# Automated dimensionality reduction in EEG based Brain Computer Interface

Asiri Nanayakkara† and Zahmeeth Sakkaff ‡
Artificial Intelligence and Applied Electronics Research Unit, Institute of Fundamental Studies, Hantana Road, Kandy 20000, Sri Lanka
Email : † asiri@ifs.ac.lk , ‡ zahmeeth@gmail.com

***Abstract* - A simple method is developed for selecting effective channels and feature dimensions automatically at the training stage of BCI systems. The method is applied on feature vectors constructed with all the EEG channels used for recording. Performance was evaluated with EEG data which was preprocessed by band pass filtering and feature vectors constructed by band powers. The classification method used in the evaluation is k Nearest Neighbor classifier, which is sensitive to number of dimensions in the feature vectors. It was found that new method can effectively select features and reduce channels and thereby improve accuracy and efficiency of BCI systems.**

***Keywords* - Brain Computer Interface (BCI), Dimensionality reduction, Electroencephalography (EEG)**

## I. INTRODUCTION

For most of the severely paralyzed patients, Brain Computer Interface (BCI) systems can provide an effective way of communicating with the outside world [1]-[4]. Since BCI usually does not rely upon any muscle movements and uses only the changes in electric or magnetic fields in the brain which are generated by certain thoughts or mental activities, it is ideally suited for the users who are unable to perform any motor functions. BCI systems based on scalp-recorded EEG signals have become very popular in the BCI community due to its fine temporal resolution and inexpensive recording equipment [5]-[10].

In most EEG based BCI systems, classifiers in the system have to be trained through machine learning techniques with a training set of data before using them for testing or controlling BCI. These training EEG data sets are constructed with several training trials recorded from the same subject. After recording EEG signals from the subject, the preprocessing techniques are used to condition the EEG data by removing undesirable frequencies and certain artifacts. Feature extraction methods are then employed to construct feature vectors from conditioned EEG data such that feature vectors can be directly utilized to train the classifiers of the system. Unless removed, all the recorded channels are usually contributing for classification, regardless of their relevance to the mental tasks. Normally, feature vectors are often of high dimensionality in BCI systems [11]. As matter of fact, several features are usually extracted from a number of EEG channels and from several time segments before being concatenated into one large single feature vector. As a result, classification methods which are sensitive to higher dimensionality such as k-Nearest Neighbor classifiers are not popular in BCI [11]. Since dimensions of the feature vectors affect the efficacy of the BCI system, it is very useful to find ways to reduce dimensionality of feature vectors without loosing important features.

In order to address this issue of higher dimensionality, various special filtering techniques such as Laplacian derivations, Common Spatial Patterns (CSPs) [12]-[13], and various independent component analysis (ICA) algorithms such as Infomax, FastICA and SOBI [14]–[16] have been used by the BCI community. Recently, for dimensional reduction in Motor Imagery based BCI, spatial filters based on independent components analysis (ICA) method has been tried out with data preprocessed by principal component analysis (PCA) [17]. In that study, it was found that the PCA preprocessing was not a suitable method that could retain motor imagery information in a smaller set of components. On the other hand, 6 ICA components chosen by visual inspection performed similar to the full range of 22 components. Further, an automated selection of 8 ICA components based on a variance criterion was found to perform better than visually selected components.

In this paper a method is developed to select effective channels and improve the quality of feature vectors by reducing dimensionality. Since the method utilizes Fisher criterion function, it is called *Automatic channel selection through Fisher Criteria* (ACSFC). In this paper, the description of ACSFC is given for binary classification but the method can easily be extended for multiclass problems.

## II. METHOD

In ACSFC, the feature vectors are constructed by concatenation of sub-feature vectors after reducing their dimensions. (Note that sub-feature vector is a feature vector whose elements belong to a single channel) Dimensional reduction in ACSFC for binary classification is carried out by keeping only the elements in the sub-feature vectors which satisfy following conditions.

(a) Difference between statistical means of the feature elements of the same dimension corresponding to class 1 and class 2 is large.

(b) Standard deviation of the feature values of the same dimension corresponding to each class is small.

The above two conditions are equivalent to having large values for Fisher criterion function $J$ [18] and can be described mathematically as follows. Assume that $y_{i,j,k}$ are the sub-feature vectors corresponding to channel $k$ constructed with features $x_{i,m,j,k}$ ($m = 1,2,...M$ and $i = 1$ *or* $2$ ) extracted from training EEG data set.

$$y_{i,j,k} = \begin{bmatrix} x_{i,1,j,k} \\ x_{i,2,j,k} \\ . \\ . \\ x_{i,M-1,j,k} \\ x_{i,M,j,k} \end{bmatrix} \tag{1}$$

where $i$ is the label of the mental task, $j$ represents the trail index and $k$ is the channel index. $M$ is the number of features extracted from each channel.

Each element $x_{i,m,j,k}$ in the sub-feature vector can be considered as $j^{\text{th}}$ value of a random variable $x_{i,m,k}$ with unknown distribution. In ACSFC method, the dimensionality of the sub-feature vectors is reduced by removing $x_{i,m,k}$ 's which make Fisher criterion function $J_{m,k}$ *small* ;

$$J_{m,k} = \left[ \frac{\left|\mu_{1,m,k} - \mu_{2,m,k}\right|^2}{s^2_{1,m,k} + s^2_{2,m,k}} \right] \tag{2}$$

where $\mu_{i,m,k} = \frac{1}{N}\sum_{j=1}^{N}\left(x_{i,m,j,k}\right)$ and $s^2_{i,m,k} = \frac{1}{N}\sum_{j=1}^{N}\left(x_{i,m,j,k} - \mu_{i,m,k}\right)^2$ , and $N$ is number of trials.

In other words, reduced sub-feature vectors will only contain $x_{i,m,k}$ 's which have high $J_{m,k}$ values. Final feature vectors are then constructed by concatenation of sub-feature vectors having large $J_{m,k}$ values. If , for a given channel $k$, $J_{m,k}$ are small for all $m$ then the channel $k$ is considered as irrelevant for the mental tasks and will not be used in controlling BCI.

In order to evaluate performance of ACSFC method, band pass filtering is employed as the preprocessing method while band powers are used for the feature vector construction. The method can be generalized for other preprocessing and feature vector construction methods which do not mix channels during the process. In ACSFC method, only the effective channels will contribute to the final form of the feature vector. For simplicity, we restrict the evaluation here to binary classification. Nevertheless, the method can be easily extended to multiclass problems as well. The effectiveness of ACSFC is demonstrated by using EEG data recorded in our laboratory for the Visualization of Arrow Movements (VAM). Details of these mental tasks can be found in [19].

The EEG data was recorded with 20 electrodes using 10-20 electrode placement system. The signals were sampled with 256 Hz. Additionally a 50Hz notch filter was used to reduce noise due to power lines. The training data sets were recorded from three healthy subjects, all inexperienced in BCI training. 90 trials per subject were recorded for each mental task, totaling 180 for Baseline (BL), Imaginary Right Arrow Movement (IMRAM) and Imaginary Down Arrow Movement (IMDAM).

Two training EEG data sets corresponding three mental tasks BL, IMRAM and IMDAM were recorded as follows. Subjects were seated in an armchair looking at a black square pasted on a white blank screen which was placed approximately one and half meters away from the subject. They were asked to keep their arms and hands relaxed and to avoid eye movements during the recordings. Each trial started with a program informing the subject verbally what mental task should be imagined in the upcoming trial. Mental tasks are randomly chosen while keeping the total number of trials of each mental task in a recording session to a fixed value of 20. Seven seconds later, a short beep is made to inform the subject to start the mental task while keeping his/her eyes open. If it is observed that eye blinks or eye moments occurred during the recording of the trial, data recorded in the trial is discarded and the trial is repeated. Duration of a trial is nine seconds and at the end of nine seconds another beep is produced by the program informing the subject to stop the mental task. We have carried out six recording sessions each consists of 100 trials for all three subjects.

As mentioned earlier, in ACSFC method, determination of selection of channels and reduction of dimensions of feature vectors are carried out based on values of criterion function $J$ [18]. In the preprocessing stage EEG data corresponding to each channel of every trial and mental task is filtered separately with a band pass filter having the same lower frequency $f_1$, and upper frequency $f_2$. After preprocessing, sub-feature vectors are constructed for each channel by using band powers with the band width $\Delta f$ . The performance of ACSFC was evaluated with $f_1=1, f_2=30,$ and $\Delta f = 32$ . These sub-feature vectors are denoted by $y_{i,j,k}$ where $i$ is the label of the mental task. $j$ represents the trail index and $k$ is the channel index. For given $i$ ,$j$, and $k$, $y_{i,j,k}$ is a column vector having band power values $b_{i,m,j,k}$ ( $m = 1,2,...M$ ) as its elements.

$$y_{i,j,k} = \begin{bmatrix} b_{i,1,j,k} \\ b_{i,2,j,k} \\ . \\ . \\ b_{i,M-1,j,k} \\ b_{i,M,j,k} \end{bmatrix} \quad (3)$$

where $M$ is the number of elements in each sub-feature vector before applying ACSFC method. At this stage we apply ACSFC method to reduce the dimensionality of individual sub-feature vectors as described below.

(a) First $J_{m,k}$ corresponding to each feature $m$ of channel $k$ is calculated.
(b) If $J_{m,k} > 0.5 \times \max_{\forall m,k} \{J_{m,k}\}$, corresponding feature dimension is selected. Otherwise, feature dimension is eliminated.
(c) The final feature vector is then constructed by concatenating sub-feature vectors with reduced dimensions.

## III. Results

The performance of ACSFC method tested with three subjects is shown in Table I. Performance was calculated as percentages;

$$P = \frac{N_s}{N} \times 100 \quad (4)$$

$P$ – Performance
$N_s$ – Number of successfully identified mental tasks
$N$ – Total number of mental tasks

EEG data recorded from three subjects was preprocessed with band pass filtering and feature vectors were constructed using band powers. kNN classifier has been used for classifying the feature vectors reduced by ACSFC method. For comparison purposes performance of all three subjects without applying ACSFC method (Using all the features in all 20 channels) has been included in the Table I.

TABLE I
PERFORMANCE OF ACSFC METHOD USING kNN CLASSIFIER. NUMBER OF EEG DATA POINTS USED IN THE CALCULATION WAS 2048.

| | | | |
|---|---|---|---|
| **SUBJECT 1** | | | |
| **Mental Tasks** | BL & IMRAM | BL & IMDAM | IMRAM & IMDAM |
| **Classification** | kNN ($k$ = 5) | kNN ($k$ = 4) | kNN ($k$ = 3) |
| **Channels selected by ACSFC ( Number of features selected per channel)** | F7 (1) P4 (1) O1 (1) O2 (1) | FP1 (2) | FP1 (1) |
| **Performance with ACSFC** | 77% | 90% | 85% |
| **Performance with all 20 channels** | 72% | 85% | 82% |
| **SUBJECT 2** | | | |
| **Classification** | kNN ($k$ = 10) | kNN ($k$ = 7) | kNN ($k$ = 6) |
| **Channels selected by ACSFC ( Number of features selected per channel)** | CZ (2) T5 (5) | C3 (1) CZ (6) T5 (7) | FP1 (2) FP2 (4) |
| **Performance with ACSFC** | 95% | 98% | 90% |
| **Performance with all 20 channels** | 93% | 93% | 97% |
| **SUBJECT 3** | | | |
| **Classification** | kNN ($k$ = 10) | kNN ($k$ = 7) | kNN ($k$ = 5) |
| **Channels selected by ACSFC ( Number of features selected per channel)** | FP2 (1) FZ (3) F8 (1) C3 (1) CZ (1) T5 (8) | FP1 (4) FP2 (3) F8 (1) C3 (1) T4 (1) G (1) | FZ (3) F8 (1) T5 (7) O1(1) |
| **Performance with ACSFC** | 75% | 67% | 73% |
| **Performance with all 20 channels** | 70% | 65% | 75% |

## IV. Concluding Remarks

It is evident from the Table I that by introducing ACSFC method significantly reduce the dimensions of the feature vectors. As an example, number of dimensions in the initial feature vectors for each subject was 640. However, number of dimensions in these feature vectors has come down below 16 after applying ACSFC method. Except for two data sets, classification carried out with reduced feature vectors performed better than initial feature vectors. Nevertheless, further performance evaluation with other classifiers is required before making any general conclusions on overall performance of the ACSFC method.

## REFERENCES


[1] J. R.Wolpaw, N. Birbaumer,D. J.McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” Electroenceph. Clin. Neurophysiol., vol. 113, no. 6, pp. 767–791, June 2002.

[2] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, “Brain-computer interface - A new communication device for handicapped persons,” J. Microcomput. Applicat., vol. 16, pp. 293–299, 1993.

[3] A. Cichocki, Y. Washizawa, T. Rutkowski, H. Bakardjian, Anh-Huy Phan, Seungjin Choi, Hyekyoung Lee, Qibin Zhao, Liqing Zhang, and Yuanqing Li, “Noninvasive BCIs: Multiway Signal-Processing Array Decompositions,” IEEE Computer Society, vol.41, pp.34 – 42, 2008.

[4] N. Birbaumer, A. Kubler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann, and H. Flor. “The thought translation device (TTD) for completely paralyzed patients,” IEEE Transactions on Rehabilitation Engineering, vol.8(2), pp.190–193, June 2000.

[5] P. R. Kennedy, R. A. Bakay, M. M. Moore, and J. Goldwaithe, “Direct control of a computer from the human central nervous system,” IEEE Trans. Rehab. Eng., vol. 8, pp. 198–202, June 2000.

[6] E. E. Sutter, “The brain response interface: Communication through visually guided electrical brain responses,” J. Microcomput. Applicat., vol.15, pp. 31–45, 1992.

[7] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, “An EEG-based brain-computer interface for cursor control,” Electroenceph.Clin. Neurophysiol., vol. 78, no. 3, pp. 252–259, Mar. 1991.

[8] D. J. McFarland, G.W. Neat, and J. R.Wolpaw, “An EEG-based method for graded cursor control,” Psychobiol., vol. 21, pp. 77–81, 1993.

[9] J. Vidal, “Toward Direct Brain–Computer Communication, in Annual Review of Biophysics and Bioengineering,” L.J. Mullins, Ed., Annual Reviews, Inc., Palo Alto, Vol. 2, pp. 157-180, 1973.

[10] J. Vidal, “Real-Time Detection of Brain Events in EEG,” in IEEE Proceedings, 65-5:633-641, May 1977.

[11] F. Lotte et al., “A Review of Classification Algorithms for EEG-Based Brain-Computer Interfaces,” J. Neural Eng., vol. 4, pp. R1-R13, 2007.

[12] H. Ramoser, J.Müller-Gerking, and G. 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.

[13] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, “Spatiospectral filters for improving the classification of single trial EEG,” IEEE Transactions on Biomedical Engineering, vol. 52, no. 9, pp. 1541–548, 2005.

[14] A. J. Bell, and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” Neural Computation, vol. 7, no. 6, pp. 1129–1159, 1995.

[15] A. Hyvarinen, “A family of fixed-point algorithms for independent component analysis,” in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP ’97-, Munich, Germany), vol. 5, pp. 3917–3920, April 1997.

[16] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, “A blind source separation technique using secondorder statistics,” IEEE Transactions on Signal Processing, vol. 45, no. 2, pp. 434–444, 1997.

[17] M. Naeem, C. Brunner, and G. Pfurtscheller, “Dimensionality Reduction and Channel Selection of Motor Imagery Electroencephalographic Data,” Computational Intelligence and Neuroscience, 537504, 2009.

[18] Richard O. Duda, Peter E. Hart, and David G. Stork, “Pattern Classification,” $2^{nd}$ Edition, 680 pages, Nov. 2000, ©2001

[19] S. S. Zahmeeth, “Brain Computer Interface (BCI) based on Electroencephalographic (EEG) patterns due to new cognitive tasks,” M.Phil Thesis, University of Peradeniya, 2010.