Classification of Single-Trial EEG Based on Support Vector Clustering during Finger Movement

Boyu Wang and Feng Wan

Department of Electrical and Electronics Engineering,
Faculty of Science and Technology,
University of Macau

Abstract. Classification of electroencephalogram (EEG) is an important and challenging issue for brain computer interface (BCI) system. In this paper, an algorithm based on common spatial subspace decomposition (CSSD) and support vector clustering (SVC) is proposed to classify single-trial EEG recording during left or right finger movement. The algorithm is tested by the dataset IV of "BCI competition 2003", and the experimental result shows the proposed method, only using bereitschaftspotential (BP), rather than both BP and event-related desynchronization (ERD), has higher classification accuracy than the best one reported in the competition.

Keywords: Support vector clustering (SVC), Common spatial subspace decomposition (CSSD), Electroencephalogram (EEG), Brain computer interface (BCI).

1 Introduction

Brain computer interface (BCI) is a communication system that allows its users to control external devices with brain activity, which does not depend on the brain normal output pathways of peripheral nerves and muscles [1], [2]. Currently, the electroencephalogram (EEG) signal, one of the non-invasive measurements of brain activity, due to its excellent temporal resolution and usability, is a most prevailing signal used in BCI system. Therefore, the BCI system based on EEG is widely studied and a variety of algorithms have been proposed to indentify intended motions of the subjects in EEG recordings.

In the BCI system design, a common approach is to ask the user to perform tasks that are known to produce distinguishable brain activity in most people [3], and task involving classification of finger movements, due to its simplicity and easy to implement, has been studied by many researchers [3]-[9]. Wolpaw *et al* [1] categorized the BCI systems into five major groups, which are sensorimotor activity, P300, slow cortical potentials (SCPs), visual evoked potentials (VEPs), and activity of neural cells (ANC). For the finger movement classification task, most of the current work extract the features from movement-related potentials (MRPs), e.g., bereitschaftspotential (BP), and changes in brain rhythms, e.g., event-related desynchronization/synchronization (ERD/ERS), which can be both viewed as the first

W. Yu, H. He, and N. Zhang (Eds.): ISNN 2009, Part II, LNCS 5552, pp. 354–363, 2009.

category of electrophysiological activities, i.e. sensorimotor activity used in BCI system designs.

Conventional analysis related to movement tasks requires the subject's training to control their brain rhythms for long time, or averaging multiple trials to enhance the EEG signal. Another approach is to detect EEG related to movement task from single trial, and has attracted more and more attentions due to its simplicity and short response time. The signal-to-noise ratio (SNR) of single trial EEG, however, is rather low, and therefore lots of algorithms based on single-trial EEG have been investigated to resolve this problem.

To improve the classification accuracy of a BCI system, practitioners have proposed various methods. Basically, the literature in this field can be divided into three categories. The first approach focuses on how to detect the EEG with more information or higher SNR, which is associated with the technique related to signal acquisition. The second is studying the use of available information with more efficiency by using pre-processing, feature selection and/or extraction technique. The third is to explore the classification algorithms to distinguish the complicated features. In this paper, we propose a novel classifier based on common spatial subspace decomposition (CSSD) and support vector clustering (SVC) to identify the finger movement attempts from the single EEG trial. The proposed algorithm is tested by the dataset IV of "BCI competition 2003", and the experimental result shows that the proposed method, only using BP, rather than both BP and ERD, has higher classification accuracy than the best one reported in the competition by Tsinghua University, so that the pre-processing and the feature extraction/selection step will be simplified considerably.

2 Methodology

In the present research, most of practitioners identify the finger movement intents based on the combination of both the BP and ERD to improve the classification accuracy [4]-[6]. BP and ERD, with different frequency bands, can be viewed as different responses of sensorimotor cortex. However, utilizing both of them for classification makes the identification process more complicated and difficult to implement. In some situations, simpler method is desirable for the classification task. One feature is therefore enough if the performance of the classifier is acceptable.

CSSD, one of spatial filters, similar to the common spatial pattern (CSP) method, proposed by Yunhua Wang *et al* [10], has shown great usefulness in the finger movement classification task [4]-[6]. In this paper, CSSD is also to process and extract the feature of multichannel EEG signal. Moreover, SVC is used to design the classifier to distinguish the left and right finger movement intents.

2.1 Bereitschaftspotential

(Movement-related potentials) MRPs have bilateral distribution and present maximum amplitude at vertex. Close to the movement, they become contralaterally preponderant [11]. Bereitschaftspotential (BP), also named readiness potential (RP), as a component of MRPs, is low-frequency potential that reflects the dynamic changes in motor cortical activity prior to the movement onset. Thus, the feature extracted from BP can be utilized in the finger movement task [4]-[6]. In this paper, we also utilize the features derived from BP. (For more details about BP, one can refer to [12] and [4].)

2.2 Common Spatial Subspace Decomposition

In order to utilize more information, one should use all of the electrodes rather than only a subset of them. Thus people proposed spatial filters, which combine all the electrodes to process multi-channel EEG. CSP [13] is a method belonging to this family. Given a binary classification task, CSP seeks a projection direction which maximizes the power of one class, and simultaneously minimizes the power of the other one.

Common spatial subspace decomposition (CSSD) is a variation of CSP, and has been applied successfully in the finger movement classification task. The aim of CSSD is to separate the evoked responses and background spontaneous brain activities (specific and common activities), which are overlapped in the scalp measurement [10]. Given single-trial multichannel spatial-temporal EEG signal matrices X_L and X_R (evoked by left and right finger movements respectively) with dimension N (channels) by T (samples), they can be modeled as follows:

$$X_L = [C_L \ C_C]\begin{bmatrix} S_L \\ S_C \end{bmatrix} \qquad X_R = [C_R \ C_C]\begin{bmatrix} S_R \\ S_C \end{bmatrix} \tag{1}$$

where C_L and C_R are the spatial patterns related to left and right finger movements respectively, and C_C represents the spatial pattern specific to the background activities. Then S_L , S_R and S_C are the corresponding source activities related to the left and right hand movements, and the common condition. One can construct spatial filters F_L and F_R by using CSSD to extract source activities:

$$S_L = F_L X \qquad S_R = F_R X \tag{2}$$

Then the CSSD algorithm can be described as in the following steps:

1. Estimate the normalized spatial covariances of the single-trial multichannel EEG signal:

$$R_L = \frac{X_L X_L^{\ T}}{trace(X_L X_L^{\ T})} \qquad R_R = \frac{X_R X_R^{\ T}}{trace(X_R X_R^{\ T})} \tag{3}$$

where trace(X) denotes the summation of the diagonal elements of X. Then calculate the averaged normalized covariances $\overline{R}_L$ and $\overline{R}_R$:

$$\overline{R}_L = \frac{1}{N_L}\sum_{i}^{N_L} R_L(i) \qquad \overline{R}_R = \frac{1}{N_R}\sum_{i}^{N_R} R_R(i) \tag{4}$$

where N_L and N_R are the numbers of the trials corresponding to left and right finger movements respectively.

2. Calculate the eigenvectors U_0 and eigenvalues Σ of the matrix R:

$$\overline{R} = \overline{R}_L + \overline{R}_R = U_0 \Sigma U_0^T \tag{5}$$

3. Construct the whiten matrix:

$$P = \Sigma^{-1/2} U_0^T \tag{6}$$

4. Transform the covariance matrices:

$$Y_L = P\overline{R}_L P^T \quad Y_R = P\overline{R}_R P^T \tag{7}$$

It can be shown [14] that Y_L and Y_R share the eigenvectors, i.e.

$$Y_L = U_L \Sigma_L U_L^T \quad Y_L = U_R \Sigma_R U_R^T \tag{8}$$

$$U_L = U_R = U \text{ and } \Sigma_L + \Sigma_R = I \tag{9}$$

where Σ_L and Σ_R are the eigenvalue matrices of Y_L and Y_R respectively, and I is the identity matrix. Since the eigenvalues are ordered in reverse, the eigenvector with the largest eigenvalue for one matrix has the smallest eigenvalue for the other, and vice versa.

5. Design the spatial filter:
The first and last eigenvectors (denoted as u_L and u_R respectively) are the optimal vectors to distinguish the finger movements and the spatial filters F_L and F_R of left and right finger movements can be therefore designed as:

$$F_L = u_L^T P \quad F_R = u_R^T P \tag{10}$$

2.3 Support Vector Clustering

Inspired by support vector machine, support vector clustering (SVC) was proposed by Ben-Hur *et al.* [15], [16] to find a set of contours as clustering boundaries in the original data space. The data are mapped by means of a Gaussian kernel to a high dimensional feature space, where the minimal enclosing sphere is found [15]. When mapped back to the input space, the sphere represents a complex geometric shape as a clustering boundary.

Let $\{x_i\} \subseteq \chi$ be a data set of N points, with $\chi \subseteq \mathbb{R}^d$, the input space. Using a nonlinear transformation $\Phi : \chi \to \mathbb{F}$, where $\mathbb{F}$ is the feature space, we look for the smallest sphere of radius R, which encloses the data projection $\Phi(x_j)$. This is described by the constraints:

$$\left\| \Phi(x_j) - a \right\|^2 \le R^2, \quad \forall j \tag{11}$$

where $\|\cdot\|$ is the Euclidean norm and a is the center of the sphere. Soft constraints are incorporated by adding slack variables ξ_j :

$$\|\Phi(x_j)-a\|^2 \leq R^2+\xi_j\,, \quad \forall\, j \tag{12}$$

with $\xi_j \geq 0$. This problem can be solved by introducing the Lagrangian:

$$L = R^2 - \sum_j (R^2+\xi_j - \|\Phi(x_j)-a\|^2)\beta_j - \sum_j \xi_j \mu_j + C\sum_j \xi_j \tag{13}$$

where $\beta_j \geq 0$ and $\mu_j \geq 0$ are Lagrangian multipliers, C is a constant and $C\sum_j \xi_j$ is a penalty term. Setting to zero the derivative of L with respect to R, a and ξ_j , leads to

$$\sum_j \beta_j = 1 \tag{14}$$

$$a = \sum_j \beta_j \Phi(x_j) \tag{15}$$

$$\beta_j = C - \mu_j \tag{16}$$

Then the Karush-Kuhn-Tucker conditions yield:

$$\xi_j \mu_j = 0 \tag{17}$$

$$(R^2+\xi_j - \|\Phi(x_j)-a\|^2)\beta_j = 0 \tag{18}$$

Using these relations, the Lagrangian can be turned into the Wolfe dual form W that is a function of the variable β_j :

$$W = \sum_j \Phi(x_j)^2 \beta_j - \sum_j \beta_i \beta_j \Phi(x_i)\cdot\Phi(x_j) \tag{19}$$

Since the variables μ_j do not appear in the Lagrangian they may be replaced with the constraints:

$$0 \leq \beta_j \leq C\,, \quad j = 1,\ldots,N \tag{20}$$

The inner product $\Phi(x_i)\cdot\Phi(x_j)$ can be computed by using an appropriate Mercer kernel $K(x_i, x_j)$. Since the polynomial kernels do not permit tight contours representation of a cluster, it is suggested to choose Gaussian kernel $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/\sigma^2)$ [15], and we therefore adopt the Gaussian kernel in our experiment.

Finally, by using the kernel trick, the Lagrangian W can be written as follows:

$$W = \sum_j K(x_j, x_j)\beta_j - \sum_{i,j} \beta_i \beta_j K(x_i, x_j) \tag{21}$$

For $\beta_j = C$, the corresponding points are called bounded support vectors (BSVs), and the points with $0 < \beta_j < C$ are referred to as support vectors (SVs). SVs lie on cluster boundaries, BSVs lie outside the boundaries and all other points lie inside them. It should be noted that, due to constraint (14), when $C \geq 0$, no BSVs exist.

At each point x the distance of its image in the feature space from the center of the sphere is:

$$R^2(x) = K(x, x) - 2\sum_j \beta_j K(x_j, x) + \sum_{i,j} \beta_i \beta_j K(x_i, x_j) \tag{22}$$

One advantage of SVC is that it can form arbitrary clustering shapes other than hyperellipsoid and hypersphere. Furthermore, it has the capability to deal with the noise and outliers, and for SVC, there is no requirement for prior knowledge to determine the system topological structure [17].

3 Experiments

The algorithm proposed is evaluated on the dataset IV in BCI Competition 2003 [18], which is provided by Fraunhofer-FIRST, Intelligent Data Analysis Group, and Freie Universität Berlin, Department of Neurology, Neurophysics Group. The dataset is recorded from a normal subject during a no-feedback session. The task is to press with the index and little fingers the corresponding keys in a self-chosen order and timing 'self-paced key typing'. The EEG is collected by 28 electrodes at the positions of the international 10/20-system. The duration of the signal is 500ms ending 130 ms before a keypress, and the sample rate is 100Hz. There are 416 trials in the dataset, including 316 training trials and 100 testing trials.

3.1 Data Preprocessing

The use of a preprocessing stage before feature extraction or classification has been proven to be useful [11]. In order to increase the SNR of the EEG and utilize the information more efficiently, two types of filters, i.e. frequency and temporal filters are used.

Since the BP of finger movement dominates in the low frequency band, a low-pass filter is applied to extract the BP of the finger movement from the EEG. The cutoff frequency is 7Hz, which was used in the previous work [4]. It should be noted that the filter used here is the zero-phase filter to avoid phase shift.

For the temporal filter, there are two parameters to be determined, i.e. the starting time and the window size. In this paper, they are chosen by four-fold cross-validation on the training data. Moreover, there are another two parameters of SVC, i.e. the width σ of the Gaussian kernel and the penalty parameters C. To resolve the problem, we firstly determine the parameters of the SVC by experience, and choose the optimal parameters for the temporal filter to extract most obvious differences between the two finger movements. Based on the designed filter, we select the parameters of SVC by using grid search for simplicity. Then, we redesign the filter again based on the selected parameters of SVC. After finite times of iterations, we can obtain an approximate optimal solution of the parameters. The classification accuracies with different starting times and window sizes are shown as follow:

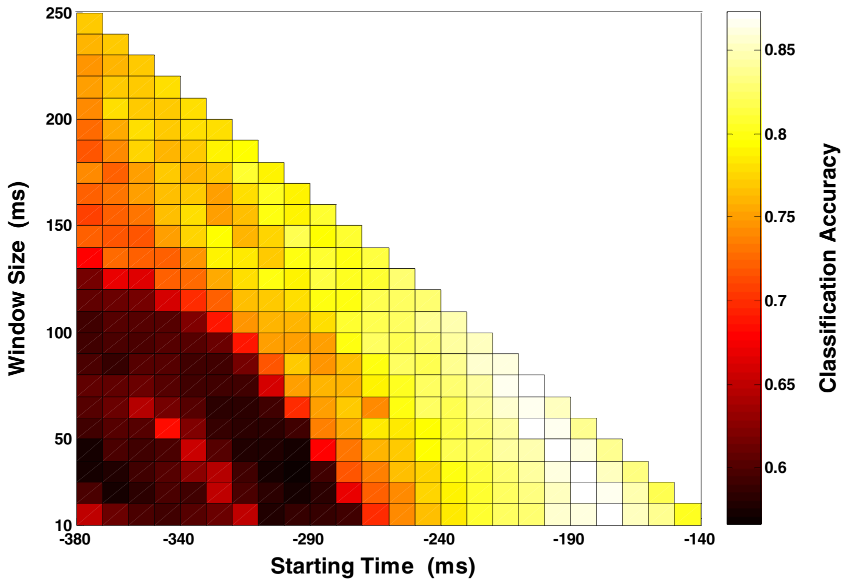

Fig. 1. The performances of the classifiers with different starting times and window sizes

Here, the parameters of SVC classifier, C and σ, are 0.1 and 1024 respectively. From Fig.1 it can be observed that the best classification accuracy is achieved at about -190 ms before a keypress, and the corresponding window size is about 60 ms.

3.2 Feature Extraction

The CSSD is applied to design a spatial filter to exploit the information from all the electrodes. Based on equation (10) we estimate the spatial filters F_L and F_R by using the training data, and then for each testing data X, we define $f = [x_L \ \ x_R]$ as the feature vector, where $x_L = F_L \cdot X$ and $x_R = F_R \cdot X$. It has been shown [4] that x_L of left trials and x_R of right trials have larger amplitudes than that of the contrary patterns, which makes the high accuracy classification possible.

3.3 SVC Discrimination

After extraction of the features, the SVC classifier is applied to classification. The key idea of the SVC classifier is to find a clustering center for each type of finger movement, and classify the movements by comparing the distance between the input data and the clustering center of each class in the feature space. SVC can form arbitrary clustering shapes and does not require prior knowledge to determine the system topological structure, therefore it is suitable for the finger movement classification task since the distribution of the extracted feature is complicated.

The explicit expression of the clustering center in the feature space cannot be solved. For each input data, the distance from the center of the sphere in the feature, however, can be calculated by (22). In this experiment, SVC algorithm is applied for each class of training data, and the corresponding parameters β_j for each class are calculated by maximizing W in (21) respectively. In the testing stage, we calculate the distance between the input point and each clustering center by using (22), and assign the label of the closest center. As mentioned above, there are still two parameters, σ and C, for SVC to be selected carefully. Table 1 shows the cross-validation performances of the SVC classifiers with different parameters. The best performance is shown in bold font. The starting and window size of the temporal filter here are -200 ms and -60 ms respectively.

Table 1. The performances of the SVC with different pairs of parameters C and σ (%)

σC	4^{-1}	1	4	4^2	4^3	4^4	4^5	4^6	4^7	4^8
0.01	50.3	50.3	55.7	57.9	70.3	68.7	75.0	77.5	77.5	77.8
0.02	50.3	50.3	55.7	58.2	71.8.	73.1	75.6	77.8	77.8	80.4
0.05	52.2	52.2	52.5	60.1	76.3	75.6	84.2	83.5	82.9	82.6
0.1	52.2	52.2	52.2	63.3	78.2	83.1	**86.4**	84.2	82.9	83.8
0.2	52.2	52.2	52.5	62.7	78.2	77.8	82.3	82.9	81.6	82.3
0.4	52.2	52.2	52.2	57.6	76.3	74.7	71.8	80.4	82.6	81.6
0.8	52.2	52.2	52.2	60.4	75.6	73.1	69.0	79.1	80.1	79.4
1	52.2	52.2	52.5	65.2	72.8	72.8	60.8	71.5	73.1	73.7
2	49.7	49.7	52.5	64.9	69.9	70.9	62.7	71.8	71.8	75.0

3.4 Experimental Results

Table 2 shows the performances of the classifiers with some selected parameters, where t1 and t2 represent the starting time and the ending time of the temporal filter respectively. The parameters of the temporal filter in fourth column of the table, i.e. t1 = -190 ms, t2 = -160 ms, were used in [4]. The best result of the classifier achieved in this paper, with t1 = -200 ms, t2 = -140 ms, C = 0.1, and σ = 1024, is 86%.

As observed in Table 2, the classification accuracy of the classifier is sensitive to the parameters of the SVC classifier. Moreover, the optimal choice of them depends on the starting time and the ending time of the temporal filter. This is mainly because that the preprocessed data are not normalized out of consideration for avoiding losing information. Therefore, how to design the optimal classifier independent of data preprocessing is the main challenge for further research.

Table 2. The effects of different parameters on classification accuracies

t1,t2(ms) / C, σ	(-350,-200)	(-250,-150)	(-200,-140)	(-190,-160)	(-190,-130)
0.01, 4^{-1}	51%	51%	51%	51%	51%
0.1, 4^5	70%	80%	**86%**	84%	82%
0.1, 4^7	74%	82%	84%	83%	81%
0.2, 4^5	54%	78%	80%	81%	79%
0.4, 4^7	59%	77%	83%	82%	77%
1, 4^8	53%	79%	76%	80%	36%

4 Conclusion

This work proposes a novel method based on CSSD and SVC to classify the single-trial EEG signal during the finger movement. The performances of the classifiers with different parameters are also investigated. Since SVC requires no prior knowledge in determining the system topological structure and clusters in the feature space, it is suitable for classifying the features extracted from EEG. The proposed method, utilizing only one feature, i.e. BP, achieves the accuracy of 86%, better than 84% reported in [4], using both features (BP and ERD) extracted from EEG. Its simplicity makes the implementation at a low cost. In addition, the method proposed here can be easily extended for multi-class EEG classification task.

As for the future work, the development of efficient algorithm to determine the optimal parameters of SVC, and the design of the classifiers which are more insensitive to different data preprocessing strategies would be a focus. Moreover, new feature extraction and representation methods immune to different subjects are also very meaningful.

Acknowledgments. This work was supported in part by the Macau Science and Technology Development Fund and the University of Macau Research Committee.

References

1. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-Computer Interface for Communication and Control. Journal of Clinical Neurophysiology 113, 767–791 (2002)
2. Blankertz, B., et al.: The BCI Competition III: Validating Alternative Approaches to Actual BCI Problems. IEEE Transaction on Neural System and Rehabilitation Engineering 14, 153–159 (2006)
3. Lehtonen, J., Jylänki, P., Kauhanen, L., Sams, M.: Online Classification of the Single EEG Trials During Finger Movements. IEEE Transactions on Biomedical Engineering 55(2), 713–720 (2008)
4. Wang, Y., et al.: BCI Competition 2003-Data Set IV: An algorithm based on CSSD and FDA for Classifying Single-Trial EEG. IEEE Transactions on Biomedical Engineering 51(6), 1081–1086 (2004)

5. Li, Y., Gao, X., Liu, H., Gao, S.: Classification of Single-Trial Electroencephalogram During Finger Movement. IEEE Transactions on Biomedical Engineering 51(6), 1019–1025 (2004)
6. Li, Y., et al.: Single Trial EEG Classification During Finger Movement Task by Using Hidden Markov Models. In: 2nd International IEEE EMBS Conference on Neural Engineering, Arlington, Virginia, pp. 625–628 (2005)
7. Ranaweera, R.D., Talavage, T.M., Krishnan, A.: Time-frequency Features Differentiate Direction of Finger Movement in Cued and Self-paced Tasks. In: 2nd International IEEE EMBS Conference on Neural Engineering, Arlington, Virginia, pp. 551–554 (2005)
8. Pfurtscheller, G., Neuper, C., Flotzinger, D., Pregenzer, M.: EEG-Based Discrimination Between Imagination of Right and Left Hand Movement. Journal of Electroencephalography and Clinical Neurophysiology 103, 642–651 (1997)
9. Blankertz, B., et al.: Boosting Bit Rates and Error Detection for the Classification of Fast-Paced Motor Commands Based on Single-Trial EEG Analysis. IEEE Transactions on Neural System and Rehabilitation Engineering 11(2), 127–131 (2003)
10. Wang, Y., Berg, P., Scherg, M.: Common Spatial Subspace Decomposition Applied to Analysis of Brain Responses Under Multiple Task Conditions: A Simulation Study. Journal of Clinical Neurophysiology 110, 604–614 (1999)
11. Bashashati, A., Fatourechi, M., Ward, R.K.: A Survey of Signal Processing Algorithms in Brain-Computer Interfaces Based on Electrical Brain Signals. Journal of Neural Engineering 4(2), R32–R57 (2007)
12. Shibasaki, H., Hallett, M.: What is the Bereitschaftspotential? Journal of Clinical Neurophysiology 117, 2341–2356 (2006)
13. Müller-Gerking, J., Pfurtscheller, G., Flyvbjerg, H.: Designing Optimal Spatial Filters for Single-Trial EEG Classification in a Movement Task. Journal of Clinical Neurophysiology 110, 787–798 (1999)
14. Fukunaga, K.: Introduction to Statistical Pattern Recognition, 2nd edn. Academic Press, San Diego (1990)
15. Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V.: Support Vector Clustering. Journal of Machine Learning Research 2, 125–137 (2001)
16. Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V.: A Support Vector Clustering Method. In: 15th International Conference on Pattern Recognition, Barcelona, pp. 724–727 (2000)
17. Xu, R., Wunsch II, D.: Survey of Clustering Algorithms. IEEE Transactions on Neural Networks 16(3), 645–678 (2005)
18. Blankertz, B., Curio, G., Müller, K.-R.: Classifying Single Trial EEG: Towards Brain Computer Interfacing. In: Dietterich, T.G., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems (NIPS 2001), vol. 14, pp. 157–164 (2001)