Acta Technica 62 (2017), No. 5A, 183–190

Research on feature extraction method of multi-channel EEG based on improved distance criterion

Yang Rong[1,2,3,4], Chen Wei[1,2,3], Shan Xinying[1,2,3],Zhang Ning[1,2,3],Bi Sheng[1]

Abstract. Sincethe brain computer interface system based on multi-channel EEG signals have some faults, such as complex data, low recognition rate and poor self-adaptive ability, etc., Hilbert-Huang transform was applied to analyze the frequency feature and extract the energy of EEG signal as the feature vector. In the meantime, an improved distance criterion was introduced as the evaluation criterion to obtain the channels corresponding to the optimal eigenvalue. The features of the signals corresponding to the optimal channels were used as the feature to recognize the different action pattern by Support Vector Machine algorithm. The experiment result has shown that the recognition based on the designed method can achieve above 90% accuracy. The improved method has effectively improved the faults of large data, long processing time, individual differences and so on.

Key words. Multi-channel eeg, hilbert-huang transform, distance criterion, optimal selection and support vector machine.

1. Introduction

Electroencephalogram (EEG) is a kind of signal that contains abundant physiological information which can reflect different thinking activities. How to extract the features that reflect specific mind work precisely and quickly is the key point of the application of EEG signals research.

Aiming at the time-varying, unstable and strong individual difference characteristics in EEG signals, researchers in China and abroad have done a lot of work

[1]Workshop 1 - National Research Center for Rehabilitation Technical Aids, Beijing 100176, China;email:chenwei@nrcrta.cn

[2]Workshop 2 - Beijing Key Laboratory of Rehabilitation Technical Aids for Old-Age Disability, Beijing 100176, China;email:shanxinying@nrcrta.cn

[3]Workshop 3 - Key Laboratory of Intelligent Control and Rehabilitation Technology of the Ministry of Civil Affair, Beijing 100176, China;email:zhangning@nrcrta.cn

[4]Corresponding author: Yang Rong ;email:yangrong@nrcrta.cn

on the feature extraction methods. At present, the main methods used in feature extraction are Wavelet Packet Transform (WPT)[1],

Common Spatial Pattern (CSP)[2], Common Spatial Subspace (CSSD)[3], and Hilbert-Huang Transform (HHT)[4]. WPT is not applicable to large data processing for the complex analysis process and long processing time because it decomposes the high and low frequency signals at the meantime during the decomposition of every grade. CSP with high individual differences is much easier to be effective by frequency filtering, unstable EEG signal and some other factors because it only considers the maximum separability of the projections which the two class tasks projects in the space. CSSD is to extract signal compositions of one certain specific task from multi-channel EEG data at a condition of multi-task, which means low stability of the constructed features, and poor separability HHT, the basic method used in this paper, is a transform method of local domain wave which has higher time-frequency resolution. The core of HHT, which is called Hilbert transform translates signals from time-domain to time-domain to preserve the dynamic characteristic of the signals maximally. Though HHT is very suitable for the analysis of EEG signals, in the feature extraction of multi-channel EEG signals, HHT has the faults of large operation load and long operation time. Therefore, how to choose the most valuable channels is the research focus of this paper.

It has been proved that the recognition result when adopting 64-channel EEG signals as the feature source is not superior to the one when adopting part-channel EEG signals, as EEG signal on each different channel has different reflection to one certain task. Therefore in this paper, all the experiment data are separated into two parts, training data and testing data. To get the most valuable channels, HHT is used to transform and extract features of all channels in the training data, and a designed distance criterion is used to get numbers of the much valuable channels. Then the EEG signals corresponding to those channels in the testing data are processed by HHT algorithm and designed classifier based SVM algorithm to realize the pattern recognition of EEG signals. The experiment results show that the designed method successfully solves the problems of strong individual difference, large operation load and long operation time.

2. HHT algorithm

HH is a new kind of processing algorithm for unsteady signal proposed by N.E.Huang in 1998. It is consisted by two steps: Empirical Mode Decomposition (EMD) and Hilbert transform. EMD is the core of HHT. It separates the signal into a finite number of Intrinsic Mode Function (IMF) with different features scale based on the internal characteristics of the signal. The Hilbert spectrum constructed by the instantaneous frequency of the IMFs solved by the Hilbert transform can reflect the energy distribution of the signal in time and frequency accurately. HHT algorithm, which has higher time-frequency resolution, is very suitable for the process of EEG signals.

2.1. EMD

IMF is derived from data itself and has multi-dimensional and self-adaptive features.

The IMFs decomposed by EMD must satisfy two characteristics to make the instantaneous frequency of EEG signal stable and meaningful [5].

(1) IMFs must satisfy the definition of narrow zone signal for Gaussian distribution: the number of all the local extremums and the zero-crossings must be equal or the difference between them is 1 at the most.

(2) IMFs must satisfy the definition of the local symmetric: the mean value determined by the local extremums must be zero at any time.

The EMD is ended until the remaining component becomes a monotone function or less than a predetermined value. By analysis of EMD, the raw EEG can be expressed as

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{1}$$

where $\sum_{i=1}^{n} c_i(t)$ is the sum of IMFs and $r_n(t)$ is the remaining component.

2.2. Hilbert transform

The Hilbert transformation of the IMFs is given as

$$y_i(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{c_i(\tau)}{t-\tau} \tag{2}$$

With the instantaneous amplitude

$$a_i(t) = \sqrt{y_i\,(t)^2 + c_i\,(t)^2} \tag{3}$$

and the instantaneous frequency

$$\omega_i(t) = \frac{d}{dt}\left(\arctan \frac{y_i(t)}{c_i(t)}\right) \tag{4}$$

the Hilbert spectrum in which the amplitude is changed according to time and instantaneous frequency, is given as

$$H(\omega, t) = \sum_{i=1}^{n} a_i(t) \exp(j \int \omega_i(t) dt) \tag{5}$$

To get the marginal spectrum, integrations are performed on the Hilbert spectrum due to the calculation of

$$h(\omega) = \int_0^T H(\omega, \mathrm{t}) dt \tag{6}$$

The marginal spectrum reflects the amplitude accumulation on each and every frequency point, which is also the total energy of each point.

Pfurtscheller G. *etal.*analyzed the imagination of left and right hand movement. They found that when the examiners were required to imagine the unilateral limb movement, the energy ofμwave8-12HZ was inhibited in cerebral motor sensing area of the contralateral limb. This phenomenon was called Event-Related Desynchronization (ERD). The simultaneous enhancement of the energy of μ wave in cerebral motor sensing area of the ipsilateral limb was called Event-Related Synchronization (ERS). Therefore, we choose the total frequency band energy between 8 and 12Hz in the EEG as the ultimate power density of the lead.

3. Feature extracion

3.1. Distance criterion

According to traditional distance criterion, each sample can be regarded as one point in the feature space. The distance between each sample is a relatively visualized criterion to analyze the sample divisibility. The larger the distance, the better the divisibility. Thus, when calculating the contribution of lead, we usually define the power density difference between different tasks with the same bandwidth.

$$h(k) = \frac{\left|\sum_{f\in f_{\sum}} P_{1,k}(f) - \sum_{f\in f_{\sum}} P_{2,k}(f)\right|}{\left|\sum_{f\in f_{\sum}} P_{1,k}(f) + \sum_{f\in f_{\sum}} P_{2,k}(f)\right|} \tag{7}$$

In equation (7), $P_{i,k}(f)$ stands for power desnsity of ith task at the kth lead, f for frequency and f for the group of all frequency points at given bandwidth. $f_{\sum}$ represents the collection of all the frequency points in the chosen frequency band. $i = 12$

From the equation above, $h(k) \in [0,1]$ and the larger the h(k), the larger the power density difference between two tasks at certain lead, i.e. the lead makes larger contribution to task recognition. It is simple yet straightforward to use distance criterion as the evaluation function. However, it does not satisfy the required additivity equation which is $M(0) = 0$ for evaluation function M.

Hence, in light of the calculation theory of entropy, we incorporate a weight factor $\alpha \in [0,1]$ into the evaluation criterion to represent concern level of different tasks for one lead.

$$h(k) = \alpha * \sum_{f\in f_{\sum}} P_{1,k}(f) - (1-\alpha) \sum_{f\in f_{\sum}} P_{2,k}(f) \tag{8}$$

Equation (8) shows the newly proposed evaluation criterion which satisfies the additivity requirement. The larger the evaluation criterion $h(k)$calculated using the equation above, the better the corresponding sample divisibility. Value of the weight factor αwill influence the classification result.

3.2. Optimal lead selection

EEG data of different channels possess effective information with different weights. Assuming there are Pchannels, marked asCH_1??$CH_2 \ldots\ldots CH_P$, m times experiment results are selected as training samples and the corresponding EEG channel lead values form the number group $N = (n_1, n_2, ...n_p)$. First, Pchannels of training samples are transformed individually using HHT and the sum of energy values between 8 and 12Hz is marked as eigenvalue set $H = (h(n_1)h(n_2), ...h(n_p))$. Then according to distance criterion, contribution values of P channels are calculated and placed in descending order. The newly derived lead value array is marked as N'. Finally, top i leads with the largest contribution value in N' are selected and marked as $N" = (n"_1, n"_2, ...n"_i)$.

3.3. Feature vector construction

Considering the large computation and high extracted eigenvalue dimension when manipulating multi-channel EEG signals using HHT algorithm, we choose i leads based on improved distance criterion. After HHT transformation of selected test samples, we get the characteristics vector set marked as
$H" = (h"(n"_1)h"(n"_2), ...h"(n"_i))$.

4. Recognition based on SVM

In this experiment SVM[6]-[7] is adopted to recognize and classify the feature of EEG signals extracted by HHT. SVM is a kind of new machine learning method based on statistical learning theory proposed by Vapnik. Through nonlinear mapping, the sample space is mapped into a high-dimensional or an infinite-dimensional feature space to transfer the nonlinear separable problem into a linear separable problem. Based on structural risk minimization rules, SVM has the optimal classification ability and extended ability which is very suitable for the classification of small sample. An SVM classifier based on the Radial Basis Function (RBF) is applied in this paper [8]. The RBF is shown as

$$K \parallel x, y \parallel) = \exp\left\{- \parallel x - y \parallel^2 / (2 * \sigma)^2\right\} \tag{9}$$

where$K \parallel x, y \parallel)$ is the monotonic function of the Euclidean distance between x and y, and σ is the width parameter of the function.

5. Experiment results

5.1. Experiment data

The EEG data used to test the new feature extraction method in this paper comes from BCI Competition ?Data set motor imagery in ECoG recordings, session-to session transfer,provided by University of Tbingen,et al. Two kinds of imagery tasks was executed in this experiment, left little finger movement imagery and tongue movement imagery. EEG signal was recorded by a 88 electrode array implanted in the motion cortex of the right cerebral. The EEG data was sampled at 1000Hz and was filtered in a 0.016-300Hz frequency band.

Each trial was lasted for 7s. The sequence of left little finger and tongue trials, as well as the duration of the breaks between consecutive trials was randomized. Each trail started with the presentation of a fixation cross at the center of the monitor at 0s. Depending on the figure shown in the monitor from 1s to 1.5s, the subject was instructed to imagine a movement of the left little finger or the tongue. From 5s to 7s, when a white screen was shown in the monitor, this kind of imagery task was finished and the subject could get some rest.

To avoid the influence of Visual Evoked Potential, data from 1.5s to 4.5s in the whole competition dataset are captured to do the off-line analysis. The whole competition dataset comprised 278 experimental trials with fixed labels (139 left little finger and 139 tongue trials).Test data is acquired one week after the training data. Thus, the entire data group is more favorable to examine the self-adaption and robustness of the recognition method.

5.2. Analysis

To validate the effectiveness of the algorithm, we manipulate and categorize 278 times EEG data, of which 100 groups are selected as training samples to get the optimal EEG channel while the other 178 groups are selected as test samples to do pattern recognition. The recognition result is listed in Table 1.

Table 1. Recognition of tested samples

Algorithm		Feature Channels	Samples	Recognition Rate
All-channel & HHT		64	178	55.8%
Traditional distance criterion & HHT		10	178	78.5%
Improved distance criterion & HHT	α=0.5	10	178	82.8%
	α=0.7	10	178	92.1%
	α=0.9	10	178	87.5%

From Table 1, the recognition rate is significantly higher when using optimal-channel EEG data than all-channel data. Moreover, improved distance criterion eclipses the traditional one in higher divisibility, which is favorable of constructing characteristic space with higher divisibility. Additionally, as the weight factor a influence the categorization correctness, comprehensive consideration of its value should be taken to get the best categorization result.

Using the proposed process method, different examiners acquire different optimal channel as EEG signal has individual differences. Therefore, this method is capable of choosing the optimal EEG channel for each individual automatically.

6. Conclusions

Our work focuses on tackling problems during multi-channel EEG signal processing, namely difficulty in selection of effective channel and large computation. Considering the time-varying, non-stable and heterogeneous features of EEG signal, we improve the traditional distance criterion and combineit with HHT algorithm. Distance criterion with the incorporation of weight factor is used as the performance evaluation function for the optimal channel. The proposed method is capable of extracting individual features and more abundant signal power information. Not only does it increase the feature extraction efficiency and categorization correctness, it also solve the heterogeneous problem. Thus, the method offers a more efficient way for the processing and application of EEG signal.

References

[1] M. Li, L. Lin, J. Yang: *Adaptive feature extraction based on best basis of wavelet packet for motor imagery EEG.* Computer measurement *19* (2011), No. 11, 2755–2762.

[2] M. Li, S. Guo, J. Yang, Y. Sun: *A novel EEG feature extraction method based on OEMD and CSP algorithm.* Journal of intelligent *30* (2016), No. 5, 2971–2983.

[3] F. Mertens, A. Lobanov: *Wavelet-based decomposition and analysis of structural patterns in astronomical images.* Astronomy *574* (2015), No. 10, 10782–10789.

[4] Y. Wang, H. Li, P. Ye: *Fault Feature Extraction of Hydraulic Pump Based on CNC De-noising and HHT.* Journal of failure analysis *15* (2015), No. 1, 139–151.

[5] K. Zheng, Y. Zhang, C. Zhao, L. Liu: *Rotary kiln cylinder deformation measurement and feature extraction based on EMD method.* Engineering letters *23* (2015), No. 4, 283–291.

[6] J. S. Paiva, J. Cardoso, T. Pereira: *Supervised learning methods for pathological arterial pulse wave differentiation: A SVM and neural networks approach.* International journal of medical informatics *109* (2018) 30–38.

[7] S. Dong, J. Sheng, Z. Liu, L. Zhong, H. Wei: *Bearing remain life prediction based on weighted complex SVM models.* Journal of Vibroengineering *18* (2016), No. 6, 3636–3653.

[8] T. Rajesh, R. S. M. Malar: *robust classification of brain tumor in mri images using salient structure descriptor and rbf kernel-svm.* Asian journal of research in social sciences and humanities *7* (2017), No. 2, 58–71.

Received November 16, 2017