32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

# Trial Pruning for Classification of Single-Trial EEG Data during Motor Imagery

Boyu Wang, Chiman Wong, Feng Wan, Peng Un Mak, Pui In Mak, and Mang I Vai

*Abstract*—**Due to the artifacts in electroencephalography (EEG) data, the performance of brain-computer interface (BCI) is degraded. On the other hand, in the motor imagery based BCI system, EEG signals are usually contaminated by the misleading trials caused by improper imagination of a movement. In this paper, we present a novel algorithm to detect the abnormal EEG data using genetic algorithm (GA). After trial pruning, a subset of the EEG data are selected, on which common spatial pattern (CSP) and Gaussian classifier are trained. The performance of the proposed method is tested on Data set IIa of BCI Competition IV, and the simulation result demonstrates a significant improvement for six out of nine subjects.**

## I. Introduction

THE brain-computer interface (BCI) is a system that forms a possible control and comunication channel for the individuals with severe motor disabilities to have effective control over external devices without using the traditional pathways such as peripheral muscle or nerves [1–3]. At present, the brain activities are often recorded noninvasively by the electroencephalogram (EEG), which has excellent temporal resolution and usability, and the EEG signal is therefore a popular choice for BCI research.

In order to control an EEG-based BCI, the user must produce different brain activity patterns, which are recorded by electrodes on the scalp, and then features are extracted from the EEG signals and translated into the control commands by classification algorithms. For a review on the feature extraction and classification technologies used in BCIs see [3–6].

Noise is ubiquitous in EEG signals due to the factors such as measurement inaccuracies, physiological variations in background EEG, muscle and eyes blink artifacts. On the other hand, for a motor imagery based BCI, the noise can also be induced by improper imagination of a movement. The artifacts can be removed by independent component analysis (ICA) [7], [8] or rejected by thresholds or criteria. However, the ICA algorithm requires visual inspection to choose the artificial components, which makes it unfeasible to be applied in an automated BCI system. Furthermore, both of the methods cannot detect the noise caused by improper imagination trials since they do not take the label information into account [9].

The performances of the classification and data-driven feature extraction methods (e.g., common spatial pattern (CSP) [10]) could be degraded due to the trials contaminated by noise. A trial pruning method based on relevant dimensionality estimation (RDE) was proposed in [9] which only handled the noise caused by the imagination improperly carried out or even related to wrong class.

In this paper, a novel trial pruning method based on genetic algorithm (GA) is presented to simultaneously remove the trials contaminated by physiological noise (e.g., artifacts) and subjects' failure to carry out mental task. The CSP and Gaussian classifier are performed on a "pruned" subset so that more robust and reliable estimation can be achieved. The performance of the proposed method is tested on Data set IIa of BCI Competition IV, and the simulation result demonstrates a significant improvement, especially for the trials contaminated by noise.

## II. Methods

Our method is proposed based on the combination of CSP-Gaussian classifier couple and GA. In particular, the CSP is trained on a subset of the EEG data and projects the EEG signals to the feature space. Then the mean and covariance of each class are estimated and the likelihood functions of the feature vectors are calculated. In the GA step, the chromosomes are updated according to the likelihood contribution for the current estimated model. Each pair of the chromosome and the corresponding mixture model will be a possible solution. The two steps are performed alternately until the stop criteria are met or the maximum number of generation is achieved. The flowchart of the genetic-based trial pruning method is shown in Fig. 1.

### A. Common Spatial Pattern (CSP)

CSP is a discriminative algorithm designed based on a decomposition of the raw EEG signals into spatial patterns so that the variances of the resulting signals carry the most discriminative information.

Let $\sum^{+} \in \mathbb{R}^{C\times C}$ and $\sum^{-} \in \mathbb{R}^{C\times C}$ (where $C$ is the number of the channels) be the estimated covariance matrices of two classes EEG signals (e.g., motor imagery of two different hands in this paper), then the CSP components can be calculated by a simultaneous diagonalization of the two covariance matrices so that the sum of the two diagonalized matrices is the unit matrix:

$$\begin{aligned} W^T \sum^{+} W &= \Lambda^{+} \\ W^T \sum^{-} W &= \Lambda^{-} \\ \Lambda^{+} + \Lambda^{-} &= I \end{aligned} \tag{1}$$

Since $\lambda_j^{+} + \lambda_j^{-} = 1 (j = 1,\ldots,C)$, if a spatial filter $w_j$ yields a

The authors are with the Department of Electrical and Electronics Engineering, Faculty of Science and Technology, University of Macau, Av. Padre Tomás Pereira,Taipa, Macau.

large value of $\lambda_j^+$ in one class, it will project the other class with low eigenvalue $\lambda_j^-$. As the eigenvalues represent the variance of each class, the resulting EEG signals filtered by $w_j$ will maximally separable by their variance. For classification the features of single-trials are chosen as the log-variance of the CSP filtered signals. For more detailed and comprehensive accounts on CSP see [10].

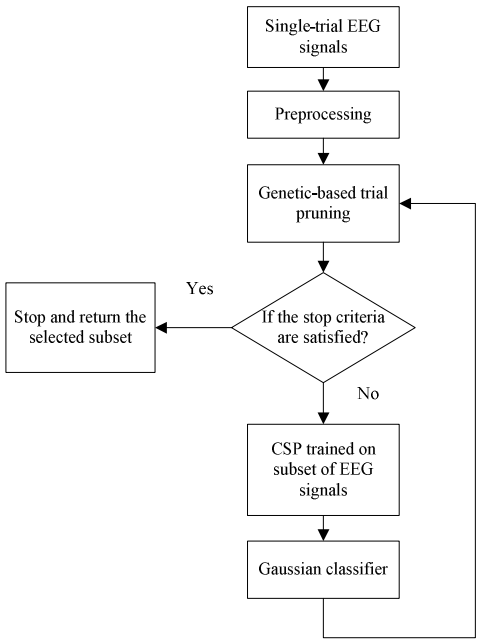


Fig. 1. Flowchart of genetic-based trial pruning method.

### B. Gaussian Classifier

After feature extraction, we use the Gaussian classifier, which is based on Bayes' decision theory, to assign data **x** to class *i*, if

$$\begin{aligned} i &= \arg\max_i p(\mathbf{x}\mid\omega_i) \\ &= \arg\max_i \left( \frac{1}{(2\pi)^{d/2}\left|\sum_i\right|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\mathbf{u}_i)^T \sum_i^{-1}(\mathbf{x}-\mathbf{u}_i) \right\} \right) \\ &= \arg\max_i \left( \frac{1}{(2\pi)^{d/2}\left|\sum_i\right|^{1/2}} \exp\{-D_i\} \right) \end{aligned} \quad (2)$$

where $\mathbf{u}_i$ and $\sum_i$ are the mean and covariance of class *i* respectively, *d* is the dimension of the feature vectors, and $D_i$ is the Mahalanobis distance between the data **x** and $\mathbf{u}_i$. Here we have assumed that each class has the same prior probability $p(\omega_i)$. The $\mathbf{u}_i$ and $\sum_i$ are estimated by maximum likelihood estimation:

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{N_i}\sum_{n=1}^{N_i}\mathbf{x}_n \quad \text{and} \quad \hat{\sum}_i = \frac{1}{N_i}\sum_{n=1}^{N_i}(\mathbf{x}_n-\hat{\boldsymbol{\mu}}_i)(\mathbf{x}_n-\hat{\boldsymbol{\mu}}_i)^T \quad (3)$$

where $N_i$ is the number of training samples of class *i*.

The reason we choose Gaussian classifier is twofold. First the variance features are approximately chi-square distributed, and therefore the log-variance features are similar to normal distribution [11], [12]. On the other hand, the Gaussian classifier can naturally produce probabilistic outputs, which are utilized as a part of objective function in the GA step.

### C. Genetic-Based Trial Pruning

As discussed above, the EEG data can be contaminated by physiological variations, artifacts, or improper imagination of the movements. Hence, training on a "cleaned" subset which contains the reliable class information can achieve a more robust estimation.

In general, the trials to be pruned are divided into two categories. The first one is the outliers, which are far from the whole data set in the feature space. This type of trials can be regarded as task-irrelated noisy trials, since the main reason of this type of trials is the background EEG noise or artifacts, and usually not related to the motor imagery task. Even though sometimes they are classified correctly, they may pull the class mean and covariance, or the decision boundary towards their location, which can affect the generalization performance. The second category is the trials which are misclassified, of which the presence is mainly caused by the subject's failure to carry out mental task, and these trials can be viewed as task-related noisy trials.

To achieve a more robust analysis result, the trials contaminated by both types of the noise should be handled carefully. From the statistical point of view, the outliers are the samples with (unnormalized) low class conditional probability density $p(\mathbf{x}\mid\omega_i)$ for both classes, and the misclassified samples have higher $p(\mathbf{x}\mid\omega_i)$ for the opposite class than the class they are assigned to. Therefore the information of labels and class conditional probabilities can be taken into account to detect and prune noisy trials.

In this paper, the subset is selected by genetic algorithm (GA). The purpose of integrating GA into CSP-Gaussian classifier couple is to take advantage of the search capability of GA. *N*-bit binary vectors are first randomly initialized, where the *N* is the number of training samples. Each bit indicated whether the sample **x** is pruned or not. Only the samples survived (the corresponding bit is 1) can contribute to the training of CSP-Gaussian classifier couples. The chromosomes are updated according to their fitness score.

To utilize the information of label and class conditional probability density of each sample, guided mutation [13] is adopted instead of basic mutation during the update procedure. In particular, the bits corresponding to the samples with low $p(\mathbf{x}\mid\omega_i)$ for both classes or the misclassified samples with high class conditional probability density difference values are set to be zero, while the samples with high $p(\mathbf{x}\mid\omega_i)$ and correctly classified with high confidence are set to be one. The trials with low conditional probability density can be regarded as the outliers caused by physiological noise such as artifacts (type I noisy trials),

while the misclassified trials can be viewed as improper imagination trials (type II noisy trials). Fig.2 demonstrates the flowchart of the guided mutation.

The fitness for the samples is defined as the sum of the difference of the posterior probabilities for each class, which is equal to difference of the normalized class conditional probability density in our case. For the samples which are labeled to the first class, the corresponding individual fitness should be $\left(p(\mathbf{x}\mid\omega_1)-p(\mathbf{x}\mid\omega_2)\right)/\left(p(\mathbf{x}\mid\omega_1)+p(\mathbf{x}\mid\omega_2)\right)$, and $\left(p(\mathbf{x}\mid\omega_2)-p(\mathbf{x}\mid\omega_1)\right)/\left(p(\mathbf{x}\mid\omega_1)+p(\mathbf{x}\mid\omega_2)\right)$ for the samples labeled to second class. Thus, the fitness of the misclassified sample is negative and the best individual is the one which maximize the fitness function. The reason we choose the difference of the posterior instead of classification accuracy is that this fitness function takes into account not only the label information but also the probabilistic information indicating how well the samples are distributed in the feature space. It should be noted that in each iteration step only the selected samples contribute the fitness.

## III. EXPERIMENTS

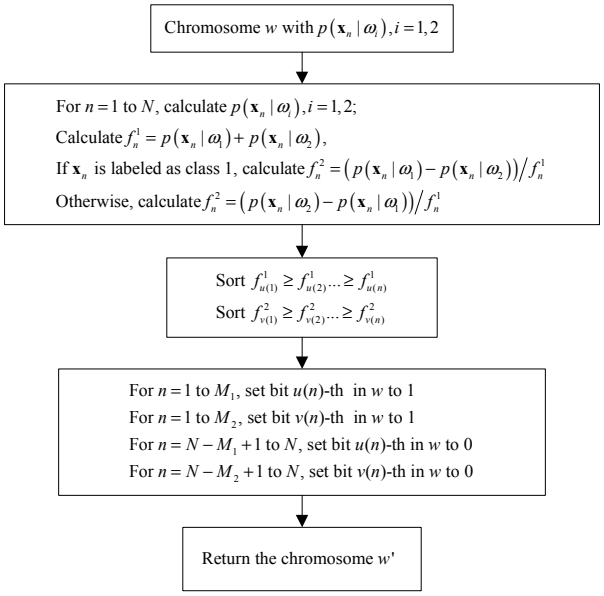


Fig. 2. Flowchart of guided mutation

### A. Data Set

The performance of the proposed method is evaluated on the data set IIa from BCI competition IV [14], provided by Graz University of Technology. The data set consists of EEG data from 9 subjects who performed the imagination of left hand, right hand, both feet, and tongue. Two sessions on different days were recorded for each subject. Each session is comprised of 6 runs separated by short breaks. One run consists of 48 trials with duration of 7s, yielding a total of 288 trials per session. For detailed description of this data set, see [15]. In this study, only two classes, namely the EEG signals of left hand and right hand motor imagery are used. In addition, instead of providing continuous classification output for each sample, we only consider the discrete classification for each trial.

### B. Parameter Setting

Some parameters for the proposed algorithms should be set carefully. Specifically, we choose the time segment from 2.5s to 4.5s of each trial as the training data, which was also adopted by the winner of the BCI competition. Before feature extraction, the EEG signals are filtered by 8-30Hz band pass filter.

The features of the EEG signals are extracted by CSP. Since the total number of parameters of the Gaussian model for each class is $D\times(D+1)/2+D$, where $D$ is the dimension of the feature space, to avoid the overfitting we only choose the first pair of the spatial filters to extract the features ($D$=2).

The parameters for the proposed genetic-based method are selected by experience. In particular, the number of the individuals is 20, and the generation number is 30. Roulette wheel rank weighting and single-point crossover are used for crossover operation. For guided mutation, the pruning number $M_1$ and $M_2$ are set to be 5% of the number of the total samples and 10% of the number of misclassified samples of each iteration, since if the pruning level is high, it may lead to overfitting and break the convergence of GA.

### C. Results

The performances of the proposed method based on trial pruning and the classical CSP-Gaussian classifier couple trained on whole data set are demonstrated in Table 1. We also compare the different pruning strategy, i.e. pruning the outliers (type I noisy trials) only ($M_2$ = 0) and pruning improper imagination trials (type II noisy trials) only ($M_1$= 0) respectively. The best accuracy for each subject is shown in bold type.

It can be observed that for the proposed method, the performances are improved for most of the subjects (six out of nine), and the classification accuracies of only two subjects slightly decrease. Besides, for the subjects who perform the imagination task well, the improvement is not significant, but for the subjects with relative low classification accuracy (<90%), the improvement is remarkable (five out of six subjects benefit from the proposed method). Therefore, the proposed method can provide more robust and reliable estimation for the trials contaminated by noise.

The combination of the type I and type II noisy trial pruning provide the best performance for the nine subjects, while each of the single pruning method cannot improve the overall classification accuracies significantly, especially for the method only pruning the type II noisy trials. One possible reason is that the outliers can change the distribution of the data which can be eliminated by the proposed method, while the type II noisy trials may be caused by the intrinsic factors such as subject's failure to perform the imagination or inseparability of the extracted feature. Thus, the pruning strategy is an efficient method of automatic detection of the outliers. For the type II noisy trials, tuning the preprocessing

TABLE I
CLASSIFICATION ACCURACY (%) OF EACH METHOD

| Methods | Subjects | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 |
| Classical CSP-Gaussian classifier couple | 85.42 | 54.17 | **92.36** | 69.44 | 52.78 | 62.50 | **77.08** | 92.36 | **90.28** |
| Method pruning improper imagination trials only ($M_1 = 0$) | 85.42 | 56.25 | 90.28 | 64.58 | 54.86 | 61.11 | 72.22 | **93.75** | 91.67 |
| Method pruning outliers only ($M_2 = 0$) | 86.11 | 60.42 | 91.67 | 71.53 | 56.25 | 63.19 | 76.39 | 91.67 | 89.58 |
| Proposed trial pruning based methods | **88.19** | **63.19** | 91.67 | **73.61** | **60.42** | **65.28** | 75.69 | **93.75** | **90.28** |

parameters such as training time segment, frequency band of the temporal filter may improve the performance, and the proposed method can also alleviate the negative effect of the misclassified samples.

To further validate the robustness of the proposed method, we add ten rejected trials containing artifacts in each subject's training set and assign their labels randomly. Table II shows the performances of the methods for different data set, where data set I is the normal set, and data set II is the training set containing rejected trials. It can be observed that the performance of the estimation on the pruned subset only slightly degrades, much better than the classical one. Since the proposed method can automatically detect the noisy samples, especially the trials containing artifacts, it is more robust and reliable than the classical method.

TABLE II
CLASSIFICATION ACCURACY (%) FOR DIFFERENT DATA SETS

| Methods | Data set I | Data set II |
|---|---|---|
| Classical CSP-Gaussian classifier couple | 75.15 | 68.90 |
| Proposed trial pruning based methods | 78.01 | 75.62 |

## IV. CONCLUSIONS

In this paper, a genetic-based trial pruning method is proposed to improve the performance of EEG classification. The motivation of this work is to automatically prune the trials contaminated by noisy data caused by artifacts and improper motor imagery without any additional channel rejection operation (i.e., ICA) or visual inspection of an expert. After the trial pruning, the CSP and the classifier are trained on a subset and achieves a robust performance. Moreover, through the guided mutation, both class conditional probabilities and label information are taken into account, which can make full use of the information of the data distribution.

The future work will focus on validating the proposed method on more data sets. In addition, more effective algorithm to handle the type II noisy trials is under development. One promising approach is to reweight the trials [9], which could be integrated into our method. Besides, we are also exploring automated parameter selection techniques for the genetic-based method, as well as other fitness function which could be more powerful to deal with this task.

## ACKNOWLEDGMENT

The authors gratefully acknowledge support from the Macau Science and Technology Department Fund (Grant FDCT/036/2009/A) and the University of Macau Research Fund (Grants RG059/08-09S/FW/FST and RG080/09-10S/WF/FST).

## REFERENCES

[1] J. R. Wolpaw *et al.*, "Brain–computer interface technology: a review of the first international meeting," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 164–173, June 2000.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interface for communication and control," *Clin. Neurophsiol.,* vol. 133, pp. 767-791, June 2002.

[3] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals," *J. Neural Eng.*, vol. 4, pp. R32–R57, 2007.

[4] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, pp. R1–R13, 2007.

[5] K.-R.Müller, C.W. Anderson, and G. E. Birch, "Linear and nonlinear methods for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, pp. 165–169, June 2003.

[6] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol.11, pp. 141-144, June 2003.

[7] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[8] A. Kachenoura, L. Albera, L. Senhadji, and P. Comon, "ICA: A potential tool for BCI systems," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 57–68, Jan 2008.

[9] C. Sannelli, M. Braun, and K.-R.Müller, "Improving BCI performance by task-related trial pruning," *Neural Netw.*, vol. 22, pp. 1295-1304, Nov. 2009.

[10] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Mueller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[11] B. Blankertz *et al.*, "Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 11, pp. 127–131, June 2003.

[12] B. Blankertz *et al.*, "Invariant common spatial patterns: Alleviating nonstationarities in braincomputer interfacing," in *Advances in Neural Information Processing Systems 20*, Cambridge, MA: MIT Press, 2008.

[13] D.T. Nguyen, L. Chen, and C. K. Chan, "An outlier-aware data clustering algorithm in mixture model," in *Proc. 7th Int. Conf. Information, Communication and Signal Processing*, Macau, China, 2009, pp. 1-5.

[14] BCI competition IV website. Available: http://bbci.de/competition/iv/

[15] C. Brunner, R. Leeb, G.R. Müller-Putz, A. Schlögl, and G.Pfurtscheller, "BCI Competition 2008 — Graz data set A," Available: http://bbci.de/competition/iv/desc_2a.pdf