A General Aggregate Model for Improving Multi-Class Brain-Computer Interface Systems’ Performance

Tuan Hoang, Dat Tran, Xu Huang and Wanli Ma

Abstract—This paper proposes a general aggregate model for improving performance of multi-class Brain-Computer Interface (BCI) systems. In BCI systems, activation and delay are well known issues in conducting experiments. The delay of meaningful brain signal depends on subjects, tasks and experimental design. Therefore, within a trial it is not easy to identify where meaningful brain signal starts and ends. Most of current methods estimate the delay and extract a portion of meaningful brain signal in a trial and use this signal as a representative for the whole trial. Instead of doing so, our proposed aggregate model divides a trial into overlapping frames and treat them equally. These frames are classified and their results are then aggregated together to form classification result of the trial. From the general aggregate model, we derive two specific aggregate models using two state-of-the-art Common Spatial Patterns (CSP)-based methods for feature extraction. We performed experiments on Dataset 2a in BCI Competition IV to evaluate the proposed models. This dataset was designed for motor imagery classification with 4 classes. Preliminary experimental results show that our proposed aggregate models are up to 8% better than the original CSP-based methods. Furthermore, we show that our aggregate model can be easily extended to online BCI systems.

I. INTRODUCTION

Activation and delay are well known issues [1][4] in conducting experiments in Brain-Computer Interface (BCI) systems. It is an effect where time of stimulation and response are different. Unlike other popular visual or audio signals such as speech or face to which human-beings can manually recognize and so quite accurately estimate the delay of the signal, brain signals are very difficult to control, read and interpret. In theory, when receiving stimulation input, brain needs time to process and then produce output. Macaluso et. al. in their work [1] proves that the delay depends on subjects, types of target (visual or tactile) and types of motor activity. Their work however mainly focuses on estimating the delay which is the start time but not the end time which helps extract the meaningful portion of the trial. Our experiments using functional Near Infrared Spectroscopy (fNIRS) technology [2][3] show that even within a trial, subjects can lose their concentration and therefore there are multiple portions containing meaningful signal in a trial rather than one. Consequently, aggregate model can be a good one for dealing with brain signals in real and online experiments.

Aggregate model is a well-known approach in signal processing, especially in speech signal processing [5]. There are two main types of aggregate model. In the first type, a trial is divided into multiple signal frames. These frames are then extracted features and used for training classifiers. After that, the classification result of a trial is determined by aggregate function whose inputs are results obtained from classifying its frames. The aggregate function can be a min, max, average or count-based function. In the second type, aggregate model is called fusion model in which different sources, features or classifiers of the same trial are combined by some aggregate function. While the fusion model is widely used in BCI systems [6][7], there are a few works following the first type of aggregate which is successfully applied in other signals such as speech.

We propose a general aggregate model and then derive from it two specific aggregate models based on Common Spatial Patterns analysis for multi-class BCI systems. Common Spatial Pattern (CSP) is one of the state-of-the-art feature extraction methods in BCI systems. It was originally proposed by Koles [9] to analyze abnormal components in clinic research and then successfully applied to 2-class BCI systems [8][11]. The idea of CSP is to map data of two classes onto the same dimension such that variance of one class is maximized while variance of the other one is minimized. For multi-class BCI systems, the main idea of current CSP-based feature extraction methods is to convert the multi-class classification problem to a set of 2-class classification problems. The two well-known methods for this purpose are one versus the rest and combination of pairs of 2-class classification problems. We tested our models with a specific classification method Support Vector Machines (SVM), a state-of-the-art classification method for BCI systems [10]. Dataset 2a from BCI Competition IV [12] is used for conducting experiments.

The remaining of the paper is organized as follows. In Section 2, we briefly present theoretical foundation of Common Spatial Patterns analysis and its extensions for multi-class BCI systems. In Section 3, we describe our proposed aggregate model. Experimental protocols as well as methods for classification and validation are introduced in Section 4. Section 5 presents our results and related discussions. Finally, we conclude and present our future work in Section 6.

II. COMMON SPATIAL PATTERNS ANALYSIS AND ITS EXTENSIONS FOR MULTI-CLASS BCI SYSTEMS

Let $X_i = x_i(t)$ be the $i^{th}$ trial of the dataset. All trials consist of $n$ channels $x^p(t) = (x^p_1, x^p_2, \ldots, x^p_T)$ where $p = \ldots$
1, \ldots, n is the channel index, \( t \) is the time index of the signal, and \( T \) is its length or number of samples. Each trial belongs to a class \( L(X_i) \) of mental action or motor imagery. Assume that there are \( k \) classes in the dataset \( L(X_i) \in [1, k] \).

Let \( \text{Cov}(X_i) = X_iX_i^T \) be the covariance matrix of trial \( X_i \) and \( C_i \) be the estimate covariance matrix of class \( i^{th} \). We use empirical method to estimate these covariance matrices.

\[
C_i = \frac{1}{|X_q: L(X_q) = i|} \sum_{X_q: L(X_q) = i} \text{Cov}(X_q) 
\]

\[
= \sum_{X_q: L(X_q) = i} X_qX_q^T 
\]

(1) (2)

A. 2-class Common Spatial Patterns

Given 2 covariance matrices \( C_1 \) and \( C_2 \) corresponding to 2 classes of the dataset, Common Spatial Patterns (CSP) analysis aims to find a matrix to simultaneously diagonalize these two matrices,

\[
C_1 = V\lambda_1V^T, \quad (3)
\]

\[
C_2 = V\lambda_2V^T \quad (4)
\]

which satisfies the condition

\[
\lambda_1 + \lambda_2 = I. \quad (5)
\]

This problem is well known and its solution can be achieved by solving generalized eigenvalue problem. The matrix \( V \) which consists of common generalized eigenvectors is called the spatial filters of the signal. Each eigenvector in \( V \) is a spatial filter and its associated eigenvalue is the variance of the projected signal. Due to its constraint in Equation (5), if variance of one class is the largest then variance of the other class will be the smallest and vice versa. This property is therefore very useful in class discrimination. To avoid overfitting problem and reducing processing time, not all but a few spatial filters are selected for classification. Blankertz et al. [8] proposed using at most 6 spatial filters. They also noted that finding the matrix \( V \) is actually to find the common basic of the two classes.

B. Feature extraction with Common Spatial Patterns in BCI systems

In BCI systems, data is acquired by multi-channel devices. We derive common principal components \( V \) of data by applying the CSP analysis as shown above. Original data will then be mapped on new subspace as shown in Equation (6).

\[
X_i^{\text{CSP}} = V^TX_i
\]

Feature vector = \((\log(\text{var}(X_i^{\text{CSP}})))\)

(6) (7)

The feature vector of a trial as shown in Equation (7) is formed by combining the variances of all channels from mapped data. Because the variance of filtered signal is equivalent to its band power, the feature vector is equivalent to logarithmic band power. So the feature vector of a trial \( X_i \) is \( \log(\text{var}(X_i^{\text{CSP}})) \) and its size is \( q \times n \) where \( q \) is the number of the selected components and independent of length of the trial. This property is useful in allowing us to flexibly determine the length of trials, especially in real time or online BCI systems.

C. CSP-based extensions for multi-class BCI systems

The main idea of CSP-based methods for multi-class BCI systems is to convert a multi-class problem into multiple 2-class problems and then apply 2-class CSP analysis to solve these problems. Based on that idea, there are two well-known strategies. The first strategy is called one-versus-the-rest (CSP_1vsN) strategy. In this strategy, we take one class out and assume that the \( k - 1 \) remaining classes have highly similar covariance matrices. The 2-class CSP analysis is then applied to covariance matrix of the selected class and estimated covariance matrix of all the \( k - 1 \) remaining classes. There are \( k \) possible ways to select the selected class therefore we have \( k \) 2-class problems that need to be solved. Final spatial filters are a combination of selected spatial filters of these \( k \) 2-class problems. The second strategy is to consider all possible pairs of \( k \) classes (CSP_pairs). So there are totally \( \frac{k(k-1)}{2} \) 2-class problems that need to be solved. Similar to the 1vsk_CSP strategy, the final spatial filters are a union of all selected spatial filters of all these \( \frac{k(k-1)}{2} \) 2-class problems.

III. AGGREGATE MODEL FOR BCI SYSTEMS

The idea of aggregate model comes from two observations we discover when conducting experiments. The first observation is that there is delay between stimulation and response and it depends on various factors such as subjects, types of stimulation, types of tasks, and design of experiments. The second one is that it is difficult to keep subjects concentrated during experiments, especially for motor imagery tasks. Therefore instead of extracting single portion of meaningful signal from a trial, we extract multiple portions of signal from a trial which is called frame. These frames can be overlapped to make sure they can capture meaningful signal due to delay issue. They can also be separated to make sure they can capture multiple portions of meaningful signal. The two parameter window size \( w \) and window step \( s \) are used to set trade off between these two issues. Fig. 1 shows how frames are created and used. These frames are equally treated in training and testing. They has the same class label with its trial. Actually, we can think them as pseudo-trials. Feature extraction methods such as Common Spatial Patterns can then be used for feature extraction. To avoid overfitting problem, all frames of a trial must be in either training set or testing set. Classification result of a testing trial is determined by an aggregate function of all its frames’ classification results.

The aggregate function we use in this paper as shown in Equation (8), where \( X \) is the testing trial, \( F \) is a set of its frames, \( c \) is a class label and \( SVM \) is the trained classifier.

\[
L(X) = \arg_c \left( \sum_{f \in F} SVM(f) = c \right) \geq \frac{|F|}{2} \quad (8)
\]
IV. EXPERIMENTAL METHODS AND VALIDATIONS

Our main purpose is to investigate the accuracy improvement in multi-class BCI systems using our proposed aggregate model. From our proposed model in Figure 1, we developed two specific models based on one-versus-the-rest ($CSP_{1vsN}$) and pair-wise ($CSP_{pairs}$) methods for feature extraction combined with Support Vector Machines as classifiers. The Dataset 2a from BCI Competition IV [12] which is a well-known dataset for multi-class BCI systems was chosen for conducting experiments. The dataset was acquired by Graz University of Technology, Austria using Electroencephalography (EEG) technology with 22 channels at sampling frequency 250Hz. Nine subjects were asked to perform 4 classes of motor imagery tasks to move cursor left, right, down or up corresponding to imagination of movement of the left hand, right hand, both feet and tongue. There are 576 trials in total in both training and testing sets of the competition. For each trial, there are 2 seconds to help participants prepare themselves. After that, there is a cue appearing and staying on screen in 1.25s. The subjects were asked to carry out motor imagery tasks until the $6^{th}$ second.

We segmented data into lengths of 2 seconds from $t = 2.5$ second. We moved segment window by half of second time frame. All these segments were bandpass filtered with frequency cut-off at 8Hz and 30Hz before were extracted features as in Equation (refeq:19). Support Vector Machine (SVM) and its popular kernel function RBF $K(x,x') = e^{-\gamma \|x-x'\|^2}$, a state-of-the-art method for classifying in BCI systems [10], was chosen to classify data. We applied grid search to get the optimal parameters. The parameter $\gamma$ was searched in range $2^k : k = -10, -9, \ldots, 4, 5$. The trade-off parameter $C$ was searched over the grid $2^k : k = 0, 1, \ldots, 12, 13$.

To evaluate accuracy of classification on the dataset, we divided it into training and testing data sets by ratio 8:2. The test data was normalized based on distribution parameters extracted from the training data set. We performed a 5-fold cross validation test on the training data to find the optimal parameters $\gamma$ and $C$. These optimal parameters were then used to build classifiers for the entire training dataset. Finally, the classifiers were applied on the testing dataset to get accuracy results. The aggregate function we used in experiments as shown in Equation (8). Basically, if more than or equal 50% of its frames are classified to a class label, the trial will be classified with same class label. This function equals to max function with additional constraint of at least 50% frames supporting the class label. To reduce randomness due to division of data into training and testing data, we ran this process five times. The reported accuracy results were calculated by taking average of accuracies of five times running this process.

V. RESULTS AND DISCUSSION

We conducted two experiments from the same dataset. The first experiment is with $CSP_{1vsN}$ method and its aggregate model. The second one is with $CSP_{pairs}$ method and its aggregate model. Fig. 2 and Fig. 3 show experimental results.

![Fig. 2. Classification results of CSP_{1vsN} and its aggregate model.](image)

![Fig. 3. Classification results of CSP_{pairs} and its aggregate model.](image)

It can be seen that in both experiments, aggregate models are better than or at least equal to their corresponding non-aggregate models at most subjects. The range of accuracy improvement can achieve at most at nearly 9% at some
subjects such as subject 9 for CSP\textsubscript{1vsN} method and subject 2 for CSP\textsubscript{pairs} method.

Another finding is that there is not much difference in classification between CSP\textsubscript{1vsN} and CSP\textsubscript{pairs} methods as shown in Fig. 4. The CSP\textsubscript{pairs} method is slightly better than the other for most subjects. It can be explained in the fact that the core of both methods are based on the same 2-class CSP analysis. Moreover, aggregate model seems to improve better in the CSP\textsubscript{pairs} method than in the CSP\textsubscript{1vsN} method.

Fig. 4. Classification results of CSP\textsubscript{1vsN} and CSP\textsubscript{pairs} methods.

The improvement in accuracy of our proposed aggregate model leads us to another question about whether or not aggregate function is necessary. We test this hypothesis by leaving framing technique and removing aggregate function. If this hypothesis is true, we can significantly reduce length of trial in future experiments. Unfortunately, experiment results as shown in Fig. 5 and Fig. 6 show that reducing time for a trial will reduce classification results for both CSP\textsubscript{1vsN} and CSP\textsubscript{pairs} methods.

Fig. 5. Classification results of CSP\textsubscript{1vsN} on trials and frames.

It is easy to see that our proposed model in Fig. 1 can be extended to handle online multi-class BCI systems with little modification. With predefined window size \(w\) and window step \(s\), we can extract frames at specific times and put them through trained classifiers. Based on classification results of the current frame and its previous frames, aggregate function can then be applied to show current mental activity. The delay time for such prediction might be equal to time defined by window step \(s\) given that processing time does not take much time as proven in our experiments. For some first frames, due to lacking of number of frames for aggregate function, some boundary handling techniques can be applied.

VI. CONCLUSIONS AND FUTURE WORK

We have proposed aggregate model as a novel method to improve classification accuracy of multi-class BCI systems. To the best of our knowledge, this is the first work following this direction. In theory, our work is supported by well-known issues about delay between stimulation and response, and subject concentration in BCI experiments. We conducted experiments on a well-known multi-class dataset. Comparing with the two current state-of-the-art methods based on CSP analysis: One-versus-The-Rest CSP (CSP\textsubscript{1vsN}) and pairwise CSP (CSP\textsubscript{pairs}), our proposed aggregate model for these two methods are better than the original ones for most of subjects. For some subjects, the accuracy improvement can achieve up to 9\%. Our model seems to improve better for the CSP\textsubscript{pairs} method than for the CSP\textsubscript{1vsN} method. Moreover, our proposed model can be easily extended to online BCI systems with very little modification. The delay for making decision in online BCI systems depends on time set at parameter window step \(s\) given that our model processing time is very fast.

The aggregate model has two parameters window size \(w\) and window step \(s\) which are very important to set trade off between delay issue and subject’s concentration issue. Further research on this direction is very useful not only for practical experiment design but also for understanding gap between portions of meaningful signal in a trial. The model is also can be extended to handle with problem of brain signal segmentation. Future work therefore will focus on these extensions and other popular multi-class BCI datasets.

REFERENCES


