Personalized EEG Feature Extraction Method Based on Filter Bank and Elastic Network

Jian-Guo Wang[1(✉)], Zeng Chen[1], and Yuan Yao[2]

[1] School of Mechatronical Engineering and Automation, Shanghai Key Lab of Power Station Automation Technology, Shanghai University, Shanghai 200072, China
jgwang@shu.edu.cn

[2] Department of Chemical Engineering, National Tsing-Hua University, Hsinchu 30013, Taiwan

Abstract. In the practical application of the Brain Computer Interface (BCI) system, because of the diversity between the individuals in the electroencephalogram (EEG) system, the manifestation of Brain signals of each individual is different, so it is necessary to conduct personalized screening for different individuals to obtain information that is conducive to the classification of the EEG signals of the movement imagination. Because the EEG signal manifestation and corresponding rhythm range of different individuals are different, and the EEG characteristics corresponding to different frequency bands are also different, this paper proposes a personalized feature extraction method based on filter bank and elastic network. Based on several commonly used feature extraction and classification algorithms in the current BCI system, the analysis and research are carried out. The best combination method to obtain higher calculation rate and recognition accuracy provides some theoretical reference for the practical application of BCI system. Thus, the shortcomings of the CSP algorithm with better feature extraction effect are improved, and the proposed method can eliminate the individual differences of EEG signals, realize automatic feature selection, and improve classification accuracy.

Keywords: Brain Computer Interface (BCI) · Motor imagery · Elastic net · Feature extraction

1 Introduction

The brain is a system with complex structure and function. It consists of hundreds of millions of neurons, and each neuron relies on the form of electrical signals to transmit information. We call this electrical signal an EEG signal. (Electroencephalogram, EEG). Every moment of human thinking, every kind of emotion, will produce a specific EEG signal, and the EEG signals produced by different thinking states are not the same. The Brain-computer interface (BCI) is a new Human-computer interaction method based on EEG signals. It can transmit directly through the human brain signal by not transmitting through the channels composed of peripheral nerves, muscles and brain. Realize the

Y. Chen et al. (Eds.): BICT 2020, LNICST 329, pp. 116–129, 2020.
https://doi.org/10.1007/978-3-030-57115-3_10

interaction and control of the human brain and external devices [1–3]. A typical brain-computer interface system should be able to quickly and accurately extract and identify EEG information reflecting different mental states of the human brain. This requires designing a corresponding EEG signal processing method according to the specific situation. Based on this, this paper is for EEG. The relevant algorithms of signal analysis and processing have been specifically analyzed and designed.

Through the analysis of spatial patterns found a total of good results have been achieved, however, a total of lack of frequency domain information space model itself, and the classification of the EEG signals accuracy is closely related to brain electrical signal frequency band selection scope, individuality difference, because EEG signals in practice need to manually adjust the specific frequency range for each individual to obtain a higher classification accuracy, limiting its universality and practical applications. In order to solve the above problems, a method of extracting personalized features based on filter Banks and elastic networks is proposed by referring to the idea of using filter Banks to enrich the frequency domain. In this method, the original signal is first divided into 17 sub-band signals with a bandwidth of 4 Hz by filter Banks, and then features are extracted from each sub-band signal by using CSP to obtain a high-dimensional feature set that covers more frequency domain information. Elastic mesh method is used for feature selection, with elastic mesh logistic regression classifier classification error rate as an evaluation standard, by means of parameters optimization ultimately selected contain classification information more feature subset, so as to realize the automatic selection of the characteristic, avoids because of individual differences caused by manually selecting frequency range. Finally, the test data feature set corresponding to the optimal feature subset is fed into the fitted elastic network logistic regression model for classification.

2 A Comparative Study of EEG Signal Feature Extraction and Classification Algorithm

2.1 Signal Processing Algorithm in Brain-Computer Interface

BCI system is built on the basis of EEG, which can realize a new human-computer interaction mode in which the brain directly controls the external equipment or environment. Therefore, in-depth study of EEG signal processing algorithm can not only promote the development of brain cognitive science, but also have important significance for the interpretation of human consciousness and the realization of the practical application of BCI system.

The EEG collected by the signal processing algorithm is analyzed and identified to extract the information reflecting the brain's thinking state, which can be divided into three steps: preprocessing, feature extraction and classification. Preprocessing is to weaken the noise and artifact interference in the signal and improve the signal-to-noise ratio. The main methods are filter filtering, channel and frequency band selection, etc. After preprocessing, the cleaner signal is beneficial to the subsequent signal processing. Feature extraction is to extract the main information that can reflect the intention of subjects from the preprocessed data. The classifier classifies the obtained characteristic information according to a certain criterion, and then converts the classification results into corresponding control signals to realize the control of external devices. This chapter introduces in detail the principles of several common feature extraction and classification algorithms and their implementation results.

Feature Extraction Algorithm. Feature extraction is the core of signal processing in BCI system. The quality of extracted features is directly related to the classification effect of subsequent classifiers and the efficiency of the whole BCI system. Due to the large amount of original signal data, the signal characteristics are not prominent enough, so it is difficult to get good classification results by directly using the original signal for classification. The purpose of feature extraction is to extract usable information from the EEG obtained in the preprocessing link, which can represent the corresponding conscious task, so as to obtain better classification performance.

The following are some common methods for EEG feature extraction in BCI system of motion image: common space mode method, wavelet packet decomposition method and power spectrum estimation method.

1) Common space mode
Common spatial pattern (CSP) algorithm is a classical spatial filtering method that can effectively improve SNR and has been widely applied in BCI system. Practice H. Ramoser first applied to sports like to imagine in the feature extraction of EEG signals, the main idea is under the condition of the labeled training set training, find a space projection, makes classification of two classes of unknown signal after projection, a variance is the largest, another kind of minimum variance, that can maximize the distinguish between two types of samples [4, 5]. The specific implementation process is as follows:
It is assumed that in a left-handed imagination task experiment, the EEG signals sampled in the left-handed imagination are the matrix and dimension respectively, where are the number of EEG signal channels and the number of sampling points in a single training, and the specific implementation of CSP algorithm is as follows:
Suppose that in a left-right hand-imagination task experiment, the EEG signals sampled when imagining the left and right hands are the matrixes X_1 and X_r of the $N \times T$ dimension, where N is the number of EEG signal channels, T is the number of single training samples, and the specificity of the CSP algorithm Implemented as:
First, the left and right hand EEG data X_l and X_r are normalized by covariance:

$$R_l = \frac{X_l X_l^T}{trace\left(X_l X_l^T\right)} \tag{1}$$

$$R_r = \frac{\mathrm{X}_r \mathrm{X}_r^T}{trace\left(X_r X_r^T\right)} \tag{2}$$

In the formula, trace is the trace of the matrix, which is the sum of the diagonal elements of the matrix. Then calculate the average covariance matrices $\overline{R}_l$ and $\overline{R}_r$ in the same task mode. Then the eigenvalues of its mixed space covariance matrix R break down:

$$R = \overline{R}_l + \overline{R_r} = U_0 \Sigma U_0^T \tag{3}$$

Where U_0 is the eigenvector matrix and Σ is the eigenvalue diagonal matrix, then the matrices R, $\overline{R}_l$, and $\overline{R}_r$ are whitened, respectively, and the transformation matrix is:

$$P = \Sigma^{-1/2} U_0^T \tag{4}$$

$$S_1 = P\overline{R}_l P^T = U \sum\nolimits_l U^T \tag{5}$$

$$S_r = P\overline{R}_r P^T = U \sum\nolimits_r U^T = U(I - \Sigma_l)U^T \tag{6}$$

Where I is the identity matrix, $\sum_l$ and $\sum_r$ are the eigenvalue diagonal arrays. Let U_1 and U_r be the feature vectors corresponding to the largest eigenvalues of the eigenvalue diagonal arrays $\sum_l$ and $\sum_r$, respectively. Then you can construct a spatial filter:

$$W_l = W_l^T P \tag{7}$$

$$W_l = W_l^T P \tag{8}$$

Then the EEG data is filtered by the spatial filter to obtain:

$$Z = WX \tag{9}$$

Where X represents the EEG data of the $N \times T$ dimension, and each column in the W matrix is a spatial filter.

The CSP algorithm utilizes the simultaneous diagonalization of matrices to maximize the variance of the two types of EEG data, and then extract features for classification, which has been proved to be an effective feature extraction method [6]. However, due to the lack of frequency domain information in the CSP method, and when the number of signal channels is too small, this will affect its feature extraction effect to some extent, and it also limits its application in BCI to some extent [7, 8].

(2) Analysis of wavelet and wavelet packet
In the 1990 s, Mayer proposed a wavelet transform (WPT) theory based on the Fourier transform. The Fourier transform cannot analyze its time domain-frequency domain simultaneously when processing signals. The problem is improved by using a time-scale window function to analyze the characteristics of the signal, that is, to select different window functions at different frequencies, and to adjust the resolution of the window in the frequency range by changing the shape of the window by panning and stretching. To ensure that it provides the best time-frequency resolution in all frequency ranges [9]. This time-frequency analysis method, which can change both the time window and the frequency domain window, has great advantages in dealing with highly random non-stationary signals such as EEG.
(3) Power spectrum estimation method
The power spectrum estimation method is a classic simple fast frequency domain analysis method. For a random non-stationary signal such as an EEG signal, it may not be clearly expressed by mathematical expressions. At this time, the power

spectrum can be used to signal the signal. Spectrum analysis, which can show the trend of brain wave amplitude over time with the spectrum of EEG power as a function of frequency, so that the distribution and changes of EEG rhythm can be visually observed, and thus extracted in the frequency domain. Important information [10]. Power spectrum estimation can be divided into classical power spectrum estimation and modern power spectrum estimation [11].

Classification Algorithm. Classifying the extracted EEG features is the final step of the BCI system and a very critical step. It maps these EEG signals reflecting the current mode of human activity to the specified classification, and converts them into some control commands to control the external devices. The performance of the classification algorithm directly determines the performance of the entire BCI system. Generally, the performance of the classification algorithm is evaluated from the aspects of classification accuracy, operation rate, simplicity of the model and interpretability. In the following, we will mainly introduce three classification methods that are often used in EEG signal recognition: support vector machine, K-nearest neighbor, and linear discriminant analysis algorithm.

(1) Support vector machine

Support Vector Machines (SVM) is a machine learning algorithm based on statistical learning theory proposed by Vapnik. It can better deal with small sample, nonlinear and high-dimensional pattern recognition problems. Learning performance [12, 13]. Using the kernel function to project linearly inseparable samples into high-dimensional space, the "dimensionality disaster" problem is well overcome without increasing computational complexity. Based on these advantages, it has been widely used in the classification of EEG signals.

The main idea of SVM is to seek an optimal hyperplane that meets the classification accuracy requirements under the premise of ensuring that data is linearly separable in a certain feature space, and successfully distinguish the two types of data while ensuring that the classification interval is as large as possible.

Suppose a training sample set $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, corresponding to two types of labels, and its corresponding category is $\left\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\right\}$, where $\boldsymbol{y_i} \in \{1, -1\}$, $\boldsymbol{N}$ is the total number of feature vectors contained in the training sample. The discriminant function of the optimal hyperplane can be expressed as:

$$y = \omega \cdot x + b = 0 \tag{10}$$

Where ω is the weight vector and b is the classification threshold. In order to ensure that the classification surface solved can correctly classify all the samples in the dataset and the classification interval is the largest, the constraint should be satisfied: ① $y_i f(x_i) > 0, y_i[wx_i + b] - 1 \geq 0$ ② The classification interval $\frac{2}{\|\omega\|}$ is the largest. Since the EEG feature vectors we extracted is linearly inseparable, we need to introduce a kernel function to project the extracted EEG feature vector into the high-dimensional space and construct the optimal hyperplane in the space. The problem of solving the optimal hyperplane is transformed into a constraint optimization problem:

$$\min\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i,\ \text{s.t.}y_i[(w\cdot\varphi(x_i)) + b] \ge 1 - \xi_i,$$
$$\xi_i \ge 0,\quad i = 1, 2, \ldots, n \tag{11}$$

Where ξ_i is the slack variable and C is the penalty factor. The larger the value of C, the greater the penalty for misclassification. The Lagrangian multiplier method is introduced to solve this problem, and the optimal classification function is obtained as follows:

$$y = \text{sgn}\left(\sum_{i=1}^{N} a_i y_i K(x_i, x) + b\right) \tag{12}$$

In the formula, α_i is a Lagrangian multiplier, $K(x_i, x) = \varphi(x_i)\cdot\varphi(x)$ is a kernel function. To ensure the classification accuracy of the SVM, a suitable kernel function should be selected, which is commonly used in SVM. The kernel functions are RBF kernel function, linear kernel function, and Sigmond kernel function.

(2) K nearest neighbor method
The K-Nearest Neighbors (KNN) algorithm is a simple and efficient parameter-free classification model, which is widely used in various fields of pattern recognition [14]. which is to use the sample in the training sample as a template to calculate the distance between each sample x and the template in the test sample according to the distance function, and choose the distance from the unknown sample x. The k templates are taken as the k neighbors of x, and the category to which the test sample x belongs is determined according to the category corresponding to the majority of the k neighbors [15].
High-dimensional data, which is easily affected by high-dimensional disasters.

(3) Linear Discriminant Analysis
Linear Discriminant Analysis (LDA) is a classical classification algorithm proposed by R.A. fisher to transform multidimensional problems into one-dimensional problems [16, 17]. Because of its simple algorithm, fast running speed, small calculation and high robustness, it has been widely used in EEG signal processing [18, 19]. The basic idea is to find the best projection direction, and project the feature vectors corresponding to different categories into this direction, so that the points in the class are more closely clustered, and the points between different classes are more dispersed, thus making different classes Samples are separated as much as possible.

2.2 Experimental Data Source

The data set I is derived from the Data sets 1 data set provided by the 4th Brain-Computer Interface Competition data in 2008. The data set contains the EEG data obtained by the 7 healthy subjects for the motion imaging task, respectively, which are recorded as data sets. a, b, c, d, e, f, and g. Among them, each subject asked to select two experiments from the three left-handed, imagined right-handed, and imagined three-sports imagination tasks. This paper mainly analyzes the EEG signals of right and left hand movements, and selects the data of c, d, e, and g groups for left and right hand movement imaging tasks. During the experiment, each subject performed 200 motion imaging experiments, and the effective time for each motion imaging was 4 s. The experimental process of a motion imaging is shown in Fig. 1.

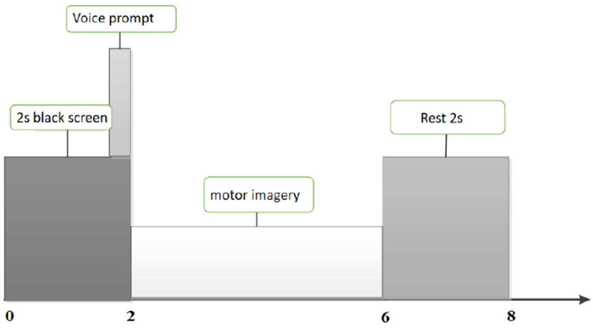

Fig. 1. Picture the process of one motion

First, the display shows a black screen for two seconds to indicate the experimenter's adjustment status, then the screen will display 2 s "ten" character to indicate that the experimenter enters the ready state, and then an arrow indicating the left or right will be randomly displayed, each time is displayed. 4 s, during which the experimenter completes the corresponding motion imaging task according to the head pointing, and the 2 s rest time will disappear after the arrow disappears, and then the next round of motion imaging task will be started. During the experiment, 59 channels of electrode caps were used to record EEG. The distribution of 59 electrodes in the brain region is shown in Fig. 2. The sampling frequency is 1000 Hz. These data are only filtered by 0.05–200 Hz. A detailed description can be found in [20].

3 An EEG Signal Analysis Method Based on Elastic Network Combined with Filter Bank

3.1 Method Introduction

How to quickly and effectively extract EEG features and improve recognition accuracy is the key to BCI technology research. Based on the comparative study of several common feature extraction methods in Sect. 2, it is found that CSP feature extraction method has achieved good results in the brain-computer interface research based on motion imagination. However, the CSP algorithm itself lacks frequency domain information, and the accuracy of classification results is closely related to the frequency range of EEG signals [21]. In practical application, due to the differences between individuals, every individual brain signal form is different, its corresponding appear rhythm signal range is

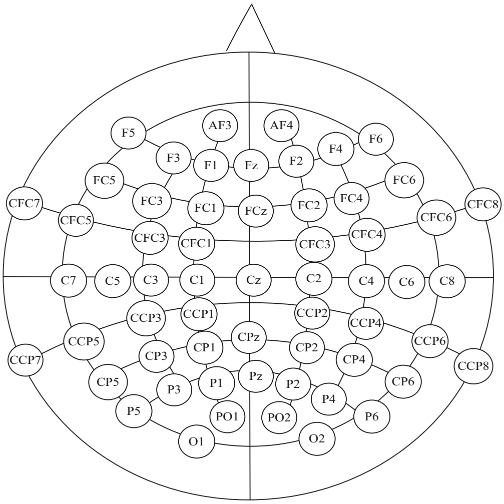

Fig. 2. 59 EEG electrode profiles

different also, in the previous studies on feature extraction using CSP, generally USES a wide range of frequencies or frequency range based on specific individual manual adjustment method for processing, in order to obtain higher classification results, but it is also a certain extent, limits the universality and practicability of the CSP method. To solve this problem, Kai proposed a feature extraction method of Filter Bank Common Spatial Pattern (FBCSP), which used Filter Bank to divide the original signal into frequency bands and then used CSP to extract features in sub-band signals, thus enriching the frequency domain information of CSP feature extraction [22, 23]. In this chapter, based on the idea of frequency band segmentation, a personalized feature extraction method based on filter bank and elastic network is proposed. In this method, the EEG is firstly divided into frequency bands by filter Banks, and the EEG signals of each sub-band are extracted by CSP spatial filter to obtain a high-dimensional feature set. Then, the elastic network method is adopted for feature selection, that is, when training the logistic regression model of elastic network with training data, the classification accuracy of classifier is taken as the evaluation standard, and the model parameters are optimized through 10-fold cross-validation to obtain the feature subset with the highest EEG signal recognition degree applicable to different individuals. Finally, the test data feature set corresponding to the selected optimal feature subset is classified using the trained elastic network logistic regression model, and the corresponding classification accuracy is obtained. The experimental results show that the proposed method is effective.

3.2 Elastic Network Algorithm

Elastic network [24] is a multivariate model analysis method which is the weighted balance of the ridge regression penalty and the lasso penalty, so it combines the advantages of the lasso and ridge regression methods. By optimizing the model parameters, it is possible to find a balance between the goodness of fit and the complexity of the model to select an optimal model, which can make the fitted model more concise and improve the recognition accuracy.

Assume that the collected EEG signal is $X = \{X_1, \ldots, X_N\}$ and its corresponding label is $G = \{g_1, \ldots, g_N\}$, where $g_i \in \{1, -1\}$, $X_i = \left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)^T$ is the predictor variable, g_i is the response variable. To simplify the calculation, assume that the predictor variable x_{ij} has been standardized which is:

$$\sum_{i=1}^{n} x_{ij} = 0, \sum_{i=1}^{n} x_{ij}^2 = 1 \tag{13}$$

Logistic regression models can be expressed as:

$$Pr(g = 1|x) = \frac{1}{1 + \mathrm{e}^{-(\beta_0 + x^T \beta)}} \tag{14}$$

$$Pr(g = -1|x) = \frac{1}{1 + \mathrm{e}^{+(\beta_0 + x^T \beta)}} = 1 - \Pr(g = 1|x) \tag{15}$$

$\beta = \left\{\beta_1, \cdots, \beta_p\right\}$ is the regression coefficient variable of the model. The corresponding log likelihood function is:

$$\begin{aligned} L(\beta_0, \beta) &= \frac{1}{N} \sum_{i=1}^{N} \left[y_i \ln p(x_i) + (1 - y_i) \ln(1 - p(x_i))\right] \\ &= \frac{1}{N} \sum_{i=1}^{N} \left[y_i\left(\beta_0 + x_i^T \beta\right) + \ln(1 - p(x_i))\right] \end{aligned} \tag{16}$$

Where $y_i = \begin{cases} 1 & g_i = 1 \\ 0 & g_i = -1 \end{cases}$, $p(x_i) = Pr(g_i = 1|x_i)$.

Applying an elastic network penalty term $P_\alpha(\beta)$ based on the maximized log likelihood function, its parameter estimates can be described as:

$$(\beta_0, \beta) = \underset{(\beta_0, \beta) \in R^{p+1}}{\arg\max} \{L(\beta_0, \beta) - \lambda P_a(\beta)\} \tag{17}$$

Where $P_\alpha(\beta) = (1 - \alpha)\frac{1}{2}\|\beta\|_{\ell_2}^2 + \alpha\|\beta\|\ell_1 = \sum_{j=1}^{p}\left[(1 - \alpha)\frac{1}{2}\beta_j^2 + \alpha\left|\beta_j\right|\right]$, λ and α are the regularization parameters. Compared with lasso and ridge regression, the elastic network method can simultaneously perform parameter estimation and feature selection.

The log-like maximum value of the above formula (17) is solved by the coordinate descent method [25] which is considered as a fast and effective calculation method. The basic idea of the coordinate descent method is to convert the multivariate problem of unrelated variables between predictors into multiple univariate sub-problems. It optimizes only one-dimensional variables at a time and the optimization coefficients can be updated in the variable cycle, so the whole iteration process will be completed soon.

Before the above formula (17) is solved by the coordinate descent method, the original form needs to be converted. Assuming that the current estimate of the parameter is $\left\{\tilde{\beta}_0, \tilde{\beta}\right\}$, Taylor expansion is performed at the current estimated point and a quadratic approximation of the log-likelihood function of Eq. (4) can be obtained:

$$L_Q(\beta_0, \beta) = -\frac{1}{2N}\sum_{i=1}^{N}\omega_i\left(z_i - \beta_0 - x_i^T\beta\right)^2 + C\left(\tilde{\beta}_0, \tilde{\beta}\right)^2 \tag{18}$$

Where $z_i = \tilde{\beta}_0 + x_i^T\tilde{\beta} + \frac{y_i - \tilde{p}(x_i)}{\tilde{p}(x_i)(1-\tilde{p}(x_i))}$ can be seen as a response. $\omega_i = P(x_i)(1 - P(x_i))$ is Weight. $C(\tilde{\beta}_0, \tilde{\beta})^2 = L\left(\tilde{\beta}_0, \tilde{\beta}\right) + \frac{1}{2N}\sum_{i=1}^{N}\left[(y_i - \tilde{p}(x_i))^2/\tilde{p}(x_i)(1 - \tilde{p}(x_i))\right]$, it is a constant only when parameter optimization is performed; it is a value calculated based on the current parameter estimation value.

The approximation form of the above Eq. (6) is equivalent to the log likelihood part of the above Eq. (5). Then the problem is transformed into a solution to the penalty-weighted least squares form of the elastic network.

$$(\beta_0, \beta) = \underset{(\beta_0, \beta)\in R^{P+1}}{\arg\min}\left\{-L_Q(\beta_0, \beta) + \lambda P_\alpha(\beta)\right\} \tag{19}$$

Solving the Eq. (7) by using the coordinate descent method:

$$\begin{aligned} &-L_Q(\beta_0, \beta) + \lambda P_\alpha(\beta) = \\ &\frac{1}{2N}\sum_{i=1}^{N}\omega_i\left(z_i - \tilde{g}(x_i)^{(j)} - x_{ij}\beta_j\right)^2 + C\left(\tilde{\beta}_0, \tilde{\beta}\right)^2 + \lambda P_\alpha(\beta) \end{aligned} \tag{20}$$

where,$\tilde{g}(x_i)^{(j)} = \tilde{\beta}_0 + \sum_{k\neq j} x_{ik}\tilde{\beta}_k$ is the contact function that removes x_{ij}.

Assuming that the current estimate of the parameter is $\left\{\tilde{\beta}_0, \tilde{\beta}\right\}$, only one dimension of the coefficient β is optimized at a time and other dimensions are considered constant. Then the j-dimension of the coefficient β can be derived:

When $\beta_j > 0$, let the derivative be equal to 0 and the coordinate update form of β_j can be obtained:

$$\beta_j = \frac{\frac{1}{N}\sum_{i=1}^{N}\omega_i x_{ij}\left(z_i - \tilde{g}(x_i)^{(j)}\right) - \lambda\alpha}{\frac{1}{N}\sum_{i=1}^{N}\omega_i x_{ij}^2 + \lambda(1-\alpha)} \quad \left(\beta_j > 0\right) \tag{21}$$

When $\beta_j < 0$, similar expressions could be obtained. In other case $\beta_j = 0$, there is a form of coordinate update described by the soft threshold operator:

$$\beta_j = \frac{S\left(\frac{1}{N}\sum_{i=1}^{N}\omega_i x_{ij}\left(z_i - \tilde{g}(x_i)^{(j)}\right), \lambda\alpha\right)}{\frac{1}{N_i}\sum_{i=1}^{N}\omega_i x_{ij}^2 + \lambda(1-\alpha)} \tag{22}$$

In general, based on the current observations, the elastic network penalty logistic regression model is established. The process of solving the coefficients β by the coordinate descent method is a series of cyclic iterative solving processes and each loop is nested with each other until convergence. The steps of the coordinate descent method are as follows:

(1) Cycling $\ell \in \{1, 2, \cdots, K, 1, 2, \cdots\}$ until β convergence;
(2) Updating the second approximation L_Q using the current parameters $\left\{\tilde{\beta}_0, \tilde{\beta}\right\}$;
(3) Solving the penalty weighted least squares problem of Eq. (5) by the coordinate descent method as shown in Eq. (8).

3.3 Encapsulated Elastic Network Feature Selection Algorithm Combined with Filter Bank

This chapter on the basis of the traditional feature extraction based on CSP was improved, is put forward based on the filter group and elastic mesh personalized feature extraction methods, this method first USES a set of filter band segmentation of EEG signals were collected separately for each sub-band using CSP spatial filter for feature extraction of EEG signals, get a high victor collection; Then use the encapsulation of elastic mesh method for feature selection: the training data is adopted to elastic mesh logistic regression model for training, the objective function is solved by coordinate descent method of model parameter, using 10 fold cross-validation to optimize a corresponding training set the highest classification accuracy of parameter estimation, corresponding to be suitable for different individuals of the EEG signal recognition feature subset supreme; Finally, the test data corresponding to the selected optimal feature subset are classified using the trained elastomeric network logistic regression model, and the corresponding classification accuracy is obtained. The flow chart of the whole method is shown in Fig. 3.

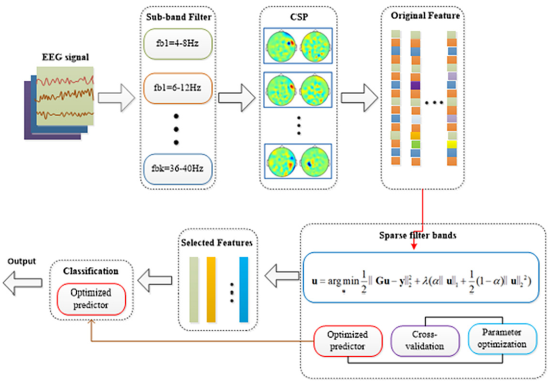

Fig. 3. The implementation process of encapsulated elastic network feature selection algorithm combined with filter bank

Specific Implementation of the Algorithm. The specific implementation of the algorithm is mainly divided into two parts: one is the training sample and the other is the test sample.

The training process is as follows:

(1) Filter bank band segmentation: the training samples are segmented into 17 sub-band signals with a bandwidth of 4 Hz and overlapping 2 Hz using a Chebyshev type II band-pass filter bank. And then the EEG signals of each sub-band are respectively input into the CSP spatial filter and the two-dimensional feature vectors of the two types of motion imaging EEG signals of each sub-band are extracted to obtain a 34-dimensional feature vector set $\boldsymbol{F} = \{\boldsymbol{f}_1, \mathrm{f}_2, \cdots, \mathrm{f}_{34}\}$.
(2) Feature selection: the elastic network logistic regression model is introduced to compress the obtained multidimensional feature variables, and the feature subset with the highest recognition is selected for the individualized EEG signals. The specific selection process is as follows: the elastic network logistic regression classifier error rate is the evaluation criteria for feature selection and 10 different values are set (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0).); the number of values is set to 100 and then 1000 feature selection subsets are obtained (the number of features may or may not be the same); a 10-fold cross-validation is performed to obtain the classification error average rate corresponding to different feature subsets and a set of features with the lowest average error rate is selected as the optimal feature subset.

The test process is as follows:

(1) As in the training process (2), the original signal is sub-band segmented using a filter bank and feature extraction is performed on each sub-band signal by CSP, so that a 34-dimensional feature variable set can be obtained;
(2) The test data corresponding to the selected optimal feature subset in the above training process is sent to the trained elastic network logistic regression classifier to perform prediction classification and the classification accuracy of the test sample is obtained.

3.4 Experimental Results and Analysis

The experimental data is processed accord to the algorithm description of the third section. In order to prove the superiority of the encapsulated elastic network feature selection method combined with the filter bank proposed in this paper, it is compared with the conventional frequency band selection method and the filtered elastic network feature selection method. The conventional frequency band selection method uses CSP to extract the features of the signal in the fixed frequency range of 440 Hz. The filtered elastic network method and the encapsulated elastic network method use the elastic network method to select the feature subsets and the final feature set is the same. The main difference is that the predictive process of the filtered elastic network method uses the classifiers as support vector machine and logistic regression and the prediction process of the encapsulated elastic network method still uses a flexible network logistic regression classifier. Based on the data of four subjects, the recognition accuracy of the five combinations obtained by the above several feature extraction and classification methods were compared. The results are shown in the Fig. 4.

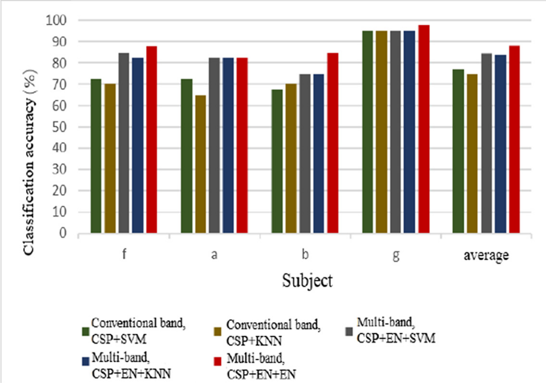

Fig. 4. Classification accuracy rate obtained by five different methods

4 Conclusion

This paper analyzes and processes the motor imagery EEG data in the brain-computer interface system. The main contents are divided into the following aspects: Firstly, based on several commonly used feature extraction and classification algorithms in the current BCI system, the analysis and research are carried out. The best combination method to obtain higher calculation rate and recognition accuracy provides some theoretical reference for the practical application of BCI system. Then some shortcomings of the CSP algorithm with better feature extraction effect are improved. A personalized feature extraction method based on filter group and elastic network is proposed, which can eliminate the individual differences of EEG signals, realize automatic feature selection, and improve classification accuracy.

References

1. Wolpaw, J.R., Birbaumer, N., Heetderks, W.J.: Brain-computer interface technology: a review of the first international meeting. Rehabil. Eng. IEEE Trans. **8**(2), 164–173 (2000)
2. Schwartz, A.B., Cui, X.T.: Brain-controlled interfaces: movement restoration with neural prosthetics. Neuron **52**(1), 205–220 (2006)
3. Curran, E.A., Stokes, M.J.: Learning to control brain activity: a review of the production and control of EEG components for driving brain-computer interface (BCI) systems. Brain Cogn. **51**(3), 326–336 (2003)
4. Ramoser, H., Muller-Gerking, J., Pfurtscheller, G.: Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans. Rehabil. Eng. A Publ. IEEE Eng. Med. Biol. Soc. **8**(4), 441–446 (2000)

5. Webb, A.R.: Introduction to Statistical Pattern Recognition. Academic Press, Cambridge (1990)
6. Mingai, L., Jingyu, L., Dongmei, H.: A method of motion imaging EEG signal recognition based on improved CSP algorithm. Chin. J. Biomed. Eng. **28**(2), 161–165 (2001)
7. Mcfarland, D.J., Anderson, C.W., Muller, K.R.: BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation. IEEE Trans. Neural Syst. Rehabil. Eng. **14**(2), 135–138 (2006)
8. Daubechies, I.: The wavelet transform, time-frequency localization and signal analysis. J. Renew. Sustain. Energy **36**(5), 961–1005 (1990)
9. Bashashati, A., Fatourechi, M., Ward, R.K.: BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation. J. Neural Eng. **4**(2), 32 (2007)
10. Li, L., Huang, S., Wu, X.: Feature extraction and classification of EEG signals based on motion imaging. Med. Equip. **32**(1), 16–17 (2011)
11. Li, Z., Xuhong, G.: Power spectrum estimation of EEG signals based on motion imaging. Electron. Meas. Technol. **35**(6), 81–83 (2012)
12. Vapnik, V.: Statistical learning theory. Ann. Inst. Stat. Math. **55**(2), 371–389 (2003)
13. Joachims, T.: Making large-scale support vector machine learning practical. In: Advances in Kernel Methods (1999)
14. Cui, Y., Ooi, B.C., Tan, K.L.: Indexing the distance: an efficient method to KNN processing. In: VLDB (2001)
15. Joshi, A.J., Papanikolopoulos, N.: Learning of moving cast shadows for dynamic environments. In: IEEE International Conference on Robotics and Automation. IEEE (2008)
16. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Ann. Hum. Genet. **7**(2), 179–188 (2012)
17. Izenman, A.J.: Linear discriminant analysis. In: Modern Multivariate Statistical Techniques (2013)
18. Power, S.D., Kushki, A., Chau, T.: Automatic single-trial discrimination of mental arithmetic, mental singing and the no-control state from prefrontal activity: toward a three-state NIRS-BCI. BMC Res. Notes **5**(1), 141 (2012)
19. Power, S.D., Kushki, A., Chau, T.: Intersession consistency of single-trial classification of the prefrontal response to mental arithmetic and the no-control state by NIRS. PLoS one **7**, e37791 (2012)
20. Blankertz, B., Dornhege, G., Krauledat, M.: The non-invasive Berlin Brain-Computer Interface: fast acquisition of effective performance in untrained subjects. Neuroimage **37**(2), 539–550 (2007)
21. Michel, C.M., Murray, M.M., Lantz, G.: EEG source imaging. Clin. Neurophysiol. **115**(10), 2195–2222 (2004)
22. Kai, K.A., Zheng, Y.C., Zhang, H.: Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: IEEE International Joint Conference on Neural Networks (2008)
23. Kai, K.A., Zheng, Y.C., Zhang, H.: Filter bank common spatial pattern (FBCSP) algorithm using online adaptive and semi-supervised learning. In: International Joint Conference on Neural Networks (2011)
24. Deng, X., Li, D., Mi, J.: Motor imagery ECoG signal classification using sparse representation with elastic net constraint. In: IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS), pp. 44–49 (2018)
25. Friedman, J., Hastie, T., Höfling, H., et al.: Pathwise coordinate optimization. Ann. Appl. Stat. **1**(2), 302–332 (2007)