Discriminative Channel Addition and Reduction for Filter Bank Common Spatial Pattern in Motor Imagery BCI

Zheng Yang Chin, Kai Keng Ang, Chuanchu Wang, Cuntai Guan

***Abstract*— An electroencephalography (EEG)-based Motor Imagery Brain-Computer Interface (MI-BCI) requires a long setup time if a large number of channels is used, and EEG from noisy or irrelevant channels may adversely affect the classification performance. To address this issue, this paper proposed 2 approaches to systematically select discriminative channels for EEG-based MI-BCI. The proposed Discriminative Channel Addition (DCA) approach and the Discriminative Channel Reduction (DCR) approach selects subject-specific discriminative channels by iteratively adding or removing channels based on the cross-validation classification accuracies obtained using the Filter Bank Common Spatial Pattern algorithm. The performances of the proposed approaches were evaluated on the BCI Competition IV Dataset 2a. The results on 2-class and 4-class MI data showed that DCA, which iteratively adds channels, selected 13~14 channels that consistently yielded better cross-validation accuracies on the training data and session-to-session transfer accuracies on the evaluation data compared to the use of a full 22-channel setup. Hence, this results in a reduced channel setup that could improve the classification accuracy of the MI-BCI after removing less discriminative channels.**

I. INTRODUCTION

For non-invasive Brain-Computer Interface (BCI), scalp electroencephalogram (EEG) signals are often used due to its relative ease of setting up and fine temporal resolution [1-3]. This allows a subject to use his brain signals for communication and controlling external devices. For a Motor Imagery-based BCI (MI-BCI), the EEG is recorded from multiple electrode sites [2],[4],[5]. The multi-channel EEG data is then processed and classified into different types of motor imagery such as the imagination of movement of the left hand, right hand, foot or tongue [6]. The translated signal could be used to control the movement of a robotic arm for MI-BCI stroke rehabilitation [7].

The multi-channel EEG could be processed using various signal processing and machine learning methods, such as the Common Spatial Pattern (CSP) algorithm [4] and the Filter Bank Common Spatial Pattern (FBCSP) algorithm [8], where the latter performs a multi-stage process of temporal filtering, CSP feature extraction and selection. However, multi-channel EEG recordings requires a long setup time, due to the preparation of conductive gel into the wet electrodes resulting in inconvenience on the subject [9]. On the other hand, a small number of channels may be subject to the risk of outliers and artifacts, adversely affecting statistical estimation [5]. Hence, various studies have proposed computational methods to reduce the number of channels used in 2-class motor imagery [3],[5],[10],[11], as well as multi-class motor imagery, where channels were manually selected [2] or were selected by analyzing Event-Related Desynchronization (ERD) / Event-Related Synchronization (ERS) effects [12]. These methods investigated the tradeoff between the number of channels used and classification performance. Experimental results show that recordings from multiple channels may include irrelevant data such as noisy electrode readings that could deteriorate classification accuracy [10],[11], and removing them could improve or maintain classification accuracy.

Given these issues related to a multi-channel EEG setup in a MI-BCI, this paper proposes 2 approaches to systematically select discriminative channels for a reduced channel setup in a 2-class and in a multiclass (4-class) MI-BCI. The proposed Discriminative Channel Addition approach and the Discriminative Channel Reduction approach, employs a subject-specific approach to add or reduce channels iteratively based on the cross-validation accuracies obtained using the FBCSP algorithm. The FBCSP algorithm has been shown to be effective not only on 4-class motor imagery (left hand, right hand, foot and tongue) on a 22-monopolar channel setup, but also effective on 2-class motor imagery (left hand, right hand) using a reduced 3-bipolar channel setup [8]. Hence, a reduced channel setup computed using the proposed approaches and the FBCSP algorithm, that could potentially remove irrelevant channels and improve classification accuracy, will be investigated in this paper.

II. MATERIALS & METHODS

A. Experimental Data

The proposed approaches were evaluated on the BCI Competition IV dataset 2a [13], where 1 training session and 1 evaluation session of EEG data from 9 subjects are provided. Each session consisted of 288 single trials, equally distributed between left hand, right hand, foot and tongue motor imagery. Figure 1 shows the structure of each trial. 22 electrodes were used to record the EEG and the montage is shown in Figure 2. The segment of 2.5s to 4.5s of EEG data after the start of each trial was used to train the FBCSP

This work was supported by the Science and Engineering Research Council of A*STAR (Agency for Science, Technology and Research), Singapore.

Z. Y. Chin, K. K. Ang, C. Wang and C. Guan are with Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way #21-01 Connexis (South Tower) Singapore 138632 (email: {zychin, kkang, ccwang, ctguan}@i2r.a-star.edu.sg).

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

algorithm. More details of the protocol are available in [6] and [13].

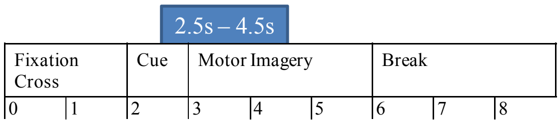

Figure 1 shows the protocol of a single trial of motor imagery. The time segment of 2.5s to 4.5s after the start of a trial was used to train the FBCSP algorithm

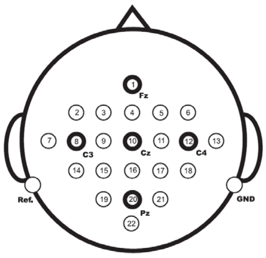

Figure 2 shows the 22-channel electrode layout for the BCI Competition IV Dataset 2a

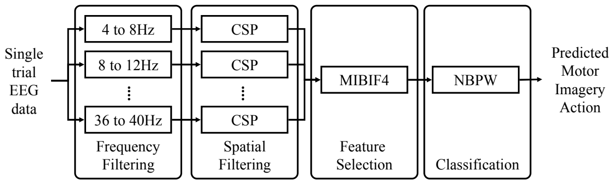

Figure 3 shows the architecture of the Filter Bank Common Spatial Pattern (FBCSP) algorithm for 2-class motor imagery EEG data. MIBIF4 and NBPW represent the Mutual Information Best Individual Feature and the Naïve Bayes Parzen Window classifier respectively

B. *Filter Bank Common Spatial Pattern (FBCSP)*

A brief description of the FBCSP algorithm and its multi-class extension is provided as follows. Readers are invited to refer to [8] for a more extensive description of the FBCSP algorithm. The FBCSP algorithm comprises 4 stages that perform an autonomous selection of subject-specific temporal-spatial discriminative EEG characteristics for 2-class MI-BCI, shown in Figure 3.

The 1st stage performs frequency filtering and artifact removal using a filter bank that decomposes the EEG measurements into 9 pass-bands from 4-8Hz, 8-12Hz…36-40Hz. Next, spatial filtering is performed in the 2nd stage by linearly transforming the EEG data using the CSP algorithm [14] that extracts m pairs of CSP features per filter band, to form the following feature vector for the i^{th} trial,

$$\mathbf{x}_i = \left[\mathbf{cf}_1, \mathbf{cf}_2, \ldots, \mathbf{cf}_9\right], \tag{1}$$

where $\mathbf{cf}_b \in \mathbb{R}^{2m}$ denotes the m pairs of CSP features for the b^{th} band-pass filtered EEG measurements, $\mathbf{x}_i \in \mathbb{R}^{1\times(9*2m)}$.

The 3rd stage performs feature selection of the extracted features using the Mutual Information Best Individual Features (MIBIF) algorithm. This algorithm selects the best k=4 features sorted by mutual information with the class label in descending order. Since CSP features are paired, the corresponding CSP features which come in pairs with the selected k features are also selected. Finally, the 4th stage performs classification using the Naïve Bayes Parzen Window (NBPW) Classifier and the classification rule is given as

$$\omega = \underset{\omega=1,2}{\arg\max}\, p\left(\omega \mid \mathbf{x}\right), \tag{2}$$

where $p(\omega|\mathbf{x})$ denotes posterior probability of the class being ω=1, 2, given the random trial $\mathbf{x} = \left[x_1, x_2, \ldots x_d\right]$ and d denotes the number of selected features from the third stage.

C. *One-Versus-Rest (OVR) Multi-class Extension*

The One-Versus-Rest (OVR) multi-class extension employs component binary NBPW classifiers which classifies CSP features that discriminates each class from all the other classes. For a 4-class MI-BCI, 4 OVR classifiers are required. The classification rule of the classifier is extended from (2) to

$$\omega = \underset{\omega=1,2,3,4}{\arg\max}\, p_{\mathrm{OVR}}\left(\omega \mid \mathbf{x}\right), \tag{3}$$

where $p_{\mathrm{OVR}}\left(\omega \mid \mathbf{x}\right)$ is the probability of classifying $\mathbf{x}$ between ω and ω'={1,2,3,4}\$\omega$; and \ denotes the set theoretic complement operation.

D. *Proposed Approaches for a Reduced Channel Setup*

The proposed approaches adopt a similar classical method in feature selection which evaluates growing feature sets (forward selection) or evaluates shrinking feature sets (backward selection) [15]. The descriptions of the two proposed approaches are as follows:

1) *Discriminative Channel Reduction (DCR)*

This approach starts from a full channel setup, with individual channels removed iteratively based on cross validation accuracies

- Step 1: Initialization

Let the set of all channels be represented by A = {1,…,n} where each element represents the respective channel number. |A| = n = 22. Denote the set of channels used in the MI-BCI as D = A initially.

- Step 2: Cross-validation with each channel removed

Compute the 10-fold cross validation accuracies $J\left(D \setminus j\right)$ with only channel $j \in D$ is removed. Select the channel k where

$$J\left(D \setminus k\right) = \max_{j\in D} J\left(D \setminus j\right). \tag{4}$$

- Step 3: Remove channel k

Update $D = D \setminus k$. Hence, this approach assumes the classification accuracy will improve or will be least adversely affected, when removing a noisy or irrelevant channel. Repeat Step 2 and Step 3 until |D| = 3 which is the minimum number of channels required for the CSP algorithm.

2) *Discriminative Channel Addition (DCA)*

This approach starts from a 3-channel setup, with individual channels added iteratively based on cross validation accuracies.

- Step 1: Initialization

Initialize set of selected channels D = {8,10,12} which represents C3, C4, Cz and $D' = A \setminus D$.

- Step 2: Cross-validation with each channel added

Compute the 10-fold cross validation accuracies with channel $j \in D'$ added. Select the channel k where

$$J\left(D \cup k\right) = \max_{j \in D'} J\left(D \cup j\right), \tag{5}$$

and $\cup$ is the set-theoretic union operation.

- Step 3:Add channel k

Update $D = D \cup k$ and $D' = A \setminus D$. Hence, this approach assumes the classification accuracy is improved the most when a relevant channel is included. Repeat Step 2 and Step 3 until $D = A$.

III. Experimental Results and Discussion

Subject-specific reduced channel configurations were computed using the proposed approaches only on the training session data and cross-validation results are presented. Session-to-session transfer results of the FBCSP algorithm, trained using the reduced channel setups on the training session data, and tested on the evaluation session data are also presented. The evaluation criteria is the classification accuracy for 10-fold cross-validation results on the training data session and maximum Kappa value [8] for the session-to-session transfer performance.

Manually selected channel subsets [2] are also included to compare performance. These subsets include (a) 20 channels {2,…,21}, (b) 17 channels {2,…18} and (c) 13 channels {2,4,6,…,14,16,18}. This paper also manually selected the smaller channel subsets around the sensorimotor cortex (d) 11 channels {2,4,6,7,8,10,12,13,14,16,18} (e) 9 channels {2,4,6,8,10,12,14,16,18} (f) 6 channels {2,4,6,14,16,18} and (g) 3 channels {8,10,12}. For 3 channel configurations, m = 1 pair of CSP features is extracted per filter band, while for > 3 channels, m = 2 pairs of CSP features are extracted per filter band [8].

A. Classification Results

The classification results versus the number of channels are presented in Figure 4 for 2-class and 4-class motor imagery respectively. As a reference for 2-class motor imagery, the 22-channel setup yielded 84.57% cross-validation accuracy and kappa = 0.70. For 4-class motor imagery, the 22-channel setup yielded 71.45% cross-validation accuracy and kappa = 0.58.

A deterioration in cross-validation accuracies below that of the 22-channel setup occurs in both proposed approaches for <6 channels. For ≥6 channels, both proposed approaches yielded higher cross-validation accuracies than the manually selected channels and the 22-channel setup. However, this was not consistently translated to session-to-session transfer performance; the manually selected channels yielded the best kappa of 0.71 and 0.72 at 13 and 17 channels in session-to-session transfer, even though its cross-validation performance was actually below that of the 22-channel setup and the proposed approaches. Note that at 13 channels, DCA and DCR yielded kappa = 0.70 and 0.69 respectively. Nevertheless, in terms of overall consistency, DCA achieved similar or better cross-validation and session-to-session transfer results compared to the manually selected channels and the 22-channel setups for ≥ 13 channels.

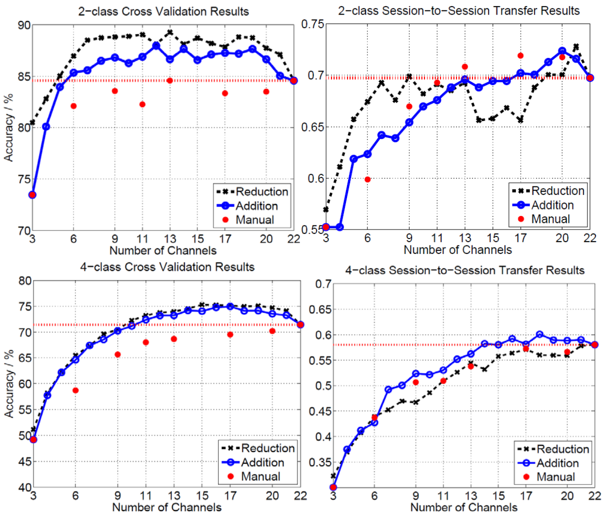

Figure 4 shows 2-class motor imagery mean accuracy results (top left: cross-validation, top right: session-to-session transfer), and 4-class motor imagery mean accuracy results (bottom left: cross-validation, bottom right: session-to-session transfer) on the BCI Competition IV Dataset 2a. The red horizontal line represents the accuracy results of the 22-channel setup.

4-class cross-validation accuracies for both proposed approaches were comparable to each other and were higher than the 22-channel setup for ≥11 channels. However, a drop in accuracies was observed for <11 channels. Similar to 2-class motor imagery, DCA was again shown to be consistent and relatively the best in both cross-validation accuracy and session-to-session transfer results for ≥14 channels where it could yield a cross-validation accuracy of >71.45% and session-to-session transfer kappa ≥0.58. Note that at 13 channels, DCA, DCR and Manual Selection yielded kappa = 0.56, 0.54 and 0.54 respectively.

B. Absolute Frequency of Selected Channels

To illustrate the channels selected, the number of times or absolute frequency of the channels selected in the 9 subjects for both approaches at $|D|$ = 13 is shown for 2-class and 4-class motor imagery in Figure 5 and Figure 6 respectively.

For 2-class motor imagery, both proposed approaches employ mostly channels around the sensorimotor area (C3 and C4). Although DCR did not impose the condition that C3, C4 must be retained, these 2 channels were still selected in a majority of the subjects. DCA uses C3, Cz and C4 as its initial set of channels; hence this could be why Channels 2 to 18 were mostly selected too. For 4-class motor imagery, sensorimotor channels 7 to 13 were selected in a majority of the subjects. Unlike the manually selected 13-channel configuration, Channel 19, 20, 21 in the parietal region and Channel 22 in the occipital region were selected in ≥ 4 subjects. Existing studies have also found that these channels were also selected in their reduced channel setups for motor imagery [3],[5],[12]. As noted in one study[5]: peripheral channels do not contain the most useful information but they could help improve classification.

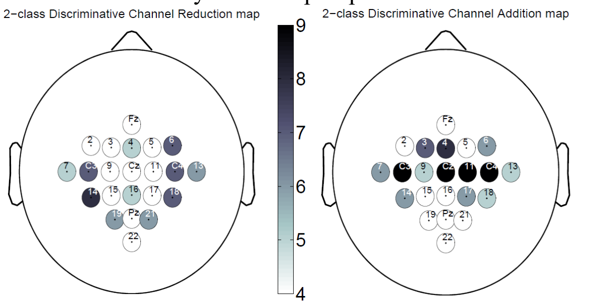

Figure 5 Absolute frequency map of selected channels of Discriminative Channel Reduction (left) and Discriminative Channel Addition (right) for 2-class motor imagery at 13 channels. Channels which have been selected in ≥ 4 subjects are shaded darker as indicated by the color bar.

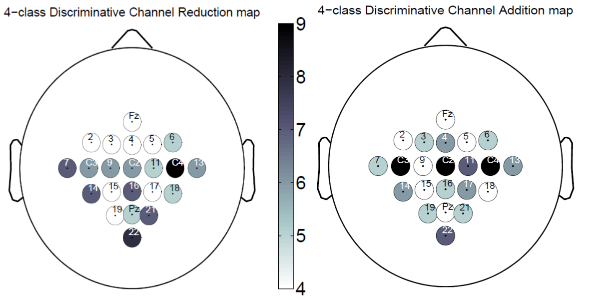

Figure 6 Absolute frequency map of selected channels of Discriminative Channel Reduction (left) and Discriminative Channel Addition (right) for 4-class motor imagery at 13 channels.

A limitation in this study is that cross-validation is heuristic and need not lead to improved classifiers [16]. For example, the manually selected 17-channels yielded lower cross-validation accuracy than the 22-channel setup for 2-class data in Figure 4 but it performed better during session-to-session transfer. Nevertheless, cross-validation can improve generalization accuracy for many real-world problems [16]. Also, other factors such as the type of classifier and initial set of electrodes employed may affect the performance of these methods. Hence the influence of these factors could be further investigated in future.

IV. CONCLUSION

This paper proposed 2 approaches to compute a subject-specific reduced channel subset for 2-class and 4-class MI-BCI using the Filter Bank Common Spatial Pattern algorithm. The proposed Discriminative Channel Reduction (DCR) removes channels iteratively starting from all the channels, based on the cross-validation accuracies. Discriminative Channel Addition (DCA) adds individual channels iteratively to an initial set of 3 channels located around the sensorimotor cortex, based on the cross-validation accuracies. In both 2-class and 4-class MI from the BCI Competition Dataset 2a, DCA, with a minimum of 13 to 14 channels, consistently yielded a cross-validation accuracy and session-to-session transfer performance which is similar or better than that of the full channel setup, DCR or manually selected channels. The results of the proposed approaches concurred with observations from existing studies [3],[10],[11]: when noisy or irrelevant channels were removed, classification results improved, but when too many channels were removed, performance deteriorated compared to using all the channels. Future work will investigate the proposed approaches on more datasets to further validate their performance.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, et al., "Brain Computer interfaces for communication and control," *Clin Neurophysiol.,* vol. 113, 2002.

[2] M. Naeem, C. Brunner, and G. Pfurtscheller, "Dimensionality Reduction and Channel Selection of Motor Imagery Electroencephalographic Data," *Computational Intelligence and Neuroscience,* vol. 2009, 2009.

[3] L. He, Y. Hu, Y. Li, et al., "Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG," *Neurocomputing,* vol. 121, pp. 423-433, 2013.

[4] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans Rehabil Eng.,* vol. 8, pp. 441-446, 2000.

[5] C. Sannelli, T. Dickhaus, S. Halder, et al., "On Optimal Channel Configurations for SMR-based Brain–Computer Interfaces," *Brain Topography,* vol. 23, pp. 186-193, 2010.

[6] C. Brunner, M. Naeem, R. Leeb, et al., "Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis," *Pattern Recogn. Lett.,* vol. 28, pp. 957-964, 2007.

[7] K. K. Ang, C. Guan, K. S. G. Chua, et al., "A Large Clinical Study on the Ability of Stroke Patients to Use an EEG-Based Motor Imagery Brain-Computer Interface," *Clinical EEG and Neuroscience,* vol. 42, pp. 253-258, 2011.

[8] K. K. Ang, Z. Y. Chin, C. Wang, et al., "Filter Bank Common Spatial Pattern algorithm on BCI Competition IV Datasets 2a and 2b," *Frontiers in Neuroscience,* vol. 6, 2012.

[9] B. Blankertz, F. Losch, M. Krauledat, et al., "The Berlin Brain-Computer Interface: Accurate Performance From First-Session in BCI-Naive Subjects," *IEEE Trans. Biomed. Eng.,* vol. 55, pp. 2452-2462, 2008.

[10] T. N. Lal, M. Schroder, T. Hinterberger, et al., "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.,* vol. 51, pp. 1003-1010, 2004.

[11] M. Arvaneh, G. Cuntai, A. Kai Keng, et al., "Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI," *IEEE Trans. Biomed. Eng.,* vol. 58, pp. 1865-1873, 2011.

[12] D. Wang, D. Miao, and G. Blohm, "Multi-class motor imagery EEG decoding for brain-computer interfaces," *Frontiers in Neuroscience,* vol. 6, 2012.

[13] M. Tangermann, K.-R. Müller, A. Aertsen, et al., "Review of the BCI Competition IV," *Frontiers in Neuroscience,* vol. 6, 2012.

[14] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans Rehabil Eng,* vol. 8, pp. 441 - 446, 2000.

[15] A. K. Jain, R. P. W. Duin, and M. Jianchang, "Statistical pattern recognition: a review," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 22, pp. 4-37, 2000.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Second Edition ed.: John Wiley & Sons, Inc, 2001.