Feature Combination and Correlation Analysis for Motor Imagery EEG

Mingai Li [a], Jinfeng Xu[b] and Xinyong Luo

College of Electronic Information &Control Engineering, Beijing University of Technology, Beijing, China

[a]limingai@bjut.edu.cn, [b]u09020126@126.com,

Keywords: motor imagery EEG, Sample Entropy, orthogonal empirical mode decomposition, correlation coefficient，feature combination

Abstract: In order to get high classification accuracy, feature combination is commonly used in analysis of motor imagery electroencephalography EEG signal, including the nonlinear analysis and traditional time-frequency analysis. In this paper, Sample entropy(SampEn) was computed and represented as the nonlinear feature of motor imagery EEG signal for it can quantify the probability of new information appeared in time series. In addition, orthogonal empirical mode decomposition (OEMD) was also employed to extract the average energy of selected intrinsic mode functions(IMF) as the time-frequency feature for motor imagery EEG signal. Based on a public dataset, many experiments were conducted. Slide window was used to select the best time period for a better performance in feature extraction, and cross validation of 10 folds was applied in all the classification procedure. The highest recognition rate using SampEn and OEMD is respectively 86.07% and 83.21% classified by incremental support vector machine (ISVM) respectively. However, the highest classification rate of combined features is 86.79% by using ISVM which is a little higher than that of SampEn. Big linear correlation between SampEn and energy of IMF explains why the classification accuracy by combining the two types of features is not as high as expected.

Introduction

With the development of robot and rehabilitation medicine, rehabilitation robot has become a new technique of motor nerve rehabilitation therapy. The benefits of rehabilitation robotics include introducing the ability to perform precise and repeatable therapeutic exercises, reduction of the physical burden of participating therapists, incorporation of interactive virtual reality systems and so on [1]. The major training modes of rehabilitation robot are passive training and initiative training. From the effect of different exercise modes, it can be concluded that initiative training performs better than the passive one due to the active participation of patients in the former training. Besides, neural circuits can be fully stimulated through the patients' active participation [2]. Surface electromyography (sEMG) and motor imagery EEG [3] are commonly used in the active trianing process with rehabilitation robots to obtain the user's motion pattern information. However，it is difficult to extract the patient's intention precisely and reliably for the extracted EEG is usually weaker than sEMG. As a result, motor imagery EEG has been widely studied by many researchers in order to find better method to recognize distinct motor imagery patterns.

EEG signals are nonlinear time series, which are generally very noisy, nonstationary, and easily contaminated by artifacts that can deteriorate classification methods. As a result, some nonlinear dynamic analysis methods are applied in feature extraction for EEG such as approximate entropy (ApEn), sample entropy (SampEn), phase space reconstruction and so on[4,5]. Sample entropy is a nonlinear analysis method which could estimate the occurrence probability of new information so that the complexity of EEG is computed [6]. As the nonlinear feature, SampEn is combined with traditonal feature extracted using HHT to get a higher correct recognition rate with better stability for motor imagination EEG [7].

Considering the problems of feature fusion and feature dimension reduction in related researches, statistical performance including auto-correlation coefficients and mutual information is introduced in feature evaluation rules. In paper [8], the feature fusion algorithm designed by using correlation coefficients performs better than the existing ones. In other related research[9], correlation coefficient

is used to reflect the influences of choosing wavelet packet generating function and to select the optimal classification frequency of electrooculogram.

Based on the above research, SampEn is calculated as the nonlinear feature of EEG which matches the physiological ERD/ERS (Event-related desynchronization/Event-related synchronization) phenomenon found by Professor Pfurtscheller [10]. The average energy feature of IMF is also extracted as traditional feature to combine with SampEn. The linear correlation between SampEn and average energy of IMF is put forward to find the reason why the correct rate of classification with combined features in different domain is not as high as expected.

Method

Sample Entropy

The physics meaning of SampEn is very clear. The larger the value of SampEn is, the greater the occurring probability of the new patterns is and the more complex the corresponding time series are.

Given a discrete time series $u(i) = \{u(1), u(1) \ldots u(N)\}$ of length N, SampEn is computed following the steps:

(1) Refactoring m dimension vector for the given time series:

$$X_i^m = \{u(i), u(i+1) \ldots u(i+m-1)\}, 1 \le i \le N-m+1 \tag{1}$$

(2) Defined maximum distance d_{ij}^m between vector X_i^m and vector X_j^m

$$d_{ij}^m = d\left[X_i^m, X_j^m\right] = \max |u(i+k) - u(j+k)|, 0 \le k \le m-1 \tag{2}$$

(3) Calculate the number of the distance between adjacent vectors in reconstructed phase space which is less than threshold r.

$$B_r^m(i) = \frac{1}{N-m-1} \sum_{j=1, j \ne i}^{N-m} H\left(d_{ij}^m - r\right) \tag{3}$$

$$B_r^m = 1/(N-m) \sum_{i=1}^{N-m} B_r^m(i) \tag{4}$$

In the above formula, $H(z)$ represents Heaviside function as followed,

$$H(z) = \begin{cases} 1, z \le 0 \\ 0, z > 0 \end{cases} \tag{5}$$

(4) Vectors of m+1 dimension $X_i^{m+1} = \{u(i), u(i+1) \ldots u(i+m)\}$ could also be reconstructed with the same method, meanwhile the formula A_r^m is also defined as B_r^m.

(5) For a limited time sequence, the sample entropy is estimated with formula (6):

$$SampEn(m, r, N) = -\ln\left(A_r^m \middle/ B_r^m\right) \tag{6}$$

Through the above steps, SampEn is computed for a time series of length N, typically the embedded dimension m=2, the threshold r =0.1~0.2*$SD(u)$,SD means the standard deviation of time series.

Orthogonal Empirical Mode Decomposition

Finite numbers of intrinsic mode functions (IMF) can be obtained through Empirical mode decomposition (EMD) by self-adaptive filtering to a signal. However, when EMD is applied to a signal, the problem of mode aliasing will occur in IMFs. Since EMD can select the high frequency component at any moment, when there is abnormity in the signal, it is not able to sift it out because it is not an orthogonal decomposition method in some degree. The method of orthogonal empirical mode decomposition (OEMD) is proposed to conquer mode aliasing and avoid the occurrence of false mode [11]. The main purpose of this method is to obtain IMF and the residual function by auto-adaptive band- pass filtering. The average energy of each IMF obtained through OEMD is computed as Eq. (7):

$$OEIMF = \frac{1}{n} \sum_{i=1}^{n} |u(i)|^2 \tag{7}$$

In formula (7), $u(i)$ with $i \in [1, n]$ means the time series of one IMF, OEIMF represents the average energy of IMF which obtained by OEMD.

Correlation coefficient

Feature distribution is usually not mutually independent. Some of the features can even be transformed to each other by way of linear or nonlinear methods[8]. Correlation coefficient can be used to measure the degree of linear correlation in order to get a better performance in feature combination.

Suppose that x_i and y_i are the random series with length n, and their correlation coefficient $\rho_{x,y}$ is defined as followings:

$$r_{x,y} = \frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 \bullet \sum(y_i - \overline{y})^2}} \tag{8}$$

The value of $\rho_{x,y}$ is within [-1, 1]. When $\rho_{x,y} = 0$, x_i is not related with y_i; if $|\rho_{x,y}| = 1$, the two series are complete related and the linear relationship is existed between them. When $|\rho_{x,y}| < 1$, part values of x may changed with y, and the larger the correlation coefficient is, the greater the change in x with y. $|\rho_{x,y}| > 0.8$ is known as highly relevant，$|\rho_{x,y}| < 0.3$ is called low correlation，other value means moderate correlation. Correlation coefficients are commonly used to characterize the linear correlation degree of random variables. The greater the correlation coefficient is, the better the linear correlation is whereas the worse the linear correlation degree is.

Feature extraction and classification

Data description

EEG dataset used in this paper is derived from the BCI competition2003 which provided by Institute for Biomedical Engineering, University of Technology Graz. It was sampled with 128Hz and filtered between 0.5Hz and 30Hz. There are 140 training trails and 140 test trails in the dataset, eachtrail lasts for 9s. Three electrodes marked with C3, Cz and C4 were used to record EEG signal.

Feature extraction and display

Motor related potential are mainly distributed in the central region (brain electrode C3, C4 and Cz). The ERD(Event-related Desynchronization) and ERS(Event-related Synchronization) phenomenon found by Professor Pfurtscheller exist in original EEG. SampEn(r=0.2, m=2) is calculated for each channel using 1s slide window moves one point ahead. The values of SampEn reflect the ERD/ERS phenomenon which is distinct from sample point 451 to 800 as shown in Figure.1. Since then, the 350 points of original EEG signal were selected for computing SampEn(N=350) and OEIMF feature.

(a) SampEn of EEG for imaging left hand movement   (b) SampEn of EEG for imaging right hand movement
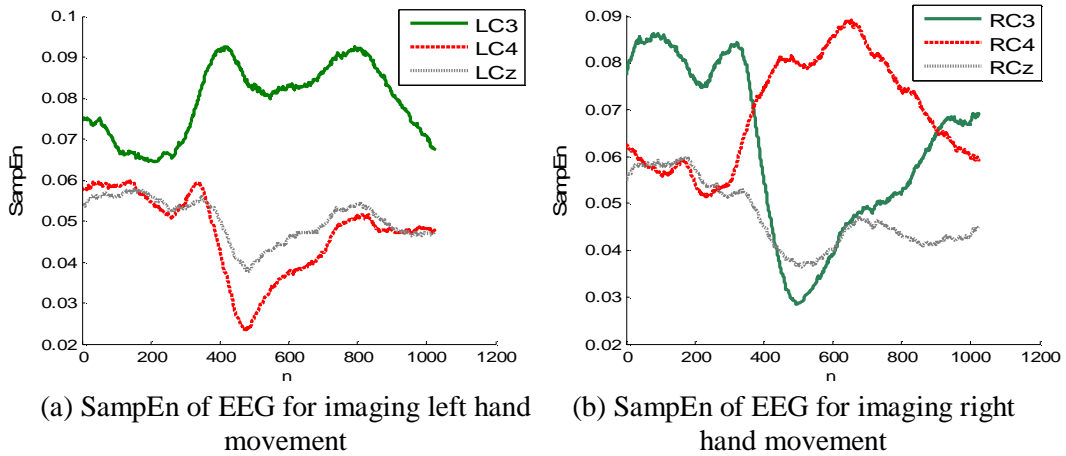
Figure.1. SampEn extracted from EEG signal using 1s slide window.

OEIMF is calculated for IMF of each channel which has the largest correlation coefficient with original EEG using formula (7). The average values of channel C3 and C4 are apparently different from each other which can be seen from Figure.2. The average value of C3 channel is higher than C4 when imaging left hand moving which is corresponding to ERS and ERD phenomenon.

Classification and correlation analysis

In this paper, classifier Incremental Support Vector Machine (ISVM) and back-propagation (BP)

neural network are independently used to testify the validity of the features SampEn and OEIMF. The dataset with 280 trials is divided randomly into ten subsets for 10-fold cross validation. Each subset has 28 trials including both left hand and right hand motor imagination tasks. Nine of the subsets are used to training the classifier and the remainder one is for test.

The recognition rates of two different motor imaging tasks are listed in table 1. The classification accuracy of SampEn and OEIFM with classifier ISVM are both higher than 82.86% in paper[12] with method of CSP and SVM and the mean value of 83.33% for left hand, 87.50% for right hand simply using SampEn[7]. But from table 1, it is obviously that the accuracy of feature fusion is just a little higher than that of SampEn even if the accuracy of OEIMF is 83.21%. However, in paper [7], the accuracy of combining HHT and SampEn is apparently higher than any of the single feature.
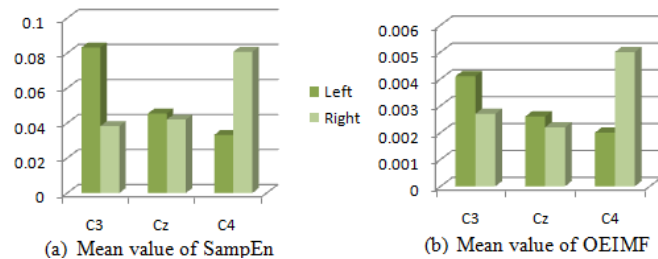

Figure.2. Feature display of SampEn and OEIMF. The mean values of two charactersare got by averaging 140 groups of left-right motor imagination.

Table 1 ISVM and BP classification accuracy for different features(%)

Feature	Classifier	
	ISVM	BP
SampEn	86.07	86.07
OEIMF	83.21	81.07
Fusion*	86.79	86.43

*Fusion means the feature combination of SampEn and OEIMF.

In order to find the reason why the recognition accuracy of feature fusion is not as high as imagined in this study, the correlation between two kinds of features is analyzed with correlation coefficient.

After computing SampEn and OEIMF features for respective 140 trials of left and right hand motor imagination, each channel has a value for SampEn or OEIMF for one trial. In order to calculate the correlation coefficient between SampEn and OEIMF, x_i time series in formula (8) is composed of the SampEn feature of one channel with the result of 140 trial (n=140) , y_i is composed of OEIMF feature in the same way.

Owing to the similarity characteristics of SampEn and OEIMF, the linear correlation coefficients of each channel were calculated for the two types of features using formula (8). Thus referring to the table 2, correlation coefficients shows that for the same channel, the correlation coefficient between the two features is higher than 0.8. The highest value 0.9868 is already very close to 1, that means an obvious linear correlation exists. As a result, it may explain why the classification accuracy of combining two kinds of characteristics is not apparently higher than before even the two kinds of features both have a good performance independently.

Table 2 Correlation coefficient ρ_{xy} between SampEn and OEIMF.

SampEn	OEIMF feature of each channel					
	LC3	LCz	LC4	RC3	RCz	RC4
LC3	**0.9868***	0.7733	0.4809	0.0070	0.0034	-0.0056
LCz	0.7454	**0.8826**	0.6646	0.1712	0.1514	0.0486
LC4	0.6202	0.7470	**0.8042**	0.1506	0.2034	0.1053
RC3	-0.0427	0.1057	0.2627	**0.8805**	0.7931	0.5865
RCz	-0.0037	0.1805	0.3113	0.8145	**0.9141**	0.7016
RC4	-0.0482	0.0031	0.8042	0.6620	0.7672	**0.9067**

*The values marked with bold and blue color in the above table are the largest correlation coefficients of corresponding channels calculated between two kinds of features. The same channel of identical motor imagination has a high correlation coefficient between two features while of different motor imagination the value is very small

Conclusions

In EEG analysis, sample entropy as a nonlinear characteristic can be used to measure the complexity of EEG and make up for the inadequacy of the time-frequency domain analysis of EEG signal.

Sample entropy can be applied to real-time brain computer interface for its fast computation. The new feature OEIMF of EEG also has its certain feasibility compare with SampEn. Essentially, the traditional time and frequency domain analysis methods such as HHT, wavelet transform, OEMD still have irreplaceable cause they can reflect the essential character of EEG. Therefore, the fusion of different characteristics for EEG is inevitable in the future. However, the research of the linear correlation coefficient cannot be ignored when combining different features because the significant correlation may exist between different categories of features. The analysis of linear and nonlinear correlation analysis offer an idea for optimizing selection of features to improve recognition rate with combined features.

To sum up, enriching feature extraction method constantly, analyzing the performance of combined features, selecting and optimizing the character of classifier are still the most important research area of EEG analysis which would promote development and the application of EEG signal in the field of rehabilitation

Acknowledgements

References

[1] Schabowsky, Christopher N, et al. Development and pilot testing of HEXORR: Hand EXOskeleton Rehabilitation Robot. *Journal of Neuroengineering & Rehabilitation* 7.3(2010):36.

[2] Liang Peng., Zeng-guang Hou. et al. Synchronous Active Interaction Control and Its Implementation for a Rehabilitation Robot. *ACTA Automation Sinic* ,vol.41.11(2015).

[3] Bao-guo Xu., Si Peng. et al. Upper-limb Rehabilitation Robot Based on Motor Imagery EEG. *RO BOT*,vol(33).3(2011):307-313.

[4] Wang, L., Xu, G., et al. (2012). Motor Imagery BCI Research Based on Sample Entropy and SV M.*Electromagnetic Field Problems and Applications (ICEF), 2012 Sixth International Conference o n* (pp.1 - 4). IEEE.

[5] Fang, Y., Chen, M. et al. (2012). Feature Extraction of Motor Imagery in BCI with Approximate Entropy. *Journal of Computational Information Systems*, vol. 8: 2485- 2491

[6] Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology Heart & Circulatory Physiology, 278* (6), H2039-H2049.

[7] Lin HB, Gong L. et al. (2015).Feature extraction of EEG signal based on improved HHT and sample entropy. *Computer Engineering and Design*(06), 1608-1613.

[8] Shi C, & Zhang M. (2014).Image steganalysis based on feature fusion by boosting feature selection algorithm . *Journalof Optoelectronics Laser* 03).

[9] Gao CF, Lv BL, & Ma JX. (2012). A new method of extracting vigilant feature from electroocul ography using wavelet packet transform. *Chinese Journal of Biomedical Engineering, 31*(05), 641-6 48.

[10] Pfurtscheller, G. & Silva, F. H. L. D. (1999). Event-related eeg/meg synchronization and desync hronization: basic principles.. *Clinical Neurophysiology, 110*(11), 1842–1857.

[11] Qin, Y., Qin, S., & Mao, Y. (2008). Fast implementation of orthogonal empirical mode decomposition and its application into harmonic detection. *Chinese Journal of Mechanical Engineering, 21*(2), 93-98.

[12] Liu, C., Zhao., et al. (2010). Classification of motor imagination based on CSP and SVM. *Journal of northeastern university,* natural science edition, 31(08), 1098-1101.