the features that preliminarily differentiate motion imagery tasks. In the second-level feature extraction, SDA network can extract more abstract category expression by self-learning, which can extract more expressive low-dimensional features. Therefore, the classification accuracy is improved significantly.

6. Conclusion

Extracting features that can express motor intentions well from complex motor imagery EEG are the key of BCI system. Although the recurrence plot method achieves good results in the classification of two types of EEG tasks, it is difficult to achieve high classification accuracy for multiple types of motor imagery EEG tasks. SDA network can automatically learn complex high-level features and enhance network generalization ability by adding noise to input. In this paper, SDA network is used to extract more abstract category attribute features from multi-scale recurrence plot based on SWT, and a two-level feature extraction method for multi-class EEGs is constructed. The proposed method achieves good classification results for four types of motor imagery tasks on Datasets 2A data set of BCI Competition IV. Compared with the single multi-scale recurrence plot method and single SDA method, the classification performance of the proposed two-level method is significantly improved, which provides a new idea for feature extraction in multi-class motor imagery BCI. However, the application of backpropagation algorithm in the fine-tuning of SDA network learning in this paper will result to the

problem of local optimal solution, which may be the reason for restricting the further improvement of classification accuracy. Therefore, the optimization of SDA parameter set will be the research content of follow-up work.

Acknowledgement

This research was financially supported by National Natural Science Fund (No. 61671338), Funded by National Engineering Research Center for Water Transport Safety (No. A2019009), Hubei Key Laboratory of Transportation Internet of Things (Wuhan University of Technology) (No. 2018IOT006) and Open Research Fund Program of Key Laboratory of Digital Mapping and Land Information Application Engineering, NASG (No. GCWD201805).

References

[1] W.Y. Hsu, Brain-computer interface: The next frontier of telemedicine in human-computer interaction, *Telematics and Informatics*, *32*(1), 2015, 180–192.

[2] L.F. Nicolas-Alanso, R. Corralejo, J. Gomez-Pilar, D. Alvarez, and R. Hornero, Adaptive stacked generalization for multi-class motor imagery-based brain computer interfaces, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *23*(4), 2015, 702–712.

[3] A.R. Hassan and M.I.H. Bhuiyan, A decision support system for automatic sleep staging from EEG signals using tunable q-factor wavelet transform and spectral features, *Journal of Neuroscience Methods*, *271*, 2016, 107–118.

[4] W.Y. Hsu and Y.N. Sun, EEG-based motor imagery analysis using weighted wavelet transform features, *Journal of Neuroscience Methods*, *176*(2), 2009, 310-318.

[5] M.-H. Yang, W.-Z. Chen, and M.-Y. Li, Multiple feature extraction based on ensemble empirical mode decomposition for motor imagery EEG recognition tasks, *Acta Automatica Sinica*, *43*(5), 2017, 743–752.

[6] A.S. Aghaei, M.S. Mahanta, and K.N. Plataniotis, Separable common spatio-spectral patterns for motor imagery BCI systems, *IEEE Transactions on Biomedical Engineering*, *63*(1), 2016, 15-29.

[7] Y. Zhang, G.X. Zhou, J. Jin, X.Y. Wang, and A. Cichocki, Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface, *Journal of Neuroscience Methods*, *255*, 2015, 85-91.

[8] L.H. He, D. Hu, M. Wan, Y. Wen, K.M. Deneen, and M.C. Zhou, Common Bayesian network for classification of EEG based multiclass motor imagery BCI, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *46*(6), 2016, 843-854.

[9] M. Niknazar, S.R. Mousavi, B. Vosoughi Vahdat, and M. Sayyah A new framework based on recurrence quantification analysis for epileptic seizure detection, *IEEE Journal of Biomedical and Health Informatics*, *17*(3), 2013, 572–578.

[10] H. Shabani, M. Mikaili, and S.M.R. Noori, Assessment of recurrence quantification analysis (RQA) of EEG for development of a novel drowsiness detection system, *Biomedical Engineering Letters*, *6*(3), 2016, 196–204.

[11] H.-R. Bian, J. Wang, C.-X. Han, B. Deng, X.-L. Wei, and Y.-Q. Che, Features extraction from EEG signals induced by acupuncture based on the complexity analysis, *Acta Physica Sinica*, *60*(11), 2011, 118701.

[12] I. Daubechies, J.F. Lu, and H.T. Wu, Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool, *Applied and Computational Harmonic Analysis*, *2*(30), 2011, 243–261.

[13] A. Mert and A. Akan, Emotion recognition based on time-frequency distribution of EEG signals using multivariate synchrosqueezing transform, *Digital Signal Processing*, *81*, 2018, 106–115.

[14] M.M. Kabir, R. Tafreshi, and D.B. Boivin, Enhanced automated sleep spindle detection algorithm based on synchrosqueezing, *Medical & Biological Engineering & Computing*, *53*(7), 2015, 635–644.

[15] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.A. Manzagol, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, *The Journal of Machine Learning Research*, *11*, 2010, 3371–3408.

[16] J.H. Li, Z. Struzik, L.Q. Zhang, and A. Cichocki, Feature learning from incomplete EEG with denoising autoencoder, *Neurocomputing*, *165*, 2015, 23–31.

[17] S.-J. Xu, L.-X. Han, and X.-Q. Zeng, Natural images classification and retrieval based on improved SDA, *Pattern Recognition and Artificial Intelligence*, *27*(8), 2014, 750–757.

[18] F. Takens, Detecting strange attractors in turbulence. Dynamical systems and turbulence, *Lecture Notes in Mathematics*, *898*, 1981, 366–381.

[19] Z. Yin, J. Li, Y. Zhang, A. Ren, K.M. Von Meneen, and L. Huang, Functional brain network analysis of schizophrenic patients with positive and negative syndrome based on mutual information of EEG time series, *Biomedical Signal Processing and Control*, *31*, 2017, 331–338.

[20] L.Y. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, *Physica D*, 110, 1997, 43–50.

[21] M. Tangermann, K.R. Muller, and A. Aertsen, Review of the BCI competition IV, *Frontiers in Neuroscience*, *6*, 2012, 55.

[22] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, Learning representations by back-propagating errors, *Nature*, *323*(6088), 1986, 533–536.

[23] P. Vincent, H. Larochelle, Y. Bengio, and P.A. Manzagol, Extracting and composing robust features with denoising auto-encoders, *Proceedings of the 25th International Conference on Machine Learning*, Helsinki, Finland: ACM, 2008, 1096–1103.

[24] H.X. Wang and D. Xu, Comprehensive common spatial patterns with temporal structure information of EEG data: Minimizing nontask related EEG component, *IEEE Transactions on Biomedical Engineering*, *59*(9), 2012, 2496–2505.

[25] Berlin Brain-Computer Interface, *BCI competition IV-final results* [Online], http://www.bbci.de (accessed Jan. 1, 2016).

[26] M. Li, L. Lin, and S. Jia, Multi-class imagery EEG recognition based on adaptive subject-based feature extraction and SVM-BP classifier, *IEEE International Conference on Mechatronics and Automation*, Beijing, China, 2011, 1184–1189.

[27] M. Li, L. Lin, S. Jia, K.K. Ang, Z.Y. Chin, C.C. Wang, *et al*, Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b, *Frontiers in Neuroscience*, *6*, 2012, 39.

[28] D. Ren, C. Zhang, S. Ren, Z. Zhang, J.H. Wang, and A.X. Lu, An improved approach of cars for Longjing tea detection based on near infrared spectra, *International Journal of Robotics & Automation*, *33*(1), 2018, 97–103.

[29] F. Qu, D. Ren, J. Wang, Z. Zhang, and L. Meng, An ensemble successive project algorithm for liquor detection using near infrared sensor, *Sensors*, *16*(1), 2016, 89.

[30] D. Ren, J. Shen, S. Ren, J.H. Wang, and A.X. Lu, An X-ray fluorescence spectroscopy pretreatment method for detection of heavy metal content in soil, *Spectroscopy and Spectral Analysis*, *38*(12), 2018, 3934–3940.

[31] P. Shao, W. Shi, and M. Hao, Indicator-kriging-integrated evidence theory for unsupervised change detection in remotely sensed imagery, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(12), 2018, 4649–4663.

[32] D. Yang, A. Lu, and J. Wang, Classification of cooked beef, lamb, and pork using hyperspectral imaging, *International Journal of Robotics & Automation*, *33*(3), 2018, 293301.

Biographies

Wenbo Wang was born in Xiangyang, Hubei, China in 1977. He received his B.S. and M.S. degrees in Applying Mathematics from the Wuhan University, Wuhan, in 2003 and the Ph.D. degree in School of Mathematics and Statistics from Wuhan University, in 2006. From 2007 to 2014, he was a research assistant with the College of Science, Wuhan University of Science and Technology. Since 2015, he has been a Professor with College of Science, Wuhan University of Science and Technology. He is the author of 5 books, more than 50 articles. His research interests include signal processing of power system, wavelet analysis theory and artificial neural network.

Lin Sun received her M.S. degree from Central China Normal University, Wuhan, China, in 2005, and a Ph.D. degree from College of Information and Communication of National University of Defense Technology, Wuhan, China, in 2012. Currently, she is an Associate Professor of Wuchang University of Technology and teaches in Wuchang University of Technology. She has published more than 40 papers in various academic journals and conference proceedings. Her current research interests cover mathematical algorithms, modelling and simulation and big data analysis.

Guici Chen received his M.S. degree from Wuhan University of Science and Technology, Wuhan, China, in 2004 and a Ph.D. degree from Huazhong University of Science and Technology, Wuhan, China, in 2010. Currently, he is a Full Professor of the College of Science, Wuhan University of Science and Technology. He has published more than 40 papers in refereed journals and international conference proceedings. His current research interests cover stability and robust control for neural networks, stochastic systems, time-delay systems, *etc.*