Selection of Effective EEG Channels in Brain Computer Interfaces based on Inconsistencies of Classifiers

Huijuan Yang, Cuntai Guan, Kai Keng Ang, Kok Soon Phua and Chuanchu Wang

Abstract— **This paper proposed a novel method to select the effective Electroencephalography (EEG) channels for the motor imagery tasks based on the inconsistencies from multiple classifiers. The inconsistency criterion for channel selection was designed based on the fluctuation of the classification accuracies among different classifiers when the noisy channels were included. These noisy channels were then identified and removed till a required number of channels was selected or a predefined classification accuracy with reference to baseline was obtained. Experiments conducted on a data set of 13 healthy subjects performing hand grasping and idle revealed that the EEG channels from the motor area were most frequently selected. Furthermore, the mean increases of 4.07%, 3.10% and 1.77% of the averaged accuracies in comparison with the four existing channel selection methods were achieved for the non-feedback, feedback and calibration sessions, respectively, by selecting as low as seven channels. These results further validated the effectiveness of our proposed method.**

I. Introduction

How to select the subject-specific, most effective EEG channels for the classification of motor imagery EEG signals is an important topic due to the noisiness of some channels. Channel selection could be an effective way to improve EEG signal quality by removing those noisy, irrelevant channels. Two obvious advantages of channel selection can be visualized. Firstly, selecting a smaller number of channels will lower down the system cost due to the expensiveness of electrodes, especially for dry electrodes. Secondly, selecting the subject-specific, informative channels will help identify the motor cortex that are correlated well with the performed motor imagery tasks. This is especially important for the stroke patients [1], who generally have lesions at a particular side of the brain, and the lesion locations are unknown before the experiments. The activation or motor imagery-related sites could be different from that of healthy subjects.

Selection of EEG channels could be treated as a feature selection problem [2], [3], e.g., wrapper-based or filter-based approaches. Feature selection using wrapper-based approaches was usually coupled with a specific classifier such as support vector machines (SVMs) [4], [5]. Its performance depended to a large extent on the applied classifier. Further, the wrapper-based approaches generally incurred high computational load due to multiple iterations of training and classifying processes. On the other hand, filter-based approaches were classifier-independent and less computational-intensive. For example, the mutual information (MI)-based method selected the channels by computing the MI between the features and class labels [6]. However, the filter-based approaches could suffer the non-optimal selection problem for a subset of features, even though each individual feature was the best [5].

The spatial pattern coefficients in the Common Spatial Pattern (CSP)-based method [7], [8] were used to select the channels, which were known to be sensitive to noise or artifacts. While regularized CSP selected the channel by sparsifying the representation of CSP [9], [10]. Along the similar lines, the sparse CSP-based channel selection selected the channels by sparsifying the CSP projection matrix [3]. In this paper, we proposed a novel channel selection method by measuring the inconsistencies from the outputs of the multiple classifiers (CS-IMC), such that an optimal set of channels for the classification of motor imagery of hand grasping EEG signals from background idling state can be selected. Comparisons with existing channel selection methods demonstrated the advantages of our proposed method, e.g., no prior knowledge on the activation region was required, lower computational complexity, better classification accuracy by selecting the same number of channels.

II. Overall Scheme and Feature Extraction

Typically, channel selection could work towards two opposite directions based on predefined criteria: a) to select the *most effective* channels one by one; b) to eliminate those *noisy* channels one by one, the final selected channels would be those leftovers. The selecting or eliminating process would be iterated till a predefined stopping criterion was satisfied, e.g., the required number of channels was selected, or a predefined classification accuracy was reached compared with the baseline accuracy by using all channels. Our proposed method fell into the second category, i.e., those noisy channels were identified based on the inconsistencies of the outputs from multiple classifiers, which were then removed one by one until the required number of selected channels was obtained. An overall schematic diagram illustrating our proposed channel selection method was shown in Fig. 1. The performance was evaluated by an $n \times k$ fold cross-validation (CV) using *LibSVM* [11] as the classifier. Observed from the figure, our proposed channel selection method consisted of: a) *best frequency bands selection* module; b) *refilter and generate covariance features* module; c) *channel selection* module; and d) *performance evaluation* module. In what follows, each module would be described in more details.

The authors are with Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore 138632. Email: {hjyang, ctguan, kkang, ksphua and ccwang}@i2r.a-star.edu.sg

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

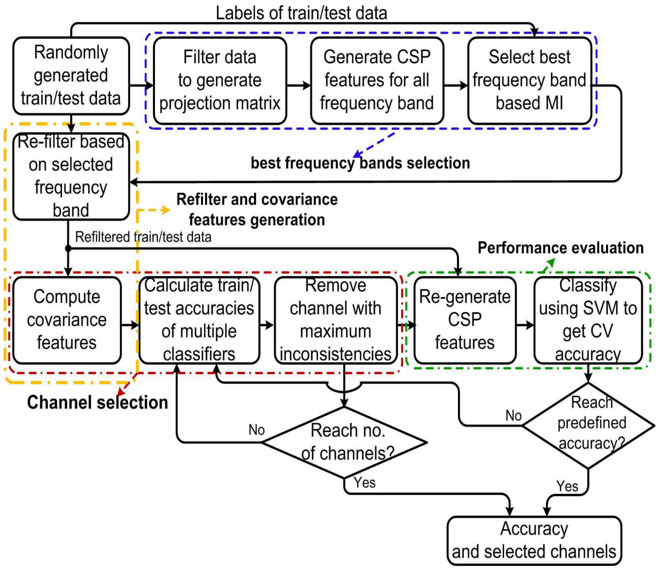


Fig. 1. An overall schematic illustration of the proposed channel selection by measuring the inconsistencies from multiple classifiers.

1) *Best frequency bands selection* module. The best frequency bands were selected based on the MI between the CSP features and class labels [12]. Both training and testing EEG data were filtered by Chebyshev filters for the frequency bands varying from 4Hz to 40Hz, with the bandwidth of each frequency band of 4Hz. The CSP projection matrix was obtained by maximizing the variance of the filtered signal under one condition while minimizing it for another condition [7], [13]. The spatial filter was formed by the first and last two columns from the projection matrix, which was then used to filter the signal. The log variance of the spatially filtered signal was used as the CSP features. The most discriminant frequency band was finally selected by maximizing the MI between the CSP features and class labels [12].

2) *Refilter and covariance features generation* module. The train and test set of EEG signals were refiltered based on the selected frequency bands using the nth order digital elliptic filter, with the resultant signal being denoted as S_{rt} and S_{rs}, respectively. Subsequently, the covariance features consisting of the variances for each column of the covariance matrices for refiltered train and test data were generated by

$$F^c_{tr} = diag(S_{rt}S^T_{rt}) \quad (1)$$
$$F^c_{te} = diag(S_{rs}S^T_{rs}) \quad (2)$$

where $diag(X)$ and X^T denoted the diagonals, and transpose of matrix X, respectively.

3) *Channel selection* module. a) Each channel was eliminated based on our proposed "channel selection based on the inconsistencies from multiple classifiers (CS-IMC)" criterion using the features of F^c_{tr} and F^c_{te} for training and testing data. The details of CS-IMC will be described in the next section. b) Updated the eliminated (c_e) and remaining (c_r) sets of channels for next iteration by c_e=$c_e \cup c_i$ and c_r=$c_r \backslash c_i$, where c_i was the current eliminated channel, "$\cup$" and "$\backslash$" represented "set union" and "set minus", respectively. c) Iterated steps a) to b) till the stopping criterion was satisfied.

4) *Performance evaluation* module. The performance was evaluated based on the remaining channels for current iteration. The CSP features were recomputed based on remaining channels c_r, i.e., $F_{csp}(c_r)$, which were then used in classification to obtain: $C_a(c_r) = \mathcal{H}_s(F_{csp}(c_r), L_b)$, where $\mathcal{H}_s$ and L_b denoted the classifier such as SVM, and class labels.

III. Channel Selection

Our idea was motivated by the observations that the features from the noisy channels did not correlate well with the corresponding motor imagery tasks. Hence, the outputs from multiple classifiers would appear to be unstable by classifying the features computed with the inclusion of these noisy channels. This motivated us to design the proposed channel selection method by eliminating one channel at a time based on the inconsistencies of the outputs from multiple classifiers, which was detailed in Table I. The inputs to the channel selection module were the refiltered training and testing EEG data for current fold. The classifiers employed were: SVM, naive bayes classifier, linear discriminant analysis and decision trees, which were denoted as $\mathcal{H}_s$, $\mathcal{H}_b$, $\mathcal{H}_l$ and $\mathcal{H}_d$, respectively. It should be noted that the finally eliminated and kept channels were based on the distribution analysis, e.g., computing histogram for each channel for a total of $n{\times}k$ fold. Similarly, the averaged classification accuracies for the elimination of each channel was obtained by averaging the accuracies for the $n{\times}k$ fold.

IV. Experimental Results

A. Data sets and data acquisition

EEG data were collected from 13 healthy subjects. Ethics approval was obtained from the institutional review board and informed consents were obtained from all the subjects. The experimental protocol was designed as follows. Each trial consisted of 12 s with 2 s of preparation, 4 s of action and 6 s of rest. Each subject underwent three sessions, i.e., calibration, feedback and non-feedback. A "smiley" or "poker" face image was shown based on the subject's performance for the feedback session, whereas no feedback was provided for non-feedback session. The models were trained based on the calibration data collected from previous sessions. Each session consisted 40 trials of actions and 40 trials of idle. In the experiments, the subjects were instructed to perform kinesthetic motor imagery of hand grasping or mental counting followed a visual cue shown in the computer screen, for motor imagery and idle, respectively. Body movements and eye blinks were constrained. EEG signals were recorded by NuAmps EEG acquisition hardware (http://www.neuroscan.com) with unipolar Ag/AgCl electrodes. Placements of the electrodes followed the international 10-20 system standard. The EEG signal was digitally sampled at 250 Hz with a resolution of 22 bits, and the voltages fell in the range of ±130 mV. The EEG signals

TABLE I

CHANNEL SELECTION BASED ON INCONSISTENCIES FROM MULTIPLE CLASSIFIERS (CS-IMC)

Inputs: the re-filtered training and testing EEG data for current fold ($n \times k$ fold CV).

Outputs: the eliminated channel ($\hat{c}_i$) and CV accuracy ($C_a(\hat{c}_i)$).

1) Computed covariance features of the refiltered training and testing sets for the currently remaining channels (c_r), denoted the obtained features as $F^c_{tr}(c_r)$ and $F^c_{te}(c_r)$.
2) Trained the models based on $F^c_{tr}(c_r)$ for multiple classifiers $\mathcal{H}_i$, where $i \in \{s, b, l, d\}$, e.g., $\mathcal{M}_s = \mathcal{H}_s(F^c_{tr}(c_r))$, which were then used to classify $F^c_{te}(c_r)$ to obtain the classification accuracies for the training ($P_{tr}(c_r, i)$) and testing ($P_{te}(c_r, i)$) data for the ith classifier.
3) Computed the inconsistencies of the training ($\mathcal{I}_{tr}(c_i)$) and testing ($\mathcal{I}_{te}(c_i)$) classification accuracies between any two classifiers by removing channel c_i, which were given by

$$\mathcal{I}_{tr}(c_i) = \frac{1}{n_p} \sum_{i,j=1, i \neq j}^{n_p} |P_{tr}(c_r, i) - P_{tr}(c_r, j)| \quad (3)$$

$$\mathcal{I}_{te}(c_i) = \frac{1}{n_p} \sum_{i,j=1, i \neq j}^{n_p} |P_{te}(c_r, i) - P_{te}(c_r, j)| \quad (4)$$

where the remaining channel by removing c_i was: $c_r = c_r \setminus c_i$; the total pair of classifiers for a total of n_r classifiers was: $n_p = \binom{n_r}{2}$; $|x|$ denoted the absolute values of x.
4) Obtained the overall inconsistencies ($I_c(c_i)$) by fusing the inconsistencies for training ($\mathcal{I}_{tr}(c_i)$) and testing ($\mathcal{I}_{te}(c_i)$) data, which was given by

$$\mathcal{I}_c(c_i) = w_1 * \mathcal{I}_{tr}(c_i) + w_2 * \mathcal{I}_{te}(c_i) \quad (5)$$

where the pair of weights (w_1, w_2)=(0.4, 0.6) was used.
5) Eliminated the channel where the overall inconsistency of the outputs from multiple classifiers was the maximum (see Eq. (5)).

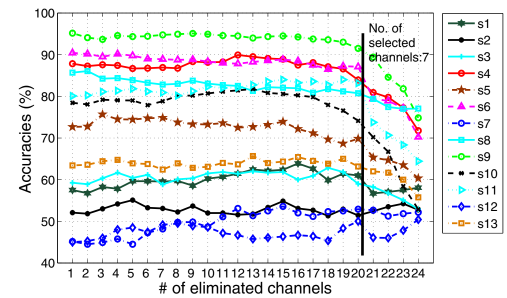
$$\hat{c_i} = \arg\min_{c_i \in \{c_r\}} (1 - \mathcal{I}_c(c_i)) \quad (6)$$

6) Recomputed the CSP features of training and testing data for currently remaining channels ($\hat{c}_r$), obtained the CV accuracy $C_a(\hat{c}_i)$.

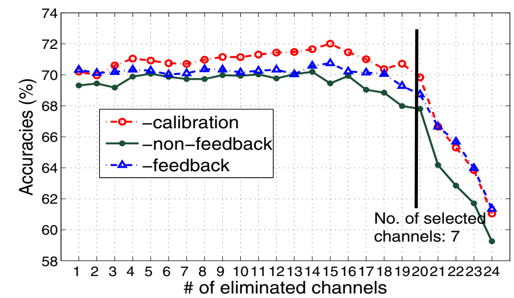
were recorded for 27 channels, which were bandpass filtered from 0.05 Hz to 40 Hz by the acquisition hardware.

B. Performance of the proposed method

Experiments were conducted to test the CV classification accuracies by eliminating each channel at a time with the 10×10 fold CV results being reported in Fig. 2. The results demonstrated that the averaged classification accuracies increased with the elimination of the noisy channels, e.g., from 1 to 15 channels, which was obvious for calibration session. The accuracies decreased significantly when the number of eliminated channels reached certain values, e.g., 20, reaching the stopping criterion. Furthermore, the performance of feedback session was superior to that of non-feedback session as expected. To further demonstrate the effectiveness of the proposed channel selection method, the distributions of the selected channels, e.g., selecting 16 channels in 10×10 fold CV, for non-feedback session were visualized in Fig. 3. The visualization results revealed that the most frequently selected channels were located at the motor area such as 'C4', and somatosensory association cortex such as 'CP4'. The high frequency that the channels such as 'PO1' and 'PO2' at visual cortex area was selected may be due to the involvements of too much visual imaging during motor imagery.



(a) all subjects in calibration session



(b) averaged across subjects for all three sessions

Fig. 2. CV classification accuracies by eliminating each channel at a time for (a) all subjects at the calibration session, and (b) group averaged across subjects for all three sessions. The vertical black line indicated the cut-off number of channels selected: 7.

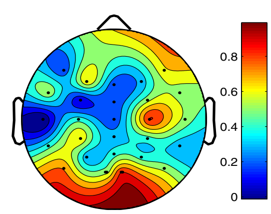
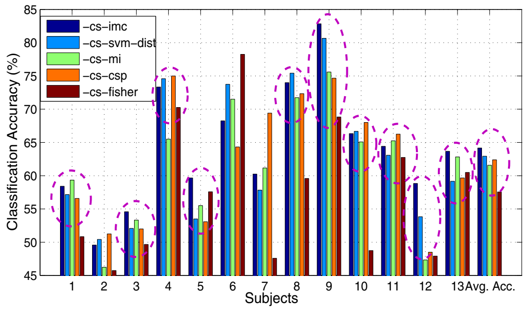


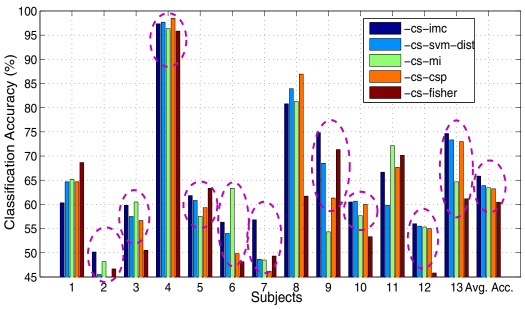
Fig. 3. Visualization of the selected channels for the non-feedback session. The channels that were colored in 'red' were the most frequently selected area.

C. Comparisons

Comparisons were made between our proposed CS-IMC method and that of other channel selection methods such as CSP (*cs-csp*) [8], mutual information (*cs-mi*) [6], fisher criterion (*cs-fisher*) and SVM-distance (*cs-svm*)-based methods [4]. The averaged accuracy increases of proposed CS-IMC compared with other methods were summarized in Table II. The comparison of CV accuracies of different methods for the non-feedback and feedback sessions were shown in Fig. 4. The results revealed that the averaged accuracies of our proposed CS-IMC by selecting 7 channels were 1.23%, 2.60% (significant with p-value: 0.045), 1.78% and 6.61% (significant with p-value: 0.007); and 1.97%, 2.40%, 2.62% and 5.39% (significant, with p-vaue: 0.021) higher than that



(a) non-feedback session



(b) feedback session

Fig. 4. Comparison of CV accuracies for different methods by selecting 7 channels. Those subjects circled in "purple" represented superior performance of our method. The last group of bar was averaged accuracies across subjects.

of *cs-svm*, *cs-mi*, *cs-csp* and *cs-fisher* methods for non-feedback and feedback sessions, respectively. The averaged accuracy of CS-IMC for the calibration session was 2.77%, 0.50% and 4.01% higher than that of *cs-mi*, *cs-csp* and *cs-fisher*, respectively. However, it was 0.21% lower than that of *cs-svm*. These results led to the means of averaged accuracy increases across methods of 4.07%, 3.10% and 1.77%, in comparison with the four existing methods for non-feedback, feedback and calibration sessions, respectively. In particular, our proposed method performed extremely good for those low-accuracy performers, e.g., those subjects whose CV accuracies were below 70%, as can be seen from Fig. 4. It is noted that the performance of the low-accuracy performers can be very low, hence, proper motor imagery strategies should be enforced to improve the performance, which may be coupled with the motor observation.

TABLE II
COMPARISON OF CV ACCURACIES WITH EXISTING METHODS

CS-IMC vs. methods	Avg. Acc. increase (%) for sessions non-feedback	feedback	calibration
cs-svm	1.23	1.97	-0.21
cs-mi	2.60*	2.40	2.77
cs-csp	1.78	2.62	0.50
cs-fisher	6.61*	5.39*	4.01
MAR.	4.07	3.10	1.77

Note: * significant at 5% significance level. MAR.: means of the averaged accuracy increases across different methods.

V. CONCLUSIONS

In this paper, we presented a novel channel selection method to select the most effective EEG channels based on the inconsistencies of the outputs from multiple classifiers, measured for both training and testing data. Our idea was motivated by the fluctuations from the outputs of the multiple classifiers by the inclusions of those noisy channels. These noisy channels were eliminated one by one till the stopping criterion was satisfied. Experiments were conducted on an EEG data set consisted of 13 healthy subjects performing motor imagery of hand grasping for three sessions. The results showed that the mean increases of 4.07%, 3.10% and 1.77% of the averaged accuracies were achieved in comparison with that of the SVM, mutual information, CSP, and fisher criterion-based channel selection methods for the non-feedback, feedback and calibration sessions, respectively. The performance decreased significantly when the number of channels dropped to certain value, e.g., less than 10 channels. Furthermore, the selected channels were located mainly on the motor cortex and somatosensory association cortex areas.

REFERENCES

[1] W. K. Tam, K. Y. Tong, F. Meng and S. Gao, "A minimal set of electrodes for motor imagery BCI to control an assistive device in chronic stroke subjects: a multi-session study," *IEEE Trans Neural Syst Rehabil Eng.,* vol. 19, no. 6, pp. 617-27, Dec. 2011.

[2] A. K. Jain, P. W., Duin Robert and J. Mao, "Statistical pattern recognition: a review," *IEEE Trans. On Patt. Anal. and Mach. Intel.,* vol. 22, no. 1, pp. 4-37, 2000.

[3] M. Arvaneh, C. Guan, K.K. Ang, C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans Biomed Eng.,* vol. 58, no. 6, pp. 1865-73, June 2011.

[4] T. N. Lal, M. Schrder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer and B. Schlkopf, "Support vector channel selection in BCI. *IEEE Trans Biomed Eng.,* vol. 51, no. 6, pp. 1003-10, June 2004.

[5] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection, " *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.

[6] T. Lan, D. Erdogmus, A. Adami, M. Pavel, S. Mathan, "Salient EEG channel selection in brain computer interfaces by mutual information maximization,"*Conf Proc IEEE Eng Med Biol Soc.*, vol. 7, pp. 7064-7, 2005.

[7] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis," *IEEE Signal Proc. Magazine*, vol. 25, no. 1, pp. 41-56, 2008.

[8] Y. Wang, S. Gao and X. Gao, "Common Spatial Pattern Method for Channel Selection in Motor Imagery Based Brain-computer Interface," *Conf Proc IEEE Eng Med Biol Soc.,* pp. 5392-5395, 2005.

[9] X. Yong, R. K. Ward, and G. E. Birch, "Sparse spatial filter optimization for EEG channel reduction in brain-computer interface," I*IEEE ICASSP'08*, pp. 417-420, 2008.

[10] J. Farquhar, N. J. Hill, T. N. Lal, and B. Schlkopf, "Regularised CSP for Sensor Selection in BCI," *Proc. of the 3rd Int. Brain-Computer Interface Workshop and Training Course*, pp. 14-15, 2006.

[11] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. on Intel. Systems and Tech.,* vol. 2, pp. 27:1-27:27, 2011. Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.

[12] K.K. Ang, Z.Y. Chin, C. Wang, C. Guan and H. Zhang, 2012, "Filter Bank Common Spatial Pattern algorithm on BCI Competition IV Datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, pp.1-9, 2012.

[13] H. Ramoser, J. Muller-Gerking and G. Pfurtscheller, "Optimal Spatial Filtering of Single Trial EEG during Imagined Hand Movements," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 4, pp. 441-6, 2000.