Improving classification accuracy using fuzzy method for BCI signals

Yu Wei[a,b], Yang Jun[a,*], Sun Lin[b] and Li Hong[c]

[a]*Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650504, China*
[b]*Department of Electrical & Electronics Engineering, Chengdu Technological University, Chengdu 611730, China*
[c]*School of information science and engineering, Yunnan University, Kunming 650091, China*

Abstract. Electroencephalograph (EEG) signals feature extraction and processing is one of the most difficult and important part in the brain-computer interface (BCI) research field. EEG signals are generally unstable, complex and have low signal-noise ratio, which is difficult to be analyzed and processed. To solve this problem, this paper disassembles EEG signals with the empirical mode decomposition (EMD) algorithm, extracts the characteristic values of the major intrinsic mode function (IMF) components, and then classifies them with fuzzy C-means (FCM) method. Also, comparison research is done between the proposed method and several current EEG classification methods. Experimental results show that the classification accuracy based on the EEG signals of the second BCI competition in 2003 is up to 78%, which is superior to those of the comparative EEG classification methods.

Keywords: Brain-computer interface (BCI), intrinsic mode function (IMF), fuzzy C-means (FCM)

1. Introduction

Brain-Computer Interface (BCI) uses the specific pattern signal which is generated when human brain response to the outside stimulation to construct a connection with the computer or external mechanical equipment directly, without the human neuromuscular transmission. One of the main challenges of BCI research is the non-stationary property of the target electroencephalograph (EEG) signal, specifically the constantly changing of the statistics characteristics of EEG. In order to effectively classify, manage and control the EEG signal, it needs to transform the original data to get the characteristics that reflect the nature of the signals. That is EEG feature extraction, which is also one of the key technologies of BCI [1–5].

Nowadays, there are mainly four kinds of methods for feature extraction of BCI, which are as follows: the pure time-domain or frequency-domain information method, the traditional combination of the time-frequency characteristics method, the model parameters method, and the wavelet analysis method. These methods all have their own characteristics: the pure time-domain or frequency-domain

---

*Corresponding author: Yang Jun, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650504, China. Tel.: +86 18608710817; Fax: +86 87165916645; Email: cdtuyw@qq.com.

method is simple to use, but it has large limitations in practical occasions, such as single information metrics and low recognition rate; the traditional time-frequency characteristics combination method is stable for analysis, but it is difficult to extract the important time-domain information of the signals, and it also needs to analyze a huge amount of data, besides, the accuracy is relatively poor; the method of model parameters possesses high recognition rate of frequency, which is suitable for short-term data analysis, but the calculation is complicated, and there have high individual difference; the wavelet transform and wavelet package transform need to preset the hierarchy number and the wavelet function before decomposing the signals, and it does not have the adaptive decomposition ability [6]. The four methods all have their limitations that need some new methods to improve the accuracy of the EEG signal processing. This paper uses the empirical mode decomposition (EMD) algorithm, which is an effective, nonlinear and non-stationary data analysis method. The algorithm essence is stabilizing the signals, decomposing the signals to some intrinsic mode function (IMF) according to different characteristic scales, and each IMF has different frequency component [7].

There are many classification algorithms of BCI system have been used in the EEG classification. The references [8–10] have used support vector machine (SVM), linear discriminant analysis (LDA) and Multi-parameter linear programming (MLP) methods for EEG classification. However, all these classification algorithms are supervised algorithms, which need training a certain number of known data set beforehand, which brings difficulties for their practical application. Fuzzy C-means (FCM), an unsupervised clustering algorithm, based on partitioning clustering, according to the principle of least square, adopts the iteration method to optimize an objective function, and finally gets the membership grade of each sample. FCM does not need training before classifying when it is used as classification method and has strong robustness for the unstable biomedical signals. Therefore, this paper chooses FCM as the final classifier [11].

This paper adopts EMD algorithm to decompose the EEG library of the second international BCI competition in 2003, extracts the mean absolute value (MAV) of IMF as the characteristic values, and classifies with FCM method.

2. Data acquisition and classification algorithm

2.1. Data acquisition

The EEG dataset used in the paper comes from the Second International BCI Competition in 2003 (provided by the University of Graz technology). The subject in the experiment is a 25 years old healthy female. The subject imagines the movement of her left and right hands, and records the EEG signals using the double guide ways of the three electrodes which located in C3, Cz, C4. The detail experiment content and parameters can be found in reference [12].

2.2. Feature extraction

Since the EEG signals which are generated when human brain imagining the left and right hands movement is quite obvious in the electrodes of C3 and C4, this paper uses EMD method to decompose the EEG signal records of C3 and C4. EMD is a local time scale algorithm, which can decompose a complex signal into several IMFs and each IMF contains different frequency component. These frequency components are not only related to the analyzing frequency, but also to the signal itself. There-

fore, EMD method is an adaptive signal processing method which is very suitable for non-linear and non-stationary signal processing [13]. EMD can be summarized as follows:

- Step 1. Given a signal $x(t)$, find all local extrema of $x(t)$ by using cubic spline curve interpolation to connect, we can get maximum envelope $x_{max}(t)$ and minimum envelope $x_{min}(t)$ of original signal $x(t)$ separately.

- Step 2. Connect the mean of maximum envelope and minimum envelope in order, to obtain the mean $m_1(t)$.

- Step 3. $x(t)$ minus $m_1(t)$ being $h_1(t)$.

- Step 4. Check whether $h_1(t)$ is an IMF. If $h_1(t)$ is an IMF then set $c_1(t) = h_1(t)$, otherwise see $h_1(t)$ as the original signal, repeat above steps.

Figures 1 and 2 separately show the EMD decomposition of C3, C4 EEG records at the same time, which extract MAV of the major IMF components as the characteristic values for the classification.
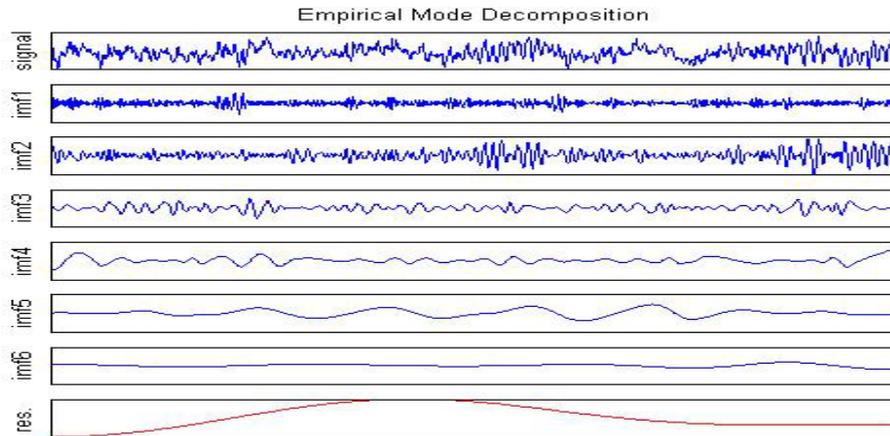

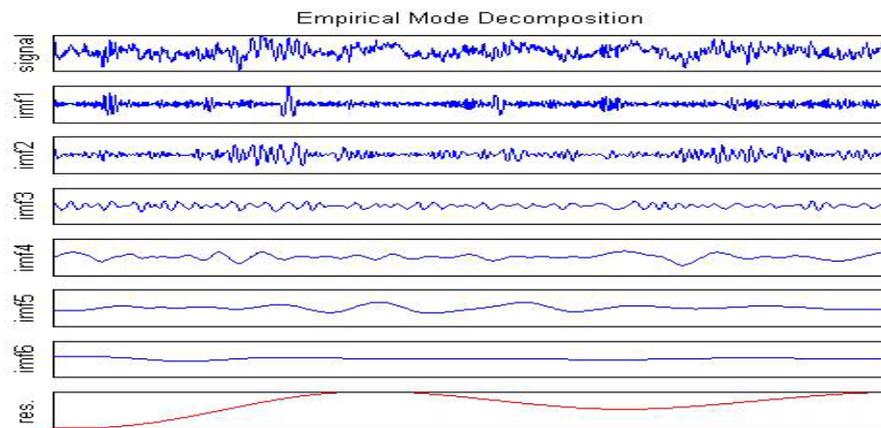
Fig. 1. EMD decomposition of C3 EEG signals.

Fig. 2. EMD decomposition of C4 EEG signals.

MAV has been applied as the characteristic values in the classification of EEG in reference [14], which has achieved good experiment effect. MAV is the average absolute value of each data sample point, as it shows in Eq. (1), in which $0 \le i \le N$, and x_i is the sample point in the segment data of j, $0 \le k \le N$.

$$MAV_k = \frac{1}{N} \sum_{i=1}^{N} |x_i|$$ (1)

2.3. Fuzzy C-means (FCM) clustering

FCM has been widely applied in the classification of biomedical signals and BCI [14]. The FCM algorithm divides the sample space $x = \{x_1, x_2, \ldots, x_n\}$ into category C ($2 \le C \le n$), in which x_i will not be divided to a certain category. The membership grade of a sample x_i to category j is u_{ij} ($0 \le u_{ij} \le 1$). The membership grade matrix of all samples is $U = \{u_{ij}\}$, which meets the requirements as Eqs. (2) and (3) [14].

$$\sum_{j=1}^{c} u_{ij} = 1$$ (2)

$$0 \le \sum_{j=1}^{c} u_{ij} \le n$$ (3)

FCM algorithm minimized the objective function J_{fcm} in the constraint condition of Eqs. (2) and (3), which is shown in Eq. (4), in which m (m>1) is the fuzzy weighted index. c_j is the center of the category j in C, $d_{ij}^2(x_i, c_j) = \|x_i - c_j\|$ is the Euclidean Distance of sample x_i to the clustering center c_j.

$$J_{fcm}(U,C) = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^m d_{ij}(x_i, c_j)$$ (4)

FCM minimizes the objective function with the iterative method, and then in the last, the optimal fuzzy partition of the sample $U = \{u_{ij}\}$ and the clustering center $C = \{c_j\}$ has been gained, as Eqs. (5) and (6).

$$c_i = \frac{\sum_{j=1}^{n} (u_{ij})^m x_j}{\sum_{j=1}^{n} (u_{ij})^m} \tag{5}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}}{d_{kj}})^{\frac{2}{m-1}}} \tag{6}$$

Table 1

A part of characteristic values of EEG

The first two IMF from:		C3 electrode		C4 electrode	
MAV value (Total 100 sets)	1	0.015782	0.029778	0.022194	0.027132
	2	0.023488	0.028498	0.024708	0.042413
	3	0.044938	0.039641	0.02378	0.029924
	4	0.021785	0.020993	0.026959	0.033329
	5	0.019608	0.023288	0.029296	0.041383
	6	…	…	…	…

3. Experimental results and analysis

In order to verify the effectiveness of the method used in the paper, this paper classify the EEG signal of imagining the left and right hands movement which is provided by the second international BCI competition in 2003. AS the ERD/ERS signals which is generated when human brain imagining the left and right hand movement is quite obvious in the electrodes of C3 and C4, this paper chooses total of 100 sets of data of C3 and C4 electrodes for EMD classification. Fourier transformation of the decomposed IMF component is used to get the spectrum energy, and then we find that the main energy is concentrated in the first two IMF components, which avoids the limitation of the traditional methods because these methods generally are based on experience to get the major components. Then the MAV of the first two IMF components are extracted as our characteristic values, Table 1 shows the IMF components characteristic values of the EEG data of the electrodes C3 and C4.

In the FCM classification procedure, the fuzzy weighted index m is set to2, the clustering number is 2, and the iteration stopping threshold is $\varepsilon = 10^{-5}$. After ten times iteration, the target function is 0.023383, and the center of the clustering C and the membership matrix U are as Eqs. (7) and (8).

$$C = \left\{ \begin{matrix} 0.030087 & 0.031648 & 0.029334 & 0.035247 \\ 0.054853 & 0.049291 & 0.040968 & 0.048083 \end{matrix} \right\} \tag{7}$$

$$U = \left\{ \begin{matrix} 0.93135 & 0.70383 & 0.92328 & \ldots 0.90618 & 0.1632 & 0.37816 \\ 0.068647 & 0.29617 & 0.076716 & \ldots 0.093818 & 0.8368 & 0.62184 \end{matrix} \right\} \tag{8}$$

Table 2

The evaluation indices of different methods

Number	Method	Classification Accuracy	Classification Sensitivity	Classification Specificity
1	MAV FCM	61%	73%	65%
2	MAV SVM	70%	75%	69%
3	EMD MAV SVM	80%	77%	67%
4	EMD MAV FCM (the proposed method)	78%	81%	76%

The classification accuracy is 78% according to the maximum membership principle. The comparison between the method used in this paper and the other three methods as follows:

– Extract the time-domain characteristics (mean absolute value, MAV) directly, and use FCM to classify the data.
– Use SVM method with Radio Basis Kernel Function (RBF) to classify the MAV data. In this method, choose the first 50 groups of data for SVM training, and use all the other data for the classification and employ the particle swarm optimization algorithm (PSO) for SVM parameter optimization.
– Use EMD to classify the data, extract the major IMF characteristic values and use SVM as the classifier.

The evaluation indices including classification accuracies, sensitivities and specificities of the above three methods compared with the proposed method are shown in Table 2. Firstly, the evaluation indices of the proposed method are obviously higher than it of the traditional time-domain characteristic and FCM methods (method 1). Secondly, the result of method 3 is higher than it of method 2, which proves the effectiveness of EMD algorithm. Finally, the result of the proposed method is close to it of the method 3. Considering the relatively longer execution (classification) time and the algorithm training cost of SVM, the proposed method has much more competitiveness. Besides, it is hard to get the data label in real EEG signal classification, so the method in this paper does have some advantages in EEG signal processing.

4. Conclusion

Because of the non-stationary, complex and low signal noise ratio characteristics of EEG signals, this paper uses the non-stationary, nonlinear EMD algorithm to decompose the signals, extracts the MAV of the first two IMF component which contains the main energy as the characteristic values, and applies the unsupervised FCM algorithm to classify the characteristics. In the last, a comparative study with the existing classification methods is given. The experiment result shows that the method proposed in this paper has higher classification accuracy for the EEG signal of the second BCI competition in 2003, which achieves the research purpose.

References

[1] G. Tapan, P.B. Ketan and A. Sneh, A comparative study of wavelet families for EEG signal classification, Neurocomputing **17** (2011), 3051–3057.

[2] P. Shenoy, M. Krauledat, B. Blankertz, R. Rao and K. Muller, Towards Adaptive Classification for BCI, Journal of Neural Engineering **1** (2006), R13–R23.

[3] D. Ian, J.N. Slawomir and W. Kevin, Brain computer interface control via functional connectivity dynamics, Pattern recognition **45** (2011), 2123–2136.

[4] B. Nicolas, L. Fabien and L. Anatole, Exploring two novel features for EEG-based brain-computer interfaces: Multifracal cumulants and predictive complexity, Neurocomputing **79** (2011), 87–94.

[5] Y.H. Wei, H.L. Chao, J.H. Hsien and H.C. Po, Wavelet-based envelope features with automatic EOG artifact removal: Application to single-trial EEG data, Expert Systems with Applications **39** (2011), 2743–2749.

[6] D.P. Mandic, Time-frequency analysis of eeg asymmetry using bivariate empirical mode decomposition, IEEE Transactions on Neural Systems And Rehabilitation Engineering **19** (2011), 366–373.

[7] P. Wang, J.Z. Shen and J.H. Shi, Feature extraction of eeg for imagery left-right hands movement, Journal of Transduction Technolog **9** (2010), 78–82.

[8] N. Shimo and S. Pang, Curiosity-driven multi-agent competitive and cooperative LDA learning, International Journal of Innovative Computing, Information and Control **7** (2008), 1537–1552.

[9] H.B. Zhu and H.M. Kai, Application of bpnn in classification of time intervals for intelligent intrusion detection decision response system, International Journal of Innovation Computing, Information and Control **10** (2008), 2483–2491.

[10] B.H. Chen and J.L. Hu, An improved multi-label classification method based on SVM with delicate decision boundary, International Journal of Innovation Computing, Information and Control **4** (2010), 1605–1614.

[11] W.Y. Hsu and C.Y. Lin, Unsupervised fuzzy c-means clustering for motor image EEG recognition, International Journal of Innovation Computing, Information and Control **8** (2011), 4965–4975.

[12] C. Liu, H.B. Zhao and C.S. Li, CSP/SVM-based EEG classification of imagined hand movements, Journal of Northeastern University (Natural Science) **8** (2010), 1098–1101.

[13] C. Park, D. Looney, P. Kidmose, M. Ungstrup and D.P. Mandic, Time-frequency analysis of eeg asymmetry using bivariate empirical mode decomposition, IEEE Transactions on Neural Systems and Rehabilitation Engineering **19** (2011), 366–373.

[14] A. Khorshidtalab, M.J.E. Salami and M. Hamedi, Evaluation of time-domain features for motor image movements using FCM and SVM, Computer Science and Software Engineering **6** (2012), 17–22.