2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2018)

Classification of motor imagery based on multi-decision fusion for brain computer interface

Yifan Cheng, Ning Gao, Zhaohua Lu and Qi Li *
School of Computer Science and Technology
Changchun University of Science and Technology
Changchun, China
*:liqi@cust.edu.cn

Abstract—**Brain-computer interface (BCI) based on motor imagery (MI) is considered to be a promising cognitive tool for rehabilitation therapy of movement disorders. Feature extraction and classification are important in MI-BCI, and affect the BCI's performance. Label is the classifying result of motor imagery in MI-BCI. For example, 1 represents imaging left hand movement and 2 represents imaging right hand movement in a MI-BCI system. In this paper, we combined labels from different feature extraction and classification, and fused labels to get a new label which has higher accuracy. We named the method multi-decision fusion. We used the multi-decision fusion on BCI competition's MI datasets for classification. By comparing the results with those using conventional methods of feature extraction and classification, we have verified multi-decision fusion is an effective method. Multi-decision fusion can effectively improve classification accuracy of motor imagery.**

Keywords- motor imagery (MI), brain-computer interface (BCI), feature extraction, classification, multi-decision fusion

I. Introduction

Brain-computer interface (BCI) is a direct communication pathway between brain and an external device. BCI system collects user's electroencephalography (EEG) via electrical equipment, uses these signals to analyze the intention of user, and converts user's intention to actual commands. Many studies showed that motor imagery based BCI (MI-BCI) system can help stroke patients and disabled patients regain their motor ability, and utilizes patient's intention to control machine to complete basic daily activities.

MI-BCI enables user think about moving their arms, hands, legs, tongue, and use these imagery to communicate with computers[1-2]. MI-BCI system consists of data collection, feature extraction and classification. Feature extraction and classification are very important, and they directly affect the accuracy of MI-BCI system. At present, the main methods of feature extraction are common spatial pattern (CSP), auto-regressive (AR), power spectral density (PSD) and sample entropy (SampEn) [3-6]. Main methods of classification are support vector machine (SVM) and linear discriminant analysis (LDA) [7].Conventional MI-BCI system consists of single feature extraction and single classifier. And it often works not well in practical application, because single feature can't express all the features of motor imagery. In this case, we want to combine different features to get a new feature which contains as much details as possible. One solution is data fusion. Data fusion can combine advantage of each feature extraction and classification. There are three main types of data fusion, which are pixel fusion, feature fusion and decision fusion [8]. Pixel fusion is not suitable for EEG, because EEG is real-time single data which can't be fused. For feature fusion in MI-BCI, each feature has different units. Feature is meaningless after combining. Therefore, feature fusion is impracticable. Decision fusion is fusing the results of each method of feature extraction and classifier. Each method's result is the label that represents imaging right hand movement or left hand movement. Decision fusion is a feasible way to improve MI-BCI's performance. Our study proposes a new method of decision fusion to increase the accuracy of MI-BCI system.

II. Methods

A. Data source and description

The dataset are from BCI competition II. This dataset was recorded from a normal subject (female, 25 years old) during a feedback session. The subject sat in a relaxing chair with armrests. The task was to control a feedback bar by imagining left or right hand movements. The order of left and right cues was random.

The experiment consists of 7 runs with 40 trials each. All runs were conducted on the same day with several minutes break in between. Given are 280 trials of 9s length. The first 2s was quiet, at t=2s an acoustic stimulus indicates the beginning of the trial, the trigger channel went from low to high, and a cross "+" was displayed for 1s; then at t=3s, an arrow (left or right) was displayed as cue (Fig.1, Fig.2) . The recording was made using a G.tec amplifier and Ag/AgCl electrodes. Three bipolar EEG channels (anterior '+', posterior '-') were measured over C3, Cz and C4. The EEG was sampled with 128Hz; it was filtered between 0.5 and 30Hz (more details [9-12]). We divided the dataset into 7 groups of sub-datasets, which are train dataset (90 trials), test dataset1, test dataset2...test dataset6(each test dataset contain 30trials). The number of trials imaging left hand movement is equal to the number of trials imaging right hand movement in each sub-dataset. There are 45 trials imagining left hand movement and 45 trials imaging right hand movement in the train dataset. In each test dataset, there are 15 left trials and 15 right trials.

978-1-5386-7604-2/18/$31.00 ©2018 IEEE

B. Feature extraction and classification

CSP is the major method of feature extraction for MI-BCI system, but it needs user's prior knowledge. For example, we need to know appropriate filter band for user in advance. And CSP's performance is easily disturbed by noise. In order not to rely on prior knowledge and make the MI-BCI system generic, we selected AR, PSD, SampEn as methods of feature extraction, and we used SVM and LDA as classifier. Combine of each feature extraction method and classifier is MI-BCI system's conventional data analysis method.

Auto-regressive (AR) is a linear prediction. It is given N data, and the data before or after the N point can be derived from the model.

If $\{x_t, t = 0, \pm 1, \pm 2, \ldots$ is a time series, and the white noise sequence is $\{\varepsilon_t, t = 0, \pm 1, \pm 2, \ldots$. The time series satisfying (1) is called p-order auto-regressive. (1) is a p-order auto-regressive model, denoted as AR(p).

$$x_t + a_1 x_{t-1} + a_2 x_{t-2} + \cdots \qquad \varsigma_{t-p} = \varepsilon_t \tag{1}$$

Power spectral density (PSD) is signal power in the unit band, and it represents the relation of signal power with frequency.

The power spectrum cannot be used directly for Fourier transform, and it is usually interrupted by the truncated function. We use time T to cut off the original signal. When T close to infinity, we assume that $F_T(\omega)$ is the Fourier transform of $f_T(t)$. According to Parseval's theorem, energy of $f_T(t)$ is (2), (3) is energy spectral density.

$$E_T = \int_{-\infty}^{\infty} \left| F_T(\omega) \right|^2 df \tag{2}$$

$$G(f) = \left| F_T(\omega) \right|^2 \tag{3}$$

The average power is calculated as follows:

$$S = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f_T^{\,2}(t) dt = \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} f_T^{\,2}(t) dt$$

$$= \int_{-\infty}^{\infty} \lim_{T\to\infty} \frac{\left| F_T(\omega) \right|^2}{T} df \tag{4}$$

$$P_S(\omega) = \lim_{T\to\infty} \frac{\left| F_T(\omega) \right|^2}{T} \tag{5}$$

$P_S(\omega)$ is power spectral density.

The definition algorithm of Sample entropy (SampEn) is described as follows:

For a given time series of N points $\{u(i)\}$, $\{u(i)\}$ is formed into m-dimensional vectors sequentially.

$$X_m(i) = [u(i), u(i+1), \cdots \qquad -1)] \tag{6}$$

For each I value, we calculate the distance between the vector $X_m(i)$ and the rest of the vector $X_m(j)$.

$$d[X_m(i), X_m(j)] = \max_{k=0\sim} \left| u(i+k) - u(j+\mathrm{k}) \right| \tag{7}$$

Given a similar tolerance r value, the number of $d[X_m(i), X_m(j)] < r$ is counted for each i value by us. And we calculate the ratio of this number to the total number of distances, and record as $B_i^m(r)$. Then we take the average value of $B_i^m(r)$ for all, and record as $B^m(r)$.

$$B_i^m(r) = \frac{d[X_m(i), X_m(j)] < r}{N - m - 1} \tag{8}$$

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \tag{9}$$

We increase the dimension to $m+1$, and construct a vector without $m+1$ dimension:

$$X_{m+1}(i) = [u(i), u(i+1), \cdots \qquad)] \tag{10}$$

For each I value, we calculate the distance between the vector $X_{m+1}(i)$ and the rest of the vector $X_{m+1}(j)$.

$$d[X_{m+1}(i), X_{m+1}(j)] = \max_{k=0\sim} \left| u(i+k) - u(j+\mathrm{k}) \right| \tag{11}$$

The number of $d[X_{m+1}(i), X_{m+1}(j)] < r$ is counted for each i value by us. And we calculate the ratio of this number to the total number of distances, and record as $B_i^{m+1}(r)$.

$$B_i^{m+1}(r) = \frac{d[X_{m+1}(i), X_{m+1}(j)] < r}{N - m - 1} \tag{12}$$

$$B^{m+1}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^{m+1}(r) \tag{13}$$

$$SampEn(m, r) = \lim_{N\to\infty} \{-\ln\left[B^{m+1}(r) / B^m(r) \right]\} \tag{14}$$

The sample entropy of the sequence is (12)[6]. We don't describe the details of above classifier algorithm in this paper. See reference for details [7].

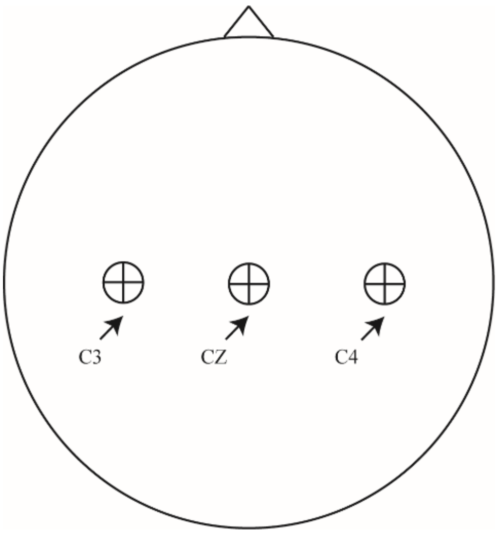

Figure. 1. Electrode positions.

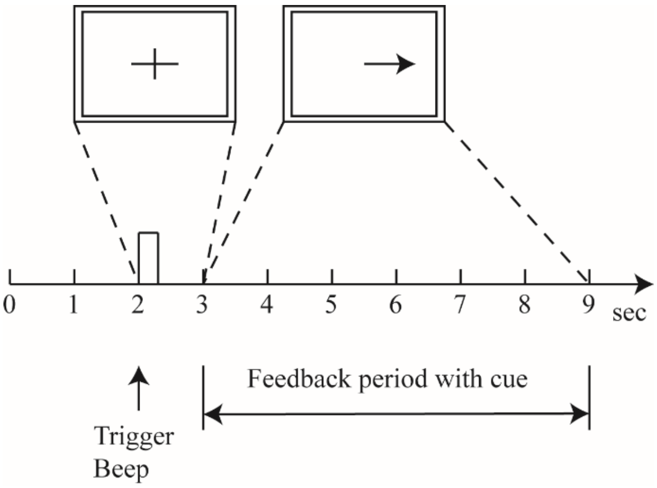

Figure. 2. Experiment's timing scheme.

C. *Multi-decision fusion*

First, in order to get SVM classifier model and LDA classifier model, we used train dataset to train SVM classifier and LDA classifier. There are 4 kinds of kernels which usually were used in SVM, and the classification performance is different when different kernel was used in SVM [13]. After a lot of studies, we found that SVM's performance is best for this datasets with RBF kernel. Therefore, we chose RBF kernel as SVM's kernel for this datasets classification. In addition, cost value (parameter C) and gamma (parameter G) are crucial parameter to SVM's performance. We could not get a more ideal accuracy with a small C value in classification. Too large a C value could lead to over-fitting. G value is a parameter that comes with RBF kernel, and it determines the distribution of data after mapping to the new feature space. If G is larger, there will be less support vectors; there will be more support vectors when G is smaller. The number of support vector affects the speed of training and prediction. In order to find the best C and G to get the best classification performance, we used 10 fold cross validation to get the best C and G. The C and the G are grouped within a certain range. The original data set was divided into 10 parts, among which 9 parts were taken as training data in turn and 1 part as test data. The selected C and G went to be tested through training set and test set, and the classification results were obtained 10 times, and the average value was calculated as the estimation of SVM algorithm precision. Finally, the C and G value with the highest classification accuracy were selected as the best C and G. We set up SVM's cost value and gamma value with best C and G, and we trained SVM classier model with train dataset to get the best-trained SVM classifier model. The first 6 train datasets were classified by the best-trained SVM classifier model and LDA classifier model. LDA doesn't work well for PSD and SampEn. Therefore, we used SVM to classify PSD feature and SampEn feature rather than LDA. We used AR+SVM, AR+LDA, PSD+SVM, SampEn+SVM as conventional MI-BCI data processing methods in this paper. AR-LDA label1 is the labels come from test dataset 1 classified by AR+LDA.AR-SVM label 2 is the labels come from test dataset 2 classified by AR+SVM. Other label's name rules are same as above label's name. Labels contain two classes: 1 represents this trial is imagining left hand movement, 2 represents imagining right hand movement. After we got 5 test datasets' labels, we fused the label like the following form:

$$newlabel(i) = a * AR - SVMlabel(i) + \mathrm{b} * AR - LDA \\ label(i) + c * PSD - SVMlabel(i) + d * SampEn - SVM \\ label(i) \qquad (1)$$

i is trial's no, a ,b ,c and d are weighting coefficient, they represent importance of each conventional data processing method's performance. If the method has higher accuracy, the method's performance is good and the method is important. And e is the threshold value which can distinguish imagining left hand movement or right hand movement. For example, if $newlabel(i) \le e$ (2),trial no. i represents imagining left hand movement, or it represents imagining right hand movement. In this dataset, PSD+LDA's performance is best, AR+LDA's performance comes second, and then after they is SampEn+SVM and AR+SVM. Table 1 shows the first 6 test datasets accuracy, and a ,b ,c , d and e satisfies the following conditions:

$$0 < a < d < b < c < 1 \qquad (3)$$

$$a + b + c + d = 1 \qquad (4)$$

$$1 \le e \le 2 \qquad (5)$$

Step size is 0.05 in (3),b ,c , d , e are same as a . We used the each conventional method's label which used first five test datasets to obtain the value of a ,b ,c , d and e , each test dataset's label can obtain several sets of a ,b ,c , d and e . We selected a common set of value as best fit set of value that we need to use it in the sixth test dataset. And then we verified that the decision-fusion can improve MI-BCI's accuracy.

TABLE I. Accuracies of test datasets.

	Dataset1	Dataset2	Dataset3	Dataset4	Dataset5	Dataset6
AR+SVM	70%	70%	63%	57%	57%	43%
AR+LDA	63%	73%	80%	77%	67%	80%
PSD+SVM	73%	80%	70%	64%	67%	87%
SampEn+SVM	70%	74%	64%	70%	54%	70%

TABLE II. Values of a, b, c, d and e.

a	b	c	d	e
0.05	0.3	0.45	0.2	1.45
0.05	0.35	0.45	0.15	1.45
0.05	0.4	0.45	0.1	1.45
0.1	0.3	0.4	0.2	1.4
0.1	**0.35**	**0.4**	**0.15**	**1.6**
0.1	0.35	0.4	0.15	1.75
0.1	**0.25**	**0.5**	**0.15**	**1.5**
0.15	0.25	0.4	0.2	1.45
0.15	0.25	0.4	0.2	1.55
0.15	0.25	0.4	0.2	1.35
0.15	0.25	0.4	0.2	1.65
0.15	0.25	0.4	0.2	1.70
0.15	0.25	0.4	0.2	1.75
0.15	0.25	0.4	0.2	1.80
0.05	0.3	0.45	0.2	1.40
0.05	0.3	0.45	0.2	1.50
0.05	0.3	0.45	0.2	1.55

III. RESULTS

We use four conventional methods to analysis all test datasets. Table I shows the four methods' accuracy in all test datasets. The conventional method which has highest accuracy is PSD+SVM, its average accuracy 74%. Table II shows values of a, b, c, d and e, bold is the common value that we selected as best fit value in the sixth test dataset. The best fit value is more than a set of value. Table III shows average accuracy of each method and the accuracy after multi-decision fusion. Results show that multi-decision fusion performs better than other methods.

IV. DISCUSSION

In this paper, we used AR, PSD, SampEn as methods of feature extraction. Classification methods were LDA and SVM. By fusing decision, we gained a higher accuracy than traditional feature extraction and classification. We used BCI competition datasets as experimental datasets, there are 280 trials in BCI competition datasets. If we have more datasets and make step size smaller (like 0.01 or 0.001), we can gain more precise values of a, b, c, d and e. More precise values can be used to distinguish the movement of the imaginary left hand from the movement of the imaginary right hand.

TABLE III. Average accuracy of each method and multi-decision fusion's accuracy

Method	accuracy
AR+LDA	73%
AR+SVM	60%
PSD+SVM	74%
SampEn+SVM	67%
Multi-decision fusion	82%

Therefore more precise values of a, b, c, d and e make multi-decision fusion perform better. But smaller step size brings more calculated quantity for computer, and it will take more time and make it inefficiency. Therefore, choosing a proper step size is important for multi-decision fusion. There is more than a set of common values of a, b, c, d and e, each set of values may get different result in other datasets. We should do more study on how to get more precise values of a, b, c, d and e, and make multi-decision fusion more effective and applicable. In addition, the best conventional method we used is PSD+SVM whose average accuracy is 74%, and the effect of AR+SVM is not good, and its average accuracy is 60%. By fusing above conventional method we improve the MI-BCI accuracy. If we fuse the results of other methods which's performance is better than now, we may be able to gain a higher accuracy than now.

V. CONCLUSION

In present study, we combined labels of 4 methods, and proposed a new method of classification for MI-BCI system which is named as multi-decision fusion. Multi-decision fusion can fuse different method's results and gain a new result. The new result has higher accuracy than each conventional method. In the next work, we will optimize our approach and use multi-decision fusion in online MI-BCI's classification to test its performance. We will try to fuse other methods of feature extraction and classification to gain a greater result.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (61773076 and 61806025), Scientific Research Project of Jilin Provincial Department of Education during the 13th Five-Year Plan Period (2019---) and Jilin Scientific and Technological Development Program (20180519012JH).

REFERENCES

[1] Pfurtscheller G, Neuper C. Motor imagery and direct brain-computer communication. Proceedings of the IEEE. 2001;89(7):1123-34.

[2] Wei P, He W, Zhou Y, Wang L. Performance of motor imagery brain-computer interface based on anodal transcranial direct current stimulation modulation. IEEE Trans Neural Syst Rehabil Eng. 2013;21(3):404-15.

[3] Robinson N, Vinod AP, Guan C, Kai KA, Peng TK, editors. A Wavelet-CSP method to classify hand movement directions in EEG based BCI system. Communications and Signal Processing; 2012.

[4] Barachant A, Bonnet S, Congedo M, Jutten C. Multiclass brain-computer interface classification by Riemannian geometry. IEEE Transactions on Biomedical Engineering. 2012;59(4):920-8.

[5] Trejo LJ, Rosipal R, Matthews B. Brain-computer interfaces for 1-D and 2-D cursor control: designs using volitional control of the EEG spectrum or steady-state visual evoked potentials. IEEE Transactions on Neural Systems & Rehabilitation Engineering. 2006;14(2):225-9.

[6] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology Heart & Circulatory Physiology. 2000;278(6):H2039.

[7] Sebald DJ, Bucklew JA. Support vector machine techniques for nonlinear equalization. Signal Processing IEEE Transactions on. 2000;48(11):3217-26.

[8] Xu J, Li JX, Xu S. Data fusion for target tracking in wireless sensor networks using quantized innovations and Kalman filtering. Science China(Information Sciences). 2012;55(3):530-44.

[9] Pfurtscheller G, Neuper C, Schlögl A, Lugger K. Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. IEEE Transactions on Rehabilitation Engineering. 1998;6(3):316-25.

[10] Schloegl A, Lugger K, Pfurtscheller G, editors. Using adaptive autoregressive parameters for a brain-computer-interface experiment. Engineering in Medicine and Biology Society, 1997 Proceedings of the International Conference of the IEEE; 1997.

[11] Neuper C, Schlogl A, G. Enhancement of left-right sensorimotor EEG differences during feedback-regulated motor imagery. Journal of Clinical Neurophysiology. 1999;16(4):373-82.

[12] Schlögl A, Neuper C, Pfurtscheller G. Estimating the mutual information of an EEG-based Brain-Computer Interface. Biomedizinische Technik/biomedical Engineering. 2002;47(1-2):3.

[13] Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning. 2002;46(1-3):389-422.