Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Spatio-Spectral & Temporal Parameter Searching using Class Correlation Analysis and Particle Swarm Optimization for a Brain Computer Interface

Abdul R Satti, Damien Coyle and Girijesh Prasad
Intelligent Systems Research Centre, School of Computing and Intelligent Systems
University of Ulster
Derry, UK
Satti-ar@email.ulster.ac.uk

***Abstract*— Distinct features play a vital role in enabling a computer to associate different electroencephalogram (EEG) signals to different brain states. To ease the workload on the feature extractor and enhance separability between different brain states, numerous parameters, such as separable frequency bands, data acquisition channels and time point of maximum separability are chosen explicit to each subject. Recent research has shown that using subject specific parameters for the extraction of invariant characteristics specific to each brain state can significantly improve the performance and accuracy of a brain-computer interface (BCI). This paper focuses on developing a fast autonomous user-specific tuned BCI system using Particle Swarm Optimization (PSO) to search for optimal parameter combination based on the analysis of the correlation between different classes i.e., the R-Squared (R^2) correlation coefficient rather than assessing overall systems performance via performance measure such as classification accuracy. Experimental results utilizing eight subjects are presented which demonstrate the effectiveness of the proposed methods for fast & efficient user-specific tuned BCI system.**

***Keywords*—particle swarm optimisation, brain computer interface, parameter search, correlation coefficient.**

I. Introduction

A BCI system transforms the metabolic brain activity or electrophysiological (EEG) signals to provide a direct communication pathway between the brain and an external device. The goal of such an interface is to provide effective communication by accepting commands directly encoded in neurophysiological signals. Patients suffering from motor impairments, severe cerebral palsy and spinal cord injuries (SCI) may use a BCI system as a substitute communication pathway which relies only on the mental imagination and not on neuromuscular control [1]-[6]. Recently, there has been a tremendous growth in BCI research exploring a broad range of BCI applications in different fields: communication, environmental control, robotics and mobility, and neuroprosthetics. However, for a BCI system to provide an alternative to motor controlled devices, it is essential for the BCI to accommodate numerous critical requirements for reliability, high information transfer rates, robustness, adaptation and safety.

It has been shown that using subject-specific parameters such as discriminable frequency bands [7]-[9], time-point of maximum separability for training the classifier [9] and channel selection [11] for the extraction of invariant characteristics specific to each brain state has proven to significantly improve the performance and accuracy of a BCI [9]. However, these parameters have to be tuned explicitly to each subject that increases the computational overhead which, eventually, forms a bottleneck for the efficient and prompt use of the application. Parameter tuning can also be time consuming and tiresome on the part of the subject and the researcher.

Recently particle swarm optimization (PSO) has been used in combination with common spatial pattern (CSP) filtering to tune the frequency band, with classification accuracies (CA) during cross-validation serving as the objective measure of performance [9]. PSO can find the global optimal much quicker than an iterative search through the search space and thus has proved to be effective, however, the need to check overall system performance (i.e., overall CA) during system training can result in a significant time overhead [9]. For this reason, the correlation coefficient between different classes has been investigated in combination with the PSO for fast autonomous tuning of parameters, which include the most separable frequency bands and channels and time-point of maximum separability specific to each subject. R-Squared (R^2) represents the square of a correlation coefficient and is a measure of the proportion of variability between different classes. The value for this coefficient lies between 1 and 0; a higher coefficient value corresponds to higher correlation between classes and vice versa. Based on this fact, the time point, frequency bands and channels having minimum correlation between the two classes can be chosen (cf. Section III for brief description on R^2). It is important to note that parameter tuning is carried out before applying Common Spatial Pattern (CSP) filtering (cf. section III.D for a brief

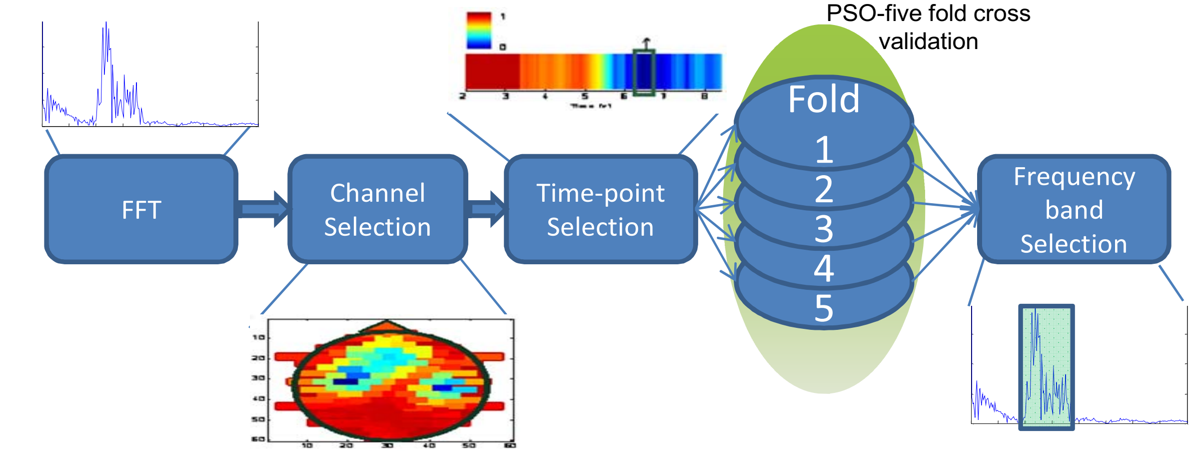

Figure 1. User-specific parameter tuning.

description on CSP) in contrast to CA based approach [9], which helps to reduce the computations to a large extent.

The remainder of the paper is organized into four sections. Section II contains details on the datasets and acquisition procedure. Section III discusses the channel, time-point and frequency band selection based on the R^2 coefficient of determination, PSO for optimized parameter search and CSP filtering. Results are presented and discussed in section IV.

II. Data Acquisition and Configuration

The datasets used in this analysis are from the BCI competition III [12]. The dataset IIIa (subjects 6-8) was recorded from three subjects using a 64 channel amplifier and was filtered between 1 and 30Hz. The subject sat on a comfortable chair and had to perform left/right hand, tongue or foot movement imagination according to the cue on the screen. Each trial begins with a blank screen and a beep sounds at t=2s. At t=3s an arrow appears on the fixation cross indicating the imagined movement to be executed. The subject performs the imagery task until t=7s. Each of these four tasks was performed 10 times in a random order [12].

The dataset IVa (subjects 1-5) was recorded with slight variation from dataset IIIa. 118 EEG channels were used for data acquisition and arrows were displayed for 3.5s indicating the corresponding task to be performed. The presentation of target arrows was stopped by periods of random length, 1.75 to 2.25 s, in which the subject could relax [12].

III. Methodology

It is important to note that the channel and time point selection is carried out before frequency band selection. As a result, the computational time is reduced by a large extent since the most separable frequency band is searched at only one time point instead of the whole trial, i.e., at the time point of minimum correlation between different classes of the selected channels. Fig. 1 shows the sequential representation of the parameter tuning model.

A. *Channel Selection*

Channel selection is an assistive process for selecting the time point of maximum separability. It has been observed that choosing the channels based on minimum correlation for different classes increases the mean separability over the selected channels, hence, gives a better estimation of the time

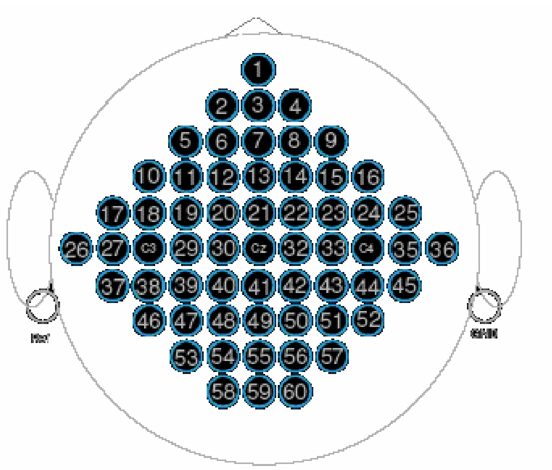

Figure 2. Position of the electrodes (Dataset IIIa)

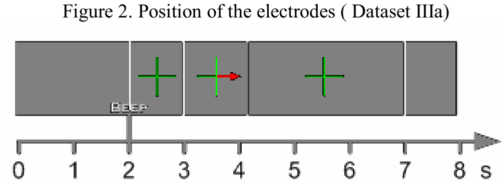

Figure 3. Training paradigm and timing (Dataset IIIa)

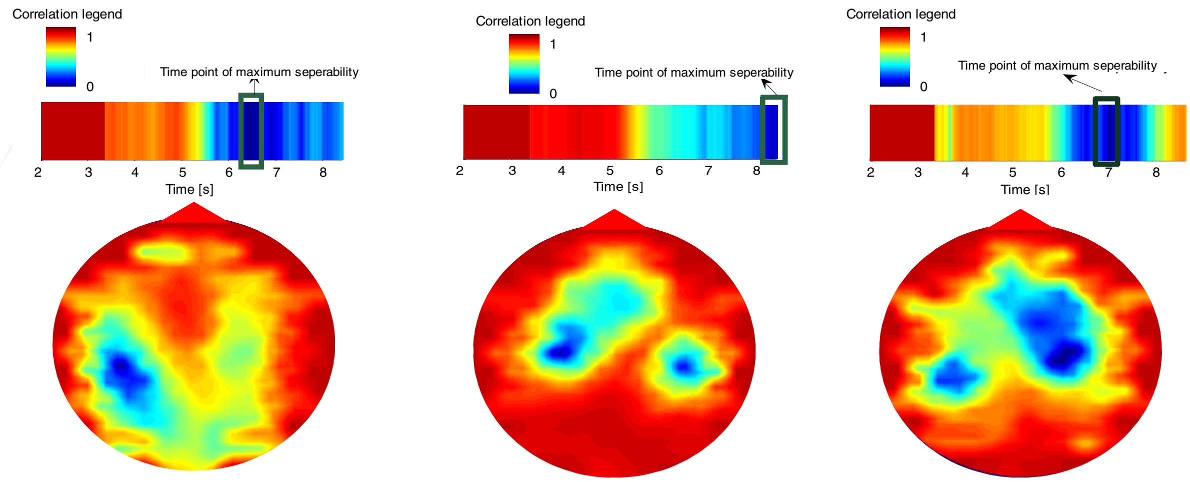

Figure 4. Topographic correlation map for subjects 1,2 and 3

point where the maximum separability might lie.

Let $fa_{t,c}$ be the mean Fast Fourier Transform (FFT) for a particular frequency over the band 8-26 Hz of one class and $fb_{t,c}$ for the other class calculated at a time point t and channel c of a trial with a window size of one second. The "separability" of the data set is measured at each time point t and every channel c by computing the R-squared coefficient of determination using the standard band of [8→26] Hz.

$$R_{t,c} = 1 - {}^{(D_{t,c})}\!/\!_{(T_{t,c})} \tag{1}$$

where,

$$T_{t,c} = \sum_{i=g}^{h} \left(fa_{t,c,i} - \left(\frac{\sum_{j=g}^{h} fa_{t,c}}{h-g} \right)^2 \right), h=26 \ \ and \ \ g=8 \ \ [8 \rightarrow 26]Hz \tag{2}$$

and

$$D_{t,c} = \sum_{i=8}^{26} (fa_{t,c,i} - fb_{t,c,i})^2 \tag{3}$$

Channel selection is based on the principle that the lower the value of R^2 between the band power of two classes, the higher the separability and thus the channel is selected. Based on the R^2 values using (1)-(3) for each channel, the decision is made in the favor of the q most uncorrelated channels. Channels are first sorted in the order of seperability, using (4) and (5):

$$\delta = \arg\min(R_{t,c}) \tag{4}$$

where δ contains the indices of the sorted channels based on maximum separability and *argmin( )* computes the indices of the minimum element. Thus the selected matrix of channels H can be computed from the R matrix as:

$$H_{t,i} = R_{t,\delta_i} \quad , \quad i = 1....q \tag{5}$$

where q is the desired number of channels.

A topographic map is displayed in Fig. 4 depicting the most separable areas for extracting meaningful discriminable information for BCI system processing (chosen based on R^2).

B. *Time point Selection*

Selected channels are then averaged over time to obtain a time point of maximum separability for the position of the time window selection. This is obtained using (6):

$$\tau_t = \frac{\sum_{i=1}^{q} k_{t,q}}{q} \quad , \quad t = 1...e \tag{6}$$

where e is the number of samples in one trial.

$$tp = \arg\min(\tau_t) \tag{7}$$

This time-point (tp) represents the least average correlated moment of time across all the trials. This time point is utilized for position of the time window (usually 1 sec) for training the classifier.

C. *PSO for frequency band selection*

Section III.A and III.B outline the methods for the channel and time point selection based on a standard band of 8→26 Hz. As outlined, performance can be optimized by chooing subject-specific bands [7]-[9]. Subject specific-frequency bands can be chosen using PSO for fast parameter search. PSO is a stochastic, population-based evolutionary computer algorithm for parameter tuning and problem solving [14]. It involves simple primitive mathematical operators, and is computationally inexpensive in terms of both memory requirements and speed [14]. Initial swarms which are a set of randomly generated solutions circulate in the search space over a number of iterations searching for the optimal solution. The algorithm below defines the steps for the implementation of PSO for a BCI frequency band selection:

1) *Define a search space:* The search space for the frequency band selection is defined as two dimensional space where one dimension represents the lower cutoff frequency and the other as the upper cutoff.
2) *Define a fitness function:* Fitness is evaluated according to the desired optimization parameter, R-square using (1).
3) *Initialize swarms:* Particles are randomly initialized in the search space.
4) *Moving particles in the search space:* Perform the following steps on each particle in parallel.
   a) *Evalute each particle's fitness:* The fitness value for each particle is calculated utlizing its current position and orientation.
   b) *Update 'pbest' and 'gbest':* update the *'pbest' (p)* which is the particle's best fitness value achieved so far and *'gbest' (g),* which is the global best fitness value for all particles using the fitness function for each particle.
   c) *Update particles velocity:* The velocity of the particles are updated as:

$$V_{t+1}^{i} = V_{t}^{i} + c_1.rand(\,).(p_i - X_t^i) + c_2.rand(\,).(p_g - X_t^i) \tag{7}$$

   where constants c_1 and c_2 represent the weights of the stochastic acceleration terms that draw each particle i toward *pbest* and *gbest* positions.
   d) *Update particles position:*

$$X_{t+1}^{i} = X_t^i + V_t^i \tag{8}$$

5) *Loop back to step 4*: restart at step 4 until all the particles converge to a common 'gbest'.

5-fold cross-validation was carried out, where the data was partitioned into a training set (50%) and a validation set (50%). PSO was implemented on each fold, thus five filters with optimum fitness value were obtained. To obtain the best fit filter, these five chosen bands were averaged using the weighted coefficient a_n i.e. the classification accuracy (CA) of the respective fold. CA is calculated using CSP for mapping into a surrogate space (cf. section III.D) where log variances in the surrogate space are classified using Linear Discriminant Analysis (LDA) [9].

If S_n represents the lower cutoff frequency for each n^{th} fold and P_n as the upper cutoff frequency, then the pass band filter [$F_s \rightarrow F_p$] is calculated as:

$$F_s = \frac{\sum_{n=1}^{fold}(a_n.S_n)}{\sum_{n=1}^{fold} a_n.} \tag{9}$$

$$F_p = \frac{\sum_{n=1}^{fold}(a_n.P_n)}{\sum_{n=1}^{fold} a_n.} \tag{10}$$

Here, 'a_n' imparts a weighted summation factor in the favor of the band achieving better classification accuracy in the cross validation (cf. Fig. 5). This frequency band (pass-band: $F_s \rightarrow F_p$) is then used to filter the data for training and testing.

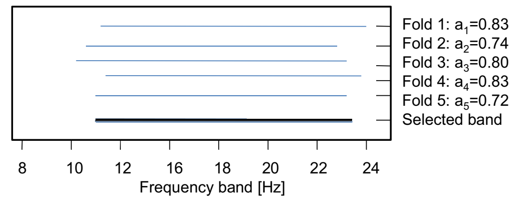

Figure 5. Band selection using (9) and (10)

D. Common Spatial Pattern

CSP filtering involves linearly projecting the multichannel EEG data into a surrogate data space by a weighted summation of the appropriate channels. This projection is based on the simultaneous diagonalisation of the covariance matrices from both classes [6].

Let $X \in R^{N\times T}$ be a single trial EEG matrix for N-channel EEG where T is the number of samples in a trial. A single trial is one specific imagined movement in a cue based paradigm depending on the direction of the arrow.

The normalized covariance for the two classes can be represented as:

$$\sum_k = \frac{1}{n}\sum_{i=1}^{n} X_i X_i{}' \quad , \quad (k \in \{1,2\}) \tag{11}$$

where n is the total number of trials for the class k.

In Matlab, simultaneous diagnolisation can simply be implemented as

$$W = eig(\textstyle\sum_k , \sum_k + \sum_{k+1}) \tag{12}$$

With the mapping matrix W, the trial X is projected as

$$Z = W \times \mathrm{X} \tag{13}$$

By construction, the variance for one of the classes is largest in the first row of Z and decreases for the subsequent rows and vice versa for the other class [15]. The appropriate number of Eigenvectors from both sides is chosen as filters; generally between 2 to 5 from either side of the Eigenvector matrix is optimal [15].

IV. Results & Discussion

Fig. 6 shows the effect of increase in mean separability as a result of choosing channels using the correlation coefficient. It is evident that choosing the most discriminable channels increases the mean separability over the channels to a large extent. Fig. 6 also illustrates the importance of not using all the channels for feature extraction and classification, as the non-event related channels add ambiguous information to the dataset, thus, resulting in separability attenuation of different classes. Channel selection has helped to eradicate a significant amount of redundant information, thus, the time point of maximum separability is effectively more prominent and

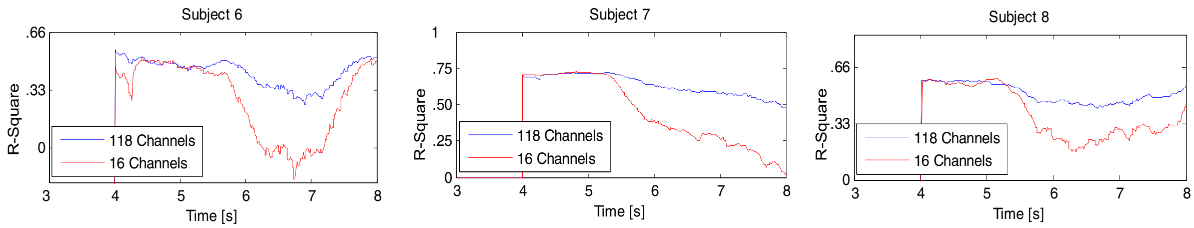

Figure 6. R-Square correlation comparison plot between 16 and 118 channels for subjects 1, 2 and 3.

TABLE I.
CLASSIFICATION ACCURACIES AND BAND SELECTION

Subject	**Standard Band (8-26 Hz)**	**Classification accuracies** [%]		**Selected Frequency Bands** [Hz]	
		R^2 *approach*	*CA Based approach*	R^2 *approach*	*CA Based approach*
1	57.14	83.57	83.57	11->17	11->18
2	95.00	91.42	95.00	8->18	10->19
3	58.57	76.57	73.57	8->20	8->23
4	59.28	91.42	96.42	10->16	11->16
5	91.42	91.42	95.71	8->24	8->23
6	91.65	83.28	90.27	8->20	12->22
7	65.65	67.28	65.65	12->22	12->24
8	100.0	100.0	100.0	8->24	8->17
Mean	**76.87**	**84.981**	**86.804**		

detectable.

Table I compares the three approaches for subject specific parameter tuning using a standard band (8-26 Hz), PSO based band selection using CA as fitness function and PSO based band selection using R^2 as the fitness function. The results signify that the R^2 approach improve the performance by ~8% for ten subjects as compared to using standard band. To determine if the improvement is statistically significant, Repeated Measures One-way Analysis of Variance (RM-ANOVA) was performed and a *P* value of 0.05 is achieved, indicating the significance of improvement. The CA based approach on the other hand improves the CA by ~10%. It is important to note here that the CA based approach is bound to give better accuracy as it computes the fitness based on CA, thus finds out the best possible band for optimum classification accuracy. On the other hand the R^2 approach computes the fitness based on the correlation difference between different bands, hence reduces the computations to a great extent by not running the whole system to compute classification accuracy. There is a difference of ~2% between the two approaches which is non-significant (p= 0.154), and therefore is justified as a tradeoff increase in computational efficiency. We previously reported in [9] that PSO has helped to improve the computational performance by a factor of 8-10, however, the R^2 approach has helped to further decrease the computational time by a factor of 1.5-2.5 relative to the PSO search performance. This drop is associated to time point of maximum separability selection at an early stage using R^2 and avoiding the CSP filtering and classification during PSO to obtain the fitness value i.e. the overall system performance.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan. "Brain-computer interfaces for communication and control", *Clinical Neurophysiology*, vol. 113, pp. 767-791, 2002.

[2] S.M. Kosslyn, G. Ganis, W.L. Thompson, "Neural foundations of imagery," *Nat Rev Neurosci,* vol. 2, pp. 635-642, 2001.

[3] K P. Liu, Ch C. Chan, T. M. Lee, C. W. Hui-Chan, " Mental Imagery for Promoting Relearning for People After Stroke: A Randomized Controlled Trial," *Arch Phys Med Rehabil* vol 85, pp. 1403-1408, 2004.

[4] S.G. Mason, A. Bashashati, M. Fatoruechi, K. F. Navarro, and G. E. Birch, "A comprehensive survey of brain interface technology designs," *Annals of Biomed. Eng.,* vol. 35, no. 2, pp. 137-169, 2007.

[5] D. Coyle. Intelligent Preprocessing and Feature Extraction Techniques for a Brain Computer Interface, *PhD Thesis, University of Ulster*, N. Ireland, 2006.

[6] D. Coyle, T.M. McGinnity and G. Prasad, "A multi-class BCI with SOFNN-based prediction preprocessing," *IEEE World Congress on Computational Intelligence*, pp. 3695-3702, 2008.

[7] M. Zhao, M. Zhou, Q. Zhu, P. Yang, "Feature Extraction and Parameters Selection of Classification Model on Brain-Computer Interface," *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering*,pp. 1249-1253, 2007

[8] P. Herman, G. Prasad, T.M. McGinnity, D. Coyle, "Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 16, No. 4, pp. 317-326, 2008.

[9] D. Coyle, G. Prasad and T.M. McGinnity, "A time-frequency approach to feature extraction for a brain-computer interface with a comparative analysis of performance measures", *EURASIP JASP, Trends in Brain-Computer Interfaces (special issue),* vol. 19, pp. 3141-3151, Nov. 2005.

[10] A. Satti, D. Coyle, G. Prasad, "Optimal Frequency Band Selection with Particle Swarm Optimization for a Brain Computer Interface," *Proceedings of the Workshop on Evolutionary Computing Lecture Series by pioneers*, pp. 15-18, 2008.

[11] D. Erdogmus, A.Adami, M. Pavel, S. Mathan, "Salient EEG Channel Selection in Brain Computer Interfaces by Mutual Information Maximization," *IEEE-EMBS ,* pp. 7064-7067, 2006

[12] B. Blankertz, BCI-Competition III, 2005 http://ida.first.fhg.de/projects/bci/competition_iii/

[13] B. Blankertz, G. Dornhege, M. Krauledat, K. Müller, and G. Curio, "The non-invasive berlin brain-bomputer interface: fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol 37, No. 2, pp:539-550, 2007

[14] J. Kennedy, R. Eberhart, " Particle swarm optimization," *Proceedings., IEEE International Conference on Neural Networks*, vol. 1, pp. 1942-1948, 1995.

[15] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis for a brain-computer interface (BCI) with subject-specific spatial patterns," *IEEE Trans. Rehab. Engng.* vol. 8, no. 4, pp. 447-456, 2000.