ROBUST COMMON SPATIAL PATTERNS ESTIMATION USING DYNAMIC TIME WARPING TO IMPROVE BCI SYSTEMS

Ahmed M.Azab⋆†*, Lyudmila Mihaylova*⋆†*, Hamed Ahmadi*†*, and Mahnaz Arvaneh*⋆†

⋆†Automatic Control and System Engineering Department, University of Sheffield, UK
†School of Computer Science and Electronic Engineering, University of Essex, UK

ABSTRACT

Common spatial patterns (CSP) is one of the most popular feature extraction algorithms for brain-computer interfaces (BCI). However, CSP is known to be very sensitive to artifacts and prone to overfitting. This paper proposes a novel dynamic time warping (DTW)-based approach to improve CSP covariance matrix estimation and hence improve feature extraction. Dynamic time warping is widely used for finding an optimal alignment between two time-dependent signals under predefined conditions. The proposed approach reduces within class temporal variations and non-stationarity by aligning the training trials to the average of the trials from the same class. The proposed DTW-based CSP approach is applied to the support vector machines (SVM) classifier and evaluated using one of the publicly available motor imagery datasets. The results showed that the proposed approach, when compared to the classical CSP, improved the classification accuracy from 78% to 83% on average. Importantly, for some subjects, the improvement was around 10%.

***Index Terms*—** Brain computer interface (BCI), Common spatial pattern (CSP), Dynamic time warping (DTW)

1. INTRODUCTION

Brain-computer interface (BCI) provides a direct communication between a person's brain and an electronic device without the need for any muscle control [1]. Electroencephalogram (EEG) is the most widely used brain signals in BCI since it is measured non-invasively with a high temporal resolution [2]. In EEG-based BCI systems, user mental states are identified by classifying features extracted from EEG. Common Spatial patterns (CSP) is one of the most popular feature extraction methods for BCIs [3]. The importance of spatial filtering arises due to the poor spatial resolution of EEG measurements, such that the EEG pattern of interest is mixed with several irrelevant but concurrent neural activities. Using spatial filtering, signals from multiple electrodes are linearly combined to increase signal to noise ratio, leading to extract more discriminative EEG features [4].

Despite being popular and effective, CSP is known to be very sensitive to artifacts and prone to overfitting [5]. There are different reasons which can lead to poor CSP features. First, EEG signals are very likely to be affected by movement artifacts and blinking during recording. Hence, it is very challenging to estimate the probability distributions for high dimensional noisy EEG signals where outliers will have significant negative effects, specially for a small training set. Second, EEG is highly non-stationary which may happen due to several factors such as the variations of users' mental states, miss concentration and fatigue, which will also lead to inaccurate CSP features.

Different algorithms have been proposed to address CSP drawbacks and improve its learning process. CSP improvement can be done either at covariance matrix estimation level [6, 7, 8] or at CSP optimization objective function level [9, 10, 11, 12]. In this paper, we are interested in methods that focus on improving covariance matrix estimation. Lu et al. in [7] improved CSP covariance matrix estimations by adding two regularization terms to increase the estimation stability. The first regularization parameter controls the degree of shrinkage of the training sample covariance matrix estimation to the generic covariance matrix estimation. Whereas, the second regularization parameter controls the degree of shrinkage toward the identity matrix. In [8], the estimation of the covariance matrix was improved by using covariance matrices of other subjects with large number of trials when only few trials from the new subject are available . The above-mentioned improved CSP algorithms have all been shown to outperform the classical CSP. However, most of the aforementioned algorithms require calculating a number of regularization parameters which is computationally expensive. Moreover, none of these algorithms consider the temporal variations between trials to reduce within session non-stationarity.

To overcome the above-mentioned drawbacks, this paper proposes a novel DTW-based CSP approach to improve CSP covariance matrix estimation, and hence improves feature extraction. To cope with the temporal non-stationarity of EEG signals, we hypothesize that the alignment of EEG trials from the same class to their average might reduce within class temporal variation and non-stationarity. Following the previous assumption, the available trials for each class are aligned to the calculated average signal of this class aiming at minimizing the non-stationarity between the available EEG trials. Us-

978-1-5386-4658-8/18/$31.00 ©2019 IEEE

ing DTW, the available trials from the same class get as close as possible to the mean of this class and also to each other. The new aligned trials are used to calculate the CSP covariance matrices. Based on our knowledge this is the first time for DTW to be used with BCI and in a such way.

The proposed approach is evaluated using one of the publicly available datasets with a moderate number of subjects. Performance of the proposed approach is also compared with the results of the classical CSP.

The remainder of this paper is organized as follows. In Section II, we will describe the baseline approach and our proposed approach. Data description and results are discussed and analyzed in Section III. Finally, conclusions and future work are drawn in Section IV.

2. METHODOLOGY

2.1. Common Spatial Patterns (CSP)

CSP linearly transforms the data from the original EEG-channels into new channels to better differentiate between two conditions by maximizing the variance of one condition while minimizing it for the other. The CSP filters are calculated based assigning a new weight for each channel depend on the projection matrix. This projection matrix will have as many filters as the number of channels and the columns of the matrix will carry the weights to make linear combinations of the original EEG channels to decide which EEG-channels carry the most useful information. The first half of the projection matrix will maximize the variance for class one and minimize it for class two, while the second half of the projection matrix will maximize the variance for class two and minimize it for class one under the assumption that the signal is band-pass filtered [13]. Based on the amount of features needed an amount of CSP-channels, also called filter pairs, are selected. The following equations show how feature extraction based on CSP works.

Let us consider, $\mathbf{X}_i \epsilon\ n\ \times t$ is the i^{th} band passed signal trial and $\mathbf{Z}\epsilon\ t\ \times n$ is the signal after spatial filtration with $\mathbf{W}\epsilon \mathrm{n} \times n$ projection matrix of CSP.

$$\mathbf{Z} = \mathbf{X}_i^T \mathbf{W}, \tag{1}$$

here, for each trial n and t are the numbers of EEG channel and time instants respectively. Let $\mathbf{C_1} \epsilon R_n \times n$ and $\mathbf{C_2} \epsilon R_n \times n$ are covariance matrix of EEG signal X for the two classes. $\mathbf{C_1}$ and $\mathbf{C_2}$ can be computed by [14]:

$$\mathbf{C_{(c)}} = \frac{1}{n_c} \sum_{i \epsilon I_c} \mathbf{X}_i \times \mathbf{X}_i^T, \qquad c = [1, 2] \tag{2}$$

here, all trials corresponding to class c are denoted by I_c , and the total number of trials for each class c is n_c. The CSP filter matrix $\mathbf{W}$ can be computed by:

$$\mathbf{C_1} \times \mathbf{W} = (\mathbf{C_1} + \mathbf{C_2}) \times \mathbf{WD}, \tag{3}$$

where, eigenvalues for $\mathbf{C_1}$ formed the $\mathbf{D}$ diagonal matrix. Normally, classification is done using m pairs of filters from $\mathbf{W}$. In this paper, we use the first three and last three rows of $\mathbf{W}$ to acquire spatial filtered signal $\mathbf{Z}^* \epsilon\ t\ \times m$ [14].

$$\mathbf{Z}^* = \mathbf{X}_i^T \mathbf{W}^*. \tag{4}$$

However, if all EEG-data points would be used, the dimensionality of the data would be too high to be used by the classifier. Therefore, the most relevant features are extracted so the feature vector $\mathbf{F} \epsilon R^{2m}$ can be computed by calculating logarithm of variance of $\mathbf{Z}$ [14].

$$\mathbf{F} = log(var(\mathbf{Z}^*)). \tag{5}$$

These features are used as the input to train the classifier, and hence the trained classifier is used to estimate the labels of unlabeled trials.

2.2. Dynamic Time Warping (DTW)

Dynamic time warping (DTW) was initially proposed to solve the time deformation problem between two patterns in speech recognition problems [15]. Subsequently, DTW has been applied to other problems such as object recognition, classification and clustering of time domain signals such as EEG, subject identification, and motion analysis [16]. For EEG, DTW is typically used as a measure of dissimilarity between two patterns after being optimally aligned [17]. In this paper, we use DTW in a different way. As opposed to the previous applications of DTW where the goal was to find the distance between two trials, in our proposed approach, we align a collection of measured trials from one class to the average of these trials. Based on our knowledge this is the first time where DTW is used in such a way with BCI.

Here, we firstly explain DTW algorithm and how it generates a pair of aligned responses, and then we present our proposed DTW-based CSP. Suppose we have two time series Q and C, of length n and m respectively, where:

$$Q = [q_1, q_2, ..., q_i, ..., q_n] \tag{6}$$

$$C = [c_1, c_2, ..., c_j, ..., c_m] \tag{7}$$

To align two sequences using DTW we construct an $n-by-m$ matrix where the (i_{th}, j_{th}) element of the matrix contains the distance $d(q_i, c_j)$ between the two points q_i and c_j using the Euclidean distance.

$$d(q_i, c_j) = (q_i - c_j), \tag{8}$$

each matrix element (i, j) corresponds to the alignment between the points q_i and c_j. A warping path $\mathbf{W}$, is a matrix whose elements defines a mapping between Q and C. The k_{th} element of $\mathbf{W}$ is defined as: $w_k = (i, j)_k$ so we have:

$$\mathbf{W} = [w_1, .., w_k, ..., w_K] \quad max(m, n) \leq K < m + n - 1 \tag{9}$$

The warping path is typically subject to the following constraints:
1- Boundary conditions: $w_1 = (1, 1)$ and $w_K = (m, n)$.
2- Continuity: Given $w_k = (a, b)$ then $w_{k-1} = (a', b')$ where $a - a' \leq 1$ and $b - b' \leq 1$.
3- Monotonicity: Given $w_k = (a, b)$ then $w_{k-1} = (a', b')$ where $a - a' \geq 0$ and $b - b' \geq 0$.
There are exponentially many warping paths that satisfy these conditions. However we are interested in the path which minimizes the warping cost. This path can be found using dynamic programming to evaluate the following recurrence which defines the cumulative distance $\gamma(i, j)$ as the distance $d(i, j)$ found in the current cell and the minimum of the cumulative distances of the adjacent elements:

$$\gamma(i,j) = d(q_i, c_j) + min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \tag{10}$$

2.3. Proposed DTW-based CSP

In this paper, we assume that a large numbers of labeled trials EEG trials are available from each subject. The set of labeled EEG trials for each subject can be presented as $d = (\mathbf{x}^l, y^l)_{l=1}^{f}$, where f is the number of trials, and $\mathbf{x}^l$ and y^l respectively denote the instances matrix and the class label, $y^l \in \{0, 1\}$, of the l^{th} trial. Each trial is a subset of $R^{h \times v}$, where h is the number of instances and v is the number of channels per trial. Typically, the classifier is trained using the available labelled training features to predict the labels of the unlabeled trials. The commonly used BCI model uses CSP algorithm to extract features. Hence, in order to overcome the problem of non-robust CSP covariance matrices estimation, we propose a novel CSP algorithm using DTW. Our proposed algorithm aligns the available trials from each class to be as much similar to their average. Performing the proposed alignment leads to create new training trials that are less dissimilar in temporal domain.

At first the average of the available trials per each class c is computed as follows:

$$\mathbf{t_c} = (1/n_c) \sum_{l=1}^{n_c} \mathbf{x}^l, \tag{11}$$

where c refers to the class label, $\mathbf{t_c}$ refers to the average of the available trials of class c, and n_c is the number of trials available from class c. After that a similarity matrix between each available trial and the average signal from the same class is computed using (8). Then the warping path for these two trials is calculated in a way to minimize the following cost function under the constraints mentioned in subsection 2.2:

$$D(\mathbf{t_c}, \mathbf{x^l}) = min\big(1/k \sum_{i=1}^{k} d\{w_k\}\big), \tag{12}$$

where D is the accumulated distance between the average of class c and each individual trial from the same class. Then

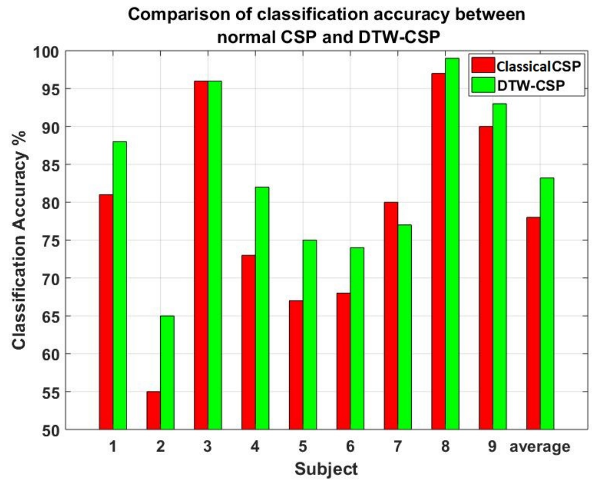

Fig. 1. Comparison of classification accuracy between classical CSP and DTW-based CSP. Interestingly, it shows that the proposed DTW-CSP algorithm outperform classical CSP for all subject except subject 7. Moreover on average classification accuracy the proposed algorithm is better than normal CSP by almost 5%.

the indices of the warping path that minimize the previous cost function for each available trial are used to construct the new aligned trial as follows:

$$\mathbf{x^l}_{aligned} = \{\mathbf{x}^l(s_1), ..., \mathbf{x}^l(s_k)\}, \tag{13}$$

where $[s_1, ..., s_k]$ are the indices of this trial instances that forming the minimum warping path calculated using (10) and satisfying (12). Those reflected instances are the instances that will make this trial to be as much similar to the reference average signal. Subsequently the covariance matrix of the new aligned raw EEG trail is calculated as follows:

$$\mathbf{\Sigma}_{i_{aligned}} = \frac{(\mathbf{x}^i_{aligned})(\mathbf{x}^i_{aligned})^T}{trace((\mathbf{x}^i_{aligned})(\mathbf{x}^i_{aligned}))}. \tag{14}$$

Finally, the average of the calculated covariance matrices of the aligned trials for each class c is computed as follows:

$$\mathbf{\Sigma}_c = (1/n_c) \sum_{i=1}^{n_c} \mathbf{\Sigma}_{i_{aligned}}. \tag{15}$$

This covariance matrix will replace the covariance matrix used to calculate normal CSP filters mentioned in (2) and hence continue till calculate the log variance features.

3. EXPERIMENTAL RESULTS

3.1. Data Description

In order to validate the proposed algorithm and compare it with the classical CSP algorithm, the both algorithms are applied to data set 2a BCI Competition IV 2008 [18]. This data

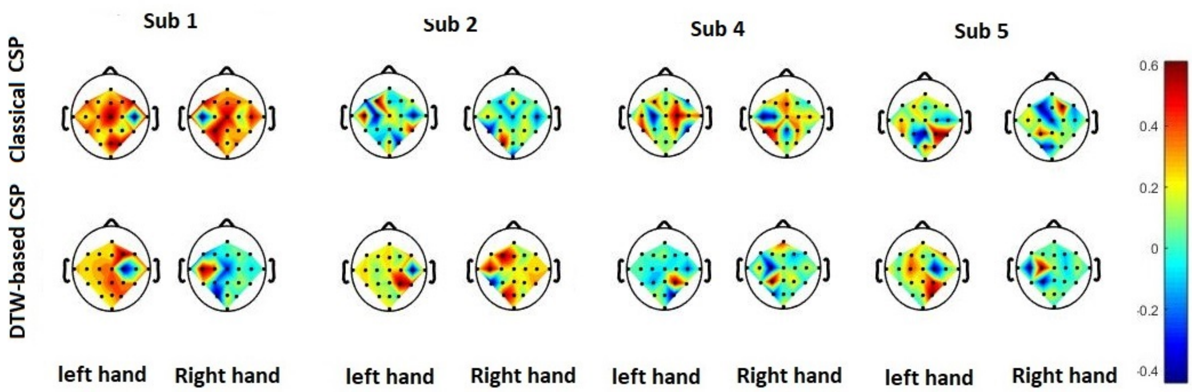

Fig. 2. Some examples of spatial filters obtained with classical CSP and DTW-CSP algorithms for different subjects. Interestingly, DTW-CSP filters are smoother and physiologically more relevant to the imagined hand. Contrary to classical CSP, DTW-CSP filters weights are more related to the expected motor cortex areas.

set consists of EEG data from 9 subjects performing 4 classes of motor imagery task. In this paper, only data from right and left hand motor imagery are used. Two sessions on different days were recorded for each subject. Each session is comprised of 6 runs, each run consists of 12 trials for each class.

EEG signals were recorded using 22 electrodes, sampled at 250 Hz, and bandpass-filtered between 0.5 Hz and 100 Hz. Moreover, a 50 Hz notch filter was applied to remove power line noise. Features were extracted using the time segment located from 0.5s to 2.5s after the cue instructing the subject to perform motor imagery (MI). The first session was used to train the BCI system once using the proposed DTW-CSP and once using classical CSP. Evaluation of classification accuracy of was done using the second session.

3.2. Evaluation and Discussion

For each subject, the CSP and the DTW-CSP filters were learnt on the available training set. The log-variances of the spatially filtered EEG signal were then used as the input features of a Support vector machines (SVM) classifier. The classification accuracy was calculated based on how accurately the labels of testing sessions trials are estimated. Fig.1 shows that, except subject 7, the DTW-CSP algorithm outperformed classical CSP. The DTW-CSP algorithm outperformed CSP by about 4% to 10% for each subject. On average, classification accuracy for all subject was increased fro 78% to 83.3%. These results confirm that using DTW reduces temporal variations and non-statinarties between trials within the same class, and hence enhance the computed features. Particularly, with a closer look at the results, suggests that the DTW-CSP algorithm is more valuable for subjects with poor and medium initial BCI performance (e.g., sub1, sub2, sub4, sub5, sub6) than subjects with initially high performance, whose performances are slightly changed. This finding makes sense as subjects with high initial accuracy already have their features well separated.

Fig. 2 shows some examples of the spatial filters obtained with classical CSP and DTW-CSP algorithms for different subjects. Notably, these pictures show that classical CSP filters appear with large weights in several unexpected locations from a neurophysiological point of view. On the other hand, DTW-CSP filters are interestingly smoother and physiologically more relevant to the imagined hand. Contrary to classical CSP, DTW-CSP filters weights are more related to the expected motor cortex areas. This is another benefit of the DTW-CSP algorithm as it is not only make the trials of the same class get closer but also lead to filters that are neurophysiologically smoother and as such more illustratable. Moreover, our approach requires much less computational time as there is no need to calculate any regularization parameters either using cross-validation or by optimizing objective functions which are computationally expensive.

4. CONCLUSION

In this paper, we proposed a novel DTW-based CSP algorithm to improve the BCI systems. The proposed algorithm was evaluated against the classical CSP. Results showed that the DTW-CSP outperformed classical CSP by almost 5% in average classification accuracy and lead to more relevant spatial filters from the neurological point of view.

Future work could deal with investigating performances of the proposed DTW-CSP algorithm with more datasets. Moreover, DTW-CSP can be addressed in the way of transfer learning to reduce BCI calibration times even for subject to subject transfer or session to session transfer. We could address this problem by aligning the few trials available from the new subject to a punch of trials previously recorded from different subjects or sessions.

5. REFERENCES

[1] Eleanor A Curran and Maria J Stokes, "Learning to control brain activity: a review of the production and control of EEG components for driving brain–computer interface (BCI) systems," *Brain and Cognition*, vol. 51, no. 3, pp. 326–336, 2003.

[2] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan, "Brain–computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[3] Herbert Ramoser, Johannes Muller-Gerking, and Gert Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[4] Benjamin Blankertz, Motoaki Kawanabe, Ryota Tomioka, Friederike Hohlefeld, Klaus-robert Müller, and Vadim V Nikulin, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," in *Advances in neural Information Processing Systems*, 2008, pp. 113–120.

[5] Wojciech Samek, Motoaki Kawanabe, and Klaus-Robert Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50–72, 2014.

[6] Hyohyeong Kang, Yunjun Nam, and Seungjin Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 683–686, 2009.

[7] Haiping Lu, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos, "Regularized common spatial patterns with generic learning for EEG signal classification," in *Annual International Conference of the IEEE,Engineering in Medicine and Biology Society, EMBC 2009.* IEEE, 2009, pp. 6599–6602.

[8] Fabien Lotte and Cuntai Guan, "Learning from other subjects helps reducing brain-computer interface calibration time," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 614–617.

[9] Fabien Lotte and Cuntai Guan, "Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.

[10] Wojciech Samek, Carmen Vidaurre, Klaus-Robert Müller, and Motoaki Kawanabe, "Stationary common spatial patterns for brain–computer interfacing," *Journal of Neural Engineering*, vol. 9, no. 2, pp. 026013, 2012.

[11] Mahnaz Arvaneh, Cuntai Guan, Kai Keng Ang, and Chai Quek, "Optimizing spatial filters by minimizing within-class dissimilarities in Electroencephalogram-based brain–computer interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 4, pp. 610–619, 2013.

[12] Fabien Lotte and Cuntai Guan, "Spatially regularized common spatial patterns for EEG classification," in *20th International Conference on Pattern Recognition (ICPR), 2010*. IEEE, 2010, pp. 3712–3715.

[13] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw, "Bci2000: a general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.

[14] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and K-R Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[15] Hiroaki Sakoe and Seibi Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[16] Pavel Senin, "Dynamic time warping algorithm review," *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, vol. 855, pp. 1–23, 2008.

[17] Takashi Yamauchi, Kunchen Xiao, Casady Bowman, and Abudullah Mueen, "Dynamic time warping: A single dry electrode EEG study in a self-paced learning task," in *International Conference on Affective Computing and Intelligent Interaction (ACII), 2015*. IEEE, 2015, pp. 56–62.

[18] C Brunner, R Leeb, G Müller-Putz, A Schlögl, and G Pfurtscheller, "BCI competition 2008–Graz data set A," .