Aler, R., Galván, I.M. (2015).Optimizing the number of electrodes and spatial filters for Brain–Computer Interfaces by means of an evolutionary multi-objective approach. *Expert Systems with Applications*, 42(15-16), pp. 6215-6223.

DOI: 10.1016/j.eswa.2015.03.008

Optimizing the Number of Electrodes and Spatial Filters for Brain-Computer Interfaces by means of an Evolutionary Multi-objective Approach

Ricardo Aler[a,*], Inés M. Galván[a]

[a] *Universidad Carlos III de Madrid, Avda. Universidad 30*
28911 Leganés, Spain

Abstract

In this paper two important issues in BCI systems are addressed. First, EEG-based BCI systems require to attach many electrodes on the scalp of subjects but their placement involves a laborious process. Second, the preprocessing of EEG signals by means of filters plays a crucial role in the success of EEG classification. An evolutionary multi-objective approach is proposed for optimizing both the number of electrodes and the classification error achieved by the spatial filter coupled to a learning classifier. Multi-objective algorithms have the advantage that they return a Pareto front, that is, a set of solutions that represent all the possible tradeoffs between the number of channels and the classification accuracy, from where the practitioner can choose. An empirical comparison with two other heuristic methods shows that the simultaneous optimization of electrodes and filters can provide better results than other approaches that reduce the number of electrodes but do not modify the filter, specially if low error solutions with few electrodes are desired.

Keywords: Brain-Computer Interfaces, Multi-Objective Optimization, Electrode Selection

1. Introduction

Brain-Computer Interface (BCI) research aims at developing a direct communication between brain and machine in order to provide an alternative control pathway for the user. One of the main aims of BCIs is to restore motor function for disabled patients, suffering for instance from Amyotrophic Lateral Sclerosis or the locked-in syndrome [16]. BCIs can provide to those patients control over wheelchairs [3, 17] or new communication channels through virtual keyboards [9, 23].

*Corresponding Author
Email addresses: `aler@inf.uc3m.es` (Ricardo Aler), `igalvan@inf.uc3m.es` (Inés M. Galván)

There are many varieties of BCIs, but the most common are the EEG-based non-invasive BCIs. In a typical BCI setting, subjects are required to perform specific mental tasks while their EEG signals are being recorded. The EEG signal is amplified and sent to the computer where it can be analyzed by algorithms. Depending on the kind of BCI, different features can be detected on the EEG signal and transformed into actions (such as moving a cursor on the screen or controlling a wheelchair). These features can be either voluntarily generated by the user (such as slow cortical potentials or sensorimotor rhythms) or elicited by visual or auditory stimulation (event-related potentials or steady-state evoked potentials) [15].

EEG-based BCIs capture the EEG signal by means of electrodes or channels disposed on the subjects scalp. The EEG signal is recorded from the electrodes, then the signal is preprocessed using different kinds of filters and, finally, it is classified. The result of the classification can be used to control a device. For instance, a BCI that is able to classify thoughts into hand-right and left-hand imaginary movement can be used to move a screen cursor right and left. Machine Learning techniques are typically used for computing the classifier from user data [8, 20]. The learning of the classifier and the adjustment of the filters is carried out off-line. They are calibrated by using the data from an acquisition/calibration session where the user is instructed to perform certain mental tasks while the EEG is recorded.

Preprocessing / filtering raw signals plays a crucial role in the classification of EEG signals [7, 26]. In particular, EEG-based BCI's suffer from low spatial resolution: only half the contribution to each electrode comes from sources within a 3 cm radius below each electrode [22]. In this context, spatial filters are useful because they can generate a more localized signal for every electrode. Previous research [4, 10, 5] has shown that hybridization of global optimization techniques and machine learning classification is useful for different issues in the BCI domain, like feature selection / extraction and improving classification. In particular, previous work by the authors [1] is focused in obtaining spatial filters for improving the performance of a classifier. In that work, the Covariance Matrix Adaptation Evolution Strategy (CMAES) [12] is used to optimize a spatial filter for a Fisher Discriminant classifier.

The issue that will be addressed in this paper is related to the reduction in the number of channels. The number of electrodes in typical EEG systems may range from 10 to 256 channels. A large number of electrodes can be useful for medical and diagnostic purposes. For real BCI applications, the classification accuracy that can be achieved by using many electrodes is highly important. But on the other hand, the placement of electrodes on the scalp is slow and it generally involves a laborious process. Therefore, it is important to find out those electrodes which have a null or very small contribution for classification. Neurological knowledge can be used for electrode selection. For instance, it is well known that hand motor imagery can be detected at the C3 and C4 locations. This information can be very useful, but it is also known that different subjects respond differently and the subject-dependent optimal location of electrodes may vary [7]. Therefore, it is important to automatically determine the degree of relevance of electrodes for the user-dependent classification task.

A way to tackle this problem is to consider initially a large number of electrodes and then use some methods to select the best set of channels for each subject. For this purpose, evolutionary algorithms have already been used within a single-objective formulation where the classification error is optimized while at the same time penalizing solutions with many electrodes [18, 14, 2]. The single-objective approach returns a single solution that represents a good trade-off between the number of electrodes and classifier accuracy.

Electrode selection can be formulated more naturally within a multi-objective framework because both goals (accuracy and number of electrodes) can be optimized simultaneously. The end result of a multi-objective algorithm is not one, but a set of solutions (the so-called non-dominated solution set or Pareto front) [24]. This front represents the best tradeoffs between the number of channels and some measure related to classification accuracy. In other words, for every possible number of electrodes, the Pareto front contains the solution that provides the optimal accuracy that can be achieved with them. This way, the practitioner does not have to commit to a single solution like in the single-objective case, but can select the electrodes that achieve the classification performance required at a particular time.

Multi-objective evolutionary algorithms (MOEAs) have already been used in the BCI domain for electrode selection. In [13] authors present a preliminary study comparing a multi-objective evolutionary algorithm based on a decomposition approach into optimization sub-problems and a Multi-objective Particle Swarm Optimization method. In [21] the decomposition approach is introduced into a Multi-objective Particle Swarm method. In those works, channel selection in the classification of continuous EEG without trial structure is carried out by minimizing two objectives: the classification error and the number of channels. In [25] a Multi-objective Binary Particle Swarm optimization is also proposed to handle the problem of channel selection for multi-channel EEG signals. Two objectives are optimized: the number of selected channels and the mutual information (instead of classification accuracy) achieved by a classifier (Support Vector Machines, Back-propagation, and Nearest Neighbor). In all the previous approaches, particles contain binary values, each one of them representing the selection (or removal) of a channel.

The goal of our work is to use a multi-objective approach to provide a set of solutions (Pareto front) that represents all the possible best trade-offs between classification error and number of electrodes. Each solution is a pair made of a collection of electrodes (like in other approaches) but also a spatial filter. Both of them are optimized simultaneously within the multi-objective framework.

The chromosome has to encode candidate solutions made of a spatial filter and selected electrodes. In order to simplify the chromosome encoding, our framework allows to take advantage that the components of the spatial filter matrix represent the weights or coefficients of the electrodes (small weights mean small influence of the electrode). Therefore, a threshold is also included in the chromosome and it is used to determine which electrodes are selected (those whose weight is larger than the threshold). Thus, only a threshold has to be encoded in the chromosome (in addition to the spatial filter matrix), instead of a longer binary electrode-selection mask, as in other approaches.

To test the performance of this approach, it will be compared with other methods that compute an optimized spatial filter for the whole set of electrodes, and then electrodes are removed according to some heuristics (but the spatial filter itself remains fixed).

The rest of the paper is organized as follows. Section 2 describes the multi-objective approach, including a brief description about how spatial filter is used to preprocess raw EEG data, the chromosome encoding, and the fitness function. Section 3 includes the experimental validation of the proposed multi-objective approach and the conclusions are summarized in Section 4.

2. Multi-objective Approach for Electrode Selection

The aim of this section is twofold. First, the main concepts related to multi-objective optimization and to preprocessing (mainly the spatial filter) and classifying EEG signals. Second, the two main components of the multi-objective optimization process used in this paper will be described: how solutions are encoded and how the actual fitness function is computed.

2.1. Introduction

The aim of this introductory section is to first describe the main concepts related to multi-objective optimization, and second, to explain the way EEG signals are preprocessed (mainly by applying the spatial filter) and classified, as done in this article.

Multi-objective optimization

Multi-objective optimization is concerned with problems where more than one goal has to be optimized. For non-trivial problems, there is no single solution that optimizes all objectives because they conflict. This is the case of this paper, where reducing the number of electrodes also increases the classification error, and viceversa. Therefore, instead of a single solution like in single-objective optimization, a set of solutions called non-dominated set or Pareto front, is obtained. Solutions in the non-dominated set are optimal in the Pareto sense: none of the objective functions can be improved without worsening some of the other objective values. Without further user preferences, all Pareto optimal solutions can be considered equally good, because none of them is better than the others according to Pareto. It is up to the user to select one of them according to his requirements at the moment of selection. An example of Pareto front can be seen later in the article in Figure 1 (left figure, label 'Multiobjective'), which represents a non-dominated set with two objectives: classification error (y-axis) and number of electrodes (x-axis). The goals to be optimized constitute the objective space. In order to find the optimal values for the objectives, the optimization algorithm can explore the so-called decision space, which in this work is made of two components: the spatial filter and a threshold that determines the subset of electrodes to be used. The objective space and the decision space are related: the classification error is computed by preprocessing the EEG signal, restricted to the subset of electrodes, by means of the spatial filter. Then, a classifier can be constructed from the preprocessed data. And finally, the classification error can be obtained. The following part of this section is dedicated to explaining this preprocessing and classification process.

Preprocessing and classifying EEG signals

Let M be a $t \times c$ data matrix that contains c time series, each one recorded from each of the c electrodes. t is the number of time instants in the signal to be classified. The spatial filter S is a linear transformation of the original data. S is represented by a $c \times c'$ matrix, where c is the number of electrodes and $c' \leq c$ is a parameter. Spatial filtering is carried out by just multiplying the original data M by matrix S: $M' = M * S$. This means that the spatial filter S transforms the c channels of M into the c' channels of M'.

Classification is carried out in two steps. First, the signal M is preprocessed by means of the spatial filter S and the Fast Fourier Transform (**FFT**). The reason for moving from the time-domain to the frequency-domain by means of the **FFT** is that EEG patterns can be better detected at the frequency domain [15]. For instance, it is known that imagination of movements can be observed around the 12Hz frequency-band, although the exact band depends on the user [1]. Second, once preprocessed, the signal is classified by a Linear Discriminant (LDA), also known as the Fisher Discriminant.

Preprocessing is carried out by first applying the spatial filter S to M as explained in the previous paragraphs, and then transforming from the time to the frequency domain (by means of the **FFT**), according to Eq. 1. **FFT** returns complex numbers (modulus and phase) but for classification purposes, only the modulus is used [7], extracted by the modulus operator $|\ |$. Rows of F from 1 to $t' = \frac{t}{2}$ represent the frequency bands from 0 Hz to $\frac{t}{2}$ Hz with a resolution of $\frac{f}{t}$ Hz, where f is the sampling frequency. That is, these rows represent the frequency bands $[0, \frac{f}{t}]$ Hz, $[\frac{f}{t}, 2 * \frac{f}{t}]$ Hz, etc.

$$F = (f_{i,j})_{t' \times c'} = |\mathbf{FFT}(M * S)| \tag{1}$$

The expression $M' = (m'_{i,j})_{t \times c'} = M * S$ in Eq. 1 transforms signal M from the c real electrodes to c' spatially filtered channels. Let $\mathbf{m'_{.j}}$ represent column j from matrix M' (or equivalently, the jth filtered channel). Eq. 2 shows that the application of the spatial filter ($M' = M * S$) can be understood as creating filtered channels $\mathbf{m'_{.j}}$ which are linear combinations of real electrodes $\mathbf{m_{.k}}$.

$$\mathbf{m'_{.j}} = \sum_{k=1}^{k=c} s_{k,j}\mathbf{m_{.k}} = s_{1,j}\mathbf{m_{.1}} + s_{2,j}\mathbf{m_{.2}} + \ldots + s_{c,j}\mathbf{m_{.c}} \tag{2}$$

Once the signal has been preprocessed, classification is carried out by a linear classifier that uses the $f_{i,j}$ in Eq. 1 as features. The $f_{i,j}$ belong to the frequency domain and represent the ith frequency-band of the jth filtered channel. Eq. 3 displays the LDA with weights $w_{i,j}$ and bias b. H is the Heavyside step function and returns either 0 and 1 (i.e. class 0 and class 1). Eq. 3 is valid for two class problems. For problems with N classes, the one versus all approach is used, and N discriminants are learned, from $D^{(1)}$ to $D^{(N)}$. Each discriminant separates one of the classes from the rest. In this case, the classification is carried out by Eq. 4.

$$D(F) = H\left(\sum_{j=1}^{j=c'}\sum_{i=1}^{i=t'} w_{i,j} f_{i,j} + b\right) \tag{3}$$

$$D(F) = \arg\max_p D^{(p)}(F) = \arg\max_p \left(\sum_{j=1}^{j=c'}\sum_{i=1}^{i=t'} w^{(p)}_{i,j} f_{i,j} + b^{(p)}\right) \tag{4}$$

Weights $w^{(p)}_{i,j}$ and biases $b^{(p)}$ can be learned from data by Fisher's Linear Discriminant Analysis (LDA) [11]. But the spatial filter S used to preprocess the data must first be adjusted in order to maximize the classification accuracy. As mentioned in the Introduction, it is also intended to minimize the number of electrodes. In this paper, a multi-objective approach has been chosen to optimize both goals.

In order to use an evolutionary algorithm, two components must be defined: how solutions (decision space) are encoded in the chromosome, and the fitness function (objective space) to be optimized. Both will be described in the next two subsections.

2.2. Solution encoding

In evolutionary algorithms, the decision space (spatial filter S and subset of electrodes) is usually encoded by an array of real numbers, also called the chromosome or the individual.

The first part of the chromosome ($c \times c'$ genes) contains a sequence of real numbers that represent matrix $S_{c\times c'}$, the spatial filter. The second part of the chromosome controls which electrodes are selected. A possibility would be to use a binary mask with as many bits as electrodes, but this could result in very long chromosomes and it would be necessary to combine real and binary parts within the same chromosome. In this paper, we will take advantage that the components of the spatial filter can be interpreted as the strength of electrodes in the filtered channels. According to Eq. 2, if the ith row of S contains small values, then the ith electrode will have small influence on the filtered channels $\mathbf{m'_j}$. The rule displayed in Eq. 5 shows how threshold s_t can be used to remove those electrodes i with weights less than threshold. Thus, s_t will be encoded in the last part of the chromosome as a single gen. For each electrode, there is a row in S. If the maximum value of a row is less than the threshold s_t then the participation of real electrode i in all filtered channels j is small. This electrode is removed by zeroing row i of S. The larger the threshold, the more electrodes are removed, but the less significant ones are removed first.

$$\text{if } \max_k(s_{ik}) < s_t \text{ then } (s_{i1}, \ldots, s_{ic'}) \leftarrow (0, \ldots, 0) \tag{5}$$

In summary, the chromosome encodes particular combination of values from the decision space $((s_{i,k})_{c\times c'}, s_t) = (S, s_t)$. Those individuals must be evaluated according to a multi-objective fitness function, which is described in the next Section.

2.3. The multi-objective fitness function

Each individual (S, s_t) in the population encodes the spatial filter to be applied to the raw data and a threshold that determines the number of channels. In order to evaluate the quality of the filter and the number of channels involved from a multi-objective approach, two objectives have been considered:

- The accuracy of the linear classifier that uses as inputs the spatially processed signals based on the filter encoded in the chromosome.
- The actual number of electrodes used

The computation of the second objective is straightforward because it is just the number of rows of matrix S where the rule in Eq. 5 does not apply (i.e. electrodes with entries in S smaller than s_t are removed). However, the computation of the first objective of fitness function is a more complex process and a detailed description can be found in [1]. Here, the main steps involve in the computation of the first objective are summarized.

The starting point is the raw data, which is made of c time series (as many as real electrodes) that contain the EEG signal recorded from a subject during an acquisition session. The raw signal belongs to the time domain: for every time instant, the signal is recorded for every electrode. In order to evaluate individual (S, s_t) these steps are followed:

1. The raw data is segmented into several chunks $M^{(n)}$, each one of dimension $t \times c$, where t is the number of time instants in the signal chunk. Consecutive $M^{(n)}$'s overlap by δt time instants. This means that if $M^{(1)}$ starts at time 1, $M^{(2)}$ starts at time $1+t-\delta t$, $M^{(3)}$ starts at time $1+2t-\delta t$ and so on. The reason for this segmentation is to generate different data samples $M^{(n)}$ from the raw data. Both t (the signal chunk size) and δt (the amount of overlap) are parameters set by the user.
2. The spatial filter S with some rows set to zero by threshold s_t (rule in Eq. 5) is applied to each $M^{(n)}$ according to Eq. 1. A matrix of features $F^{(n)} = (f_{i,j}^{(n)})_{t'\times c'}$ is created from each $M^{(n)}$.
3. The weights $w_{i,j}^{(p)}$ and biases $b^{(p)}$ of the linear discriminants $D^{(p)}$ of Eq. 4 are learned by means of Fisher's LDA from the data samples $M^{(n)}$. Let's remember that a one-versus-all approach is followed for more than two classes, therefore N linear discriminants are learned from the data.
4. The quality of the linear discriminants is measured as the mean squared error in order to provide a continuous and more precise feedback to the evolution process. It is defined as:

$$\mathbf{MSE} = E_{F^{(n)}}\left[\sum_{p=1}^{p=N}\left(\sigma(D^{(p)}(F^{(n)})) - y^{(p,n)}\right)^2\right] \tag{6}$$

where E is the average operator, N is the number of classes (or mental tasks), $D^{(p)}$ is the linear discriminant that separates class i from the rest of classes, σ is a sigmoid function between 0 and 1 ($\sigma(x) = \frac{1}{(1+e^{-x})}$), and $y^{(p,n)}$ is the desired output for data chunk $M^{(n)}$: $y^{(p,n)} = 1$ if the subject was performing mental task p during $M^{(n)}$ and 0 otherwise.

5. The results returned by the multi-objective fitness function are the **MSE** and the number of real electrodes left by threshold s_t.

3. Experimental Validation

3.1. Data sets description and Setting Parameters

In this paper, BCI-III competition data [1] has been used. They consist on three datasets acquired in the IDIAP Research Institute will be used [19] from three different subjects. Each dataset contains data for 4 non-feedback sessions (named Session1, Session2, Session3, and Session4). 32 electrodes were located on the subjectss scalp. There are 3 mental tasks: Imagination of repetitive self-paced left hand movements; imagination of repetitive self-paced right hand movements; and, generation of words beginning with the same random letter. All 4 sessions of a given subject were acquired on the same day, each lasting 4 minutes with 5-10 minutes breaks in between them. The subject performed a given task for about 15 seconds and then switched randomly to another task at the operators request. In this paper, we use the raw EEG signals provided by the competition organizers.

For each subject and all the experiments carried out in this work, the three first sessions (Session1, Session2 and Session3) were used for training and to guide the searching process, while the last one (Session4) was used for testing the classifiers.

Most of the experimental parameters are given by the provider of the data [19]. Those are the sampling frequency $f = 512$ Hz; the size of temporal windows to construct every training instance is 1s, that is $t = 512$; training instances are sampled 16 times per second (hence the amount of overlap is $\delta t = \frac{512}{16} = 32$), and the number of electrodes is $c = 32$. We have followed the suggestion of [19] and only frequencies from 8 Hz to 30 Hz have been taken into account, because it is in that range that motor imagery patterns can be detected. The only parameter for our approach is c' (the number of columns in the spatial filter). Based on previous work on the same dataset, c' has been set to 2 [1].

For each subject (1,2 and, 3) ten independent runs have been made using the multi-objective algorithm NSGA-II [6]. In this work, the MATLAB NSGA-II

[1] http://www.bbci.de/competition/iii/results/index.html#martigny.

implementation (gamultiobj) has been used. The experiments have been made under the default parameters of the algorithm, except with a population of 100 individuals and gaussian mutation.

3.2. Experimental Results and Comparative Analysis

In this Section, we report the results obtained using the NSGA-II multi-objective evolutionary algorithm and compare them with two other algorithms that follow similar heuristics but use a greedier search than the multi-objective approach. The three methods start from the same spatial filter S_0 obtained by a single-objective optimizer (the Covariance Matrix Adaptation Evolution Strategy or CMA-ES) using the maximum number of electrodes [1]. Therefore, three methods are compared: the multi-objective approach described in Section 2.3, an unsupervised greedy method, and finally a supervised one.

Unsupervised greedy method: The idea is to iteratively remove electrodes with smallest weights in the initial filter S_0 (therefore, it follows a similar heuristic to the rule in Eq. 5). Or equivalently, for each electrode, its maximum weight in S is computed as $m_i = \max_k(s_{ik})$ and then electrodes are sorted according to m_i, small values first. The electrode with smallest m_i is removed. Then, the resulting filter (with fewer electrodes $S_{(c-1)\times c'}$) is applied to the raw EEG signal, a new classifier is trained with the filtered signal, and the error is computed. The procedure is repeated until only one channel remains in the spatial filter. Thus, the error is computed for c electrodes, $c-1$ electrodes, $c-2$ electrodes, ..., down to 2 electrodes. The final result is that, for every number of electrodes (from c to 2), the classification error that can be achieved with those electrodes is obtained. It should be noticed that the initial spatial filter S_0 is not modified, except for removing electrodes (removing the ith electrode is equivalent to zeroing the ith row of S_0).

Supervised greedy method: It is similar to the unsupervised one but it is the training error (or success classification rate) that is used to decide the electrodes to be removed at every step. If for instance there are n electrodes left, all possible classifiers with $n-1$ electrodes are tested, and the electrode producing the minimum classification error decrement will be removed. Let's illustrate this with an example: if there are 4 electrodes left (x_1, x_2, x_3, and x_4), four different classifiers, C_1 to C_4, are built. Classifier C_i uses all electrodes except x_i. Then, the classification error is computed for each C_i and the electrode corresponding to the classifier with smallest error is removed. In other words, the electrode less relevant for classification is removed. As in the unsupervised case, the procedure is repeated until all channels are removed, one by one. The final result is a set of solutions where, for every possible number of electrodes, the corresponding classification error has been computed. Unlike the unsupervised method, the supervised one removes first the less important electrodes for classification by actually computing the classification error instead of using the heuristic that considers the weights in the spatial filter. In principle, this should give better results, but it is slower than the unsupervised approach, because computing the classification error is costly and this has to be done many times. Like in the unsupervised approach, the initial spatial filter S_0 is never modified, except for removing electrodes (removing the ith electrode is equivalent to zeroing the ith row of S_0).

Multi-objective approach: it is the method based on NSGA-II and described in Section 2.3, but the initial population contains the filter S_0 obtained from CMA-ES with all c electrodes. The rest of individuals in the population is random, as in standard NSGA-II. This approach should work better than the greedy ones if the spatial filter has to be re-adjusted when electrodes are removed. The greedy methods can only decide the order in which electrodes are removed, but filter S_0 is left untouched (except for the rows set to zero for removing each electrode). In some sense, the multi-objective approach combines the unsupervised and supervised ideas, because it uses a threshold (encoded in the chromosome) to remove electrodes but also uses the classification error to guide the search. But given that the spatial filter is encoded in the chromosome (i.e. is part of the decision space), it also permits to adapt the spatial filter for different sets of electrodes.

The three methods return a set of solutions $\{(n, e_n)\}_{n=1}^{n=c}$ where n is the number of electrodes and e_n is the error with that number of electrodes. In the case of the NSGA-II approach, this set of solutions is the Pareto front. In order to account for the stochastic nature of evolutionary algorithms, CMA-ES has been run 10 times, 10 initial filters $(S_0^{(r)})_{r=1}^{r=10}$ have been obtained, and then each of the three methods have been applied to these initial filters. Therefore, 10 sets of solutions have been obtained for every method.

To visualize and analyze the ten sets of solutions (for each subject), an average front has been generated by ranking the set of non-dominated solutions by the second objective (number of electrodes) and then computing the average of the ten errors for every number of electrodes. This has been done for training and test. When the multi-objective approach is used, there is no guarantee that there is a solution for every number of electrodes (i.e. a solution with 4 electrodes might not exist in the Pareto front). In that case, it is assumed that the error with n electrodes e_n is equal to the error obtained with $n-1$ electrodes, this is $e_n = e_{n-1}$. This is reasonable given that it is always possible to construct a solution with n electrodes from a solution with $n-1$ electrodes by just ignoring the additional electrode.

The differences among the average errors for the three suggested approaches (unsupervised, supervised and multi-objective) have been tested for statistical significance. This statistical study has been applied to pairs of methods: unsupervised versus supervised, multi-objective versus unsupervised, and multi-objective versus supervised. The protocol used in this study has been the following: first, it is checked if results of the ten runs follow a Normal distribution applying the Shapiro test and if the variances of the two pairs of methods are homogeneous using the Levene test (homoscedasticity). Then, if both conditions are verified, the t-test is performed to compare the average of errors for the two methods. In other cases, the Wilcoxon test (non-parametric) is applied to compare the solutions. A significance level $\alpha = 0.05$ is used in all cases.

In Figures 1, 2 and 3 the average training and test errors are shown for

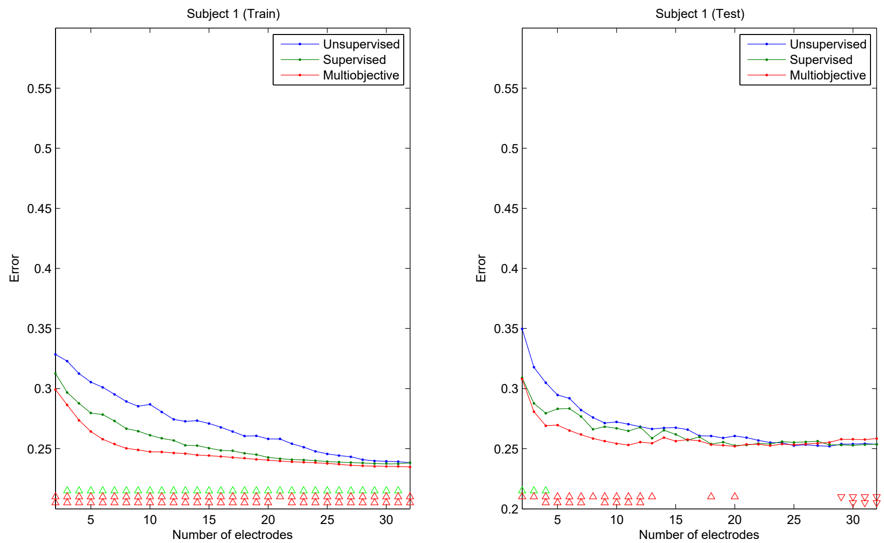

Figure 1: Unsupervised, supervised, and multi-objective average fronts for Subject 1.

subjects 1, 2, and 3, respectively. The figures include the results obtained with the multi-objective approach as well as the results provided by the two strategies used for comparison, unsupervised and supervised methods. The three lines with triangles at the bottom of the figures display the statistical significance comparisons. The upper line compares supervised vs. unsupervised and the next two lines compare multi-objective vs. unsupervised, and multi-objective vs. supervised. The symbol $\triangle$ means that the first algorithm is significantly better than the second one, while ∇ means otherwise. Absence of a triangle implies that the difference is not significant.

Observing the results provided by the multi-objective approach, it can be appreciated that the error for training and test tends to decrease as the number of electrodes increases. However, a similar level of accuracy can be reached with a number of electrodes less than 32. In particular, 90% of the reduction in training error can be achieved with only 19, 20, and 15 electrodes for subjects 1, 2, and 3, respectively. This shows that the number of electrodes to be used to get an appropriate level of accuracy can be reduced, as it was the hypothesis of this study.

The supervised approach sorts the electrodes according to classification error. The unsupervised method was thought as a faster way of sorting the electrodes because they were ordered by using only the weights in the spatial filter instead of computing the classification error at each iteration. Results show that the supervised approach is generally better than the unsupervised one, specially in subject 2. Both approaches leave the initial spatial filter unchanged, they only sort the electrodes according to some criterion. On the other hand, the

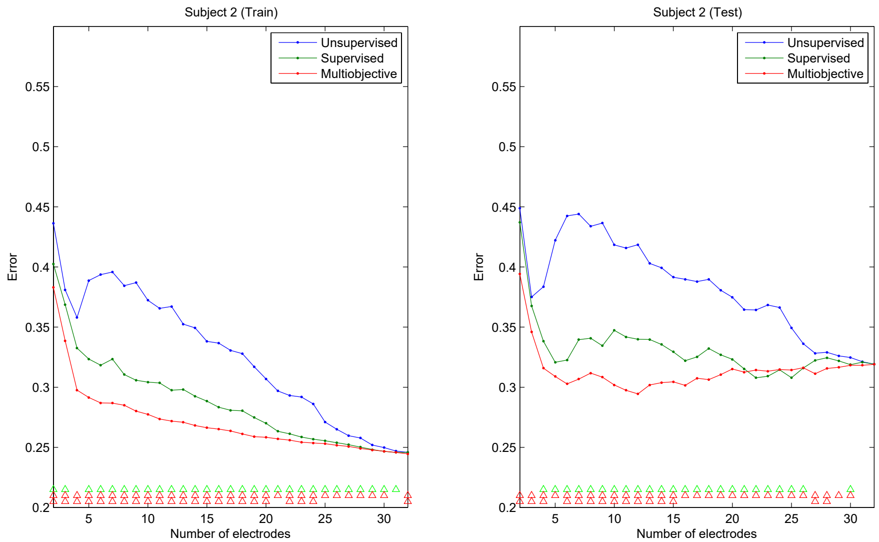

Figure 2: Unsupervised, supervised, and multi-objective average fronts for Subject 2.

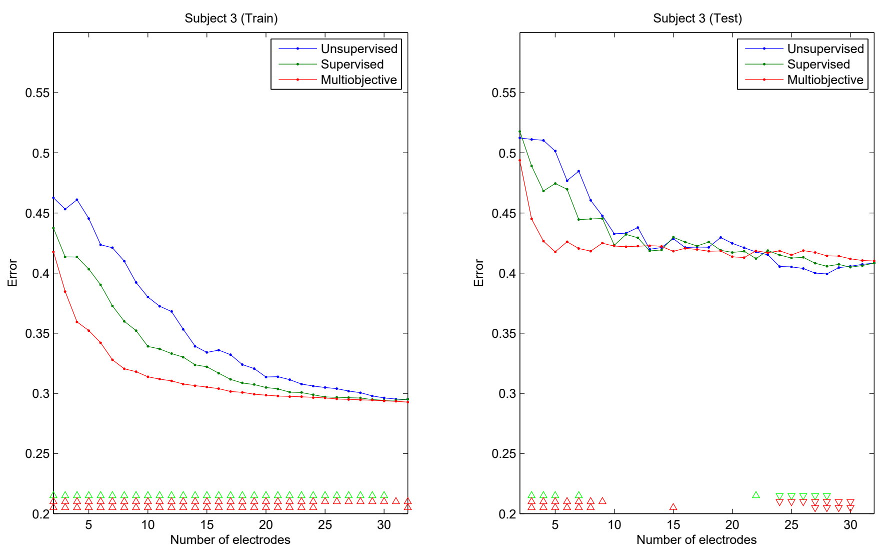

Figure 3: Unsupervised, supervised, and multi-objective average fronts for Subject 3.

multi-objective approach is designed to optimize the initial filter, which allows to improve its performance in classification. As it can be observed in Figures 1, 2, and 3 on the left (training data), the multi-objective approach is always significantly better than the supervised method for subject 1. For subjects 2 and 3, this is also the case, except when the number of electrodes is larger than 25, where there is no significant difference between both approaches.

In test, the situation is slightly different (see Figures 1, 2, and 3 on the right), which can be expected in the BCI domain, because training and test is done with different sessions with data recorded in different days. Let us remember that in this work training is done with session 1, 2, and 3 and testing with session 4. In any case, the multi-objective approach is generally better than the supervised one also in test. The main improvement can be observed on the left hand side of the fronts (small number of electrodes). For subject 1, it is significantly better with fewer than 12 electrodes (except for 2, 3, and 8 where there is no significant difference). For subjects 2 and 3 this is also the case with fewer than 15 and 8 electrodes respectively. For a larger number of channels, differences are not significant, except in a few cases where multi-objective performs worse (30 to 32 for subject 1 and 27 to 30 for subject 3). But even in those cases, the actual differences are very small (less than 0.005 in the worse case for subject 1 and 0.009 for subject 3).

In general terms, the multi-objective framework provides better Pareto fronts, specially in those regions of the front that allows to use fewer electrodes.

4. Conclusions

EEG-based BCI systems require to attach many electrodes on the scalp of subjects. However, the placement of each channel involves a laborious process. The later, coupled to the fact that not all electrodes are equally relevant for EEG classification, motivates the study of methods for selecting electrodes. On the other hand, the preprocessing of EEG signals by means of filters (spatial, spectral, ...) plays a crucial role in the success of EEG classification.

In this work, we combine the optimization of a spatial filter and the optimization of the number of electrodes within an evolutionary multi-objective framework. The approach provides a set of solutions (Pareto front), in which each solution represents the optimal classification error for each number of channels. The evolutionary search not only takes charge of the selection of relevant channels, but also the optimization of the spatial filter. Thus, the minimization of classifier error is also guaranteed.

The multi-objective approach has been compared with two other approaches (named unsupervised and supervised) which do not modify the filter, but start from a filter previously optimized for the whole set of electrodes, and then iteratively remove the less relevant channels. The criteria for electrode removal are related to the weights of channels in the spatial filter (unsupervised), or to classification error (supervised). In order for the comparison to be meaningful, the multi-objective approach also starts from this initially optimized filter, but it is allowed to modify it along the evolutive process.

As a general observation, results show that for all the approaches, reasonable levels of error can be achieved by using significantly fewer electrodes than the whole set, hence confirming the relevance of electrode reduction. With regard to differences between the approaches, results show that the multi-objective approach provides similar solutions than unsupervised and supervised methods in the right part of the fronts (i.e. large number of channels). However, when the number of electrodes is smaller (left part of the fronts), the multi-objective method provides solutions with smaller classification errors. This is due to the multi-objective framework being able to optimize the filter for each different number of electrodes.

Acknowledgments

This work has been funded by the Spanish Ministry of Science under contract TIN2011-28336 (MOVES project).

References

[1] Ricardo Aler, Inés María Galván, and José María Valls. Applying evolution strategies to preprocessing eeg signals for brain-computer interfaces. *Inf. Sci.*, 215:53–66, 2012.

[2] Adham Atyabi, Martin H. Luerssen, Sean P. Fitzgibbon, and David M. W. Powers. Evolutionary feature selection and electrode reduction for eeg classification. In *IEEE Congress on Evolutionary Computation*, pages 1–8, 2012.

[3] Tom Carlson and José del R. Millán. Brain-controlled wheelchairs: A robotic architecture. *IEEE Robot. Automat. Mag.*, 20(1):65–73, 2013.

[4] Luca Citi, Riccardo Poli, Caterina Cinel, and Francisco Sepulveda. Feature selection and classification in brain computer interfaces by a genetic algorithm. In *Late-breaking papers of the Genetic and Evolutionary Computation Conference (GECCO-2004)*, volume 400, 2004.

[5] G.P. Coelho, C.C. Barbante, L. Boccato, R. R F Attux, J.R. Oliveira, and F.J. Von Zuben. Automatic feature selection for bci: An analysis using the davies-bouldin index and extreme learning machines. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–8, 2012.

[6] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[7] G. Dornhege, M. Krauledat, K-R. Muller, and B. Blankertz. *Toward Brain-Computer Interfacing*, chapter General Signal Processing and Machine Learning Tools for BCI Analysis, pages 207–234. MIT Press, 2007.

[8] G. Dornhege, M. Krauledat, K.-R. Müller, and B. Blankertz. *Towards Brain-Computer Interfacing*, chapter General Signal Processing and Machine Learning Tools for BCI Analysis, pages 207–234. MIT Press, 2007.

[9] Josef Faller, Gernot R. Müller-Putz, Dieter Schmalstieg, and Gert Pfurtscheller. An application framework for controlling an avatar in a desktop-based virtual environment via a software ssvep brain-computer interface. *Presence*, 19(1):25–34, 2010.

[10] Mehrdad Fatourechi, Ali Bashashati, Rabab K Ward, and Gary E Birch. A hybrid genetic algorithm approach for improving the performance of the lf-asd brain computer interface. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 5, pages v–345. IEEE, 2005.

[11] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(7):179–188, 1936.

[12] N. Hansen and A. Ostermeier. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. In *Proceedings of the 1996 IEEE International Conference on Evolutionary Computation*, pages 312–317. Citeseer, 1996.

[13] Bashar Awwad Shiekh Hasan, John Q. Gan, and Qingfu Zhang. Multi-objective evolutionary methods for channel selection in brain-computer interfaces: Some preliminary experimental results. In *IEEE Congress on Evolutionary Computation*, pages 1–6, 2010.

[14] Jing Jin, Xingyu Wang, and Jianhua Zhang. Optimal selection of eeg electrodes via dpso algorithm. In *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on*, pages 5095–5099, June.

[15] A. Kubler. *Towards Brain-Computer Interfacing*, chapter An Introduction to Brain-Computer Interfacing, pages 2–25. MIT Press, 2007.

[16] A. Kubler and K-R. Müller. *Toward Brain-Computer Interfacing*, chapter An Introduction to Brain-Computer Interfacing, pages 1–26. MIT Press, 2007.

[17] Robert Leeb, Doron Friedman, Gernot R Müller-Putz, Reinhold Scherer, Mel Slater, and Gert Pfurtscheller. Self-paced (asynchronous) bci control of a wheelchair in virtual environments: a case study with a tetraplegic. *Computational intelligence and neuroscience*, 2007, 2007.

[18] Jun Lv and Meichun Liu. Common spatial pattern and particle swarm optimization for channel selection in bci. In *Innovative Computing Information and Control, 2008. ICICIC '08. 3rd International Conference on*, page 457, june 2008.

[19] J. del R. Millán. On the need for on-line learning in brain-computer interfaces. In *Proceedings of the International Joint Conference on Neural Networks*, Budapest, Hungary, July 2004. IDIAP-RR 03-30.

[20] B.S. Moon, H.C. Lee, Y.H. Lee, J.C. Park, I.S. Oh, and J.W. Lee. Fuzzy systems to process ecg and eeg signals for quantification of the mental workload. *Information Sciences*, 142(14):23 – 35, 2002.

[21] Noura Al Moubayed, Bashar Awwad Shiekh Hasan, Jonh Q. Gan, Andrei Petrovski, and John McCall. Binary-sdmopso and its application in channel selection for brain-computer interfaces. In *Proceedings of the Workshop on Computational Intelligence,*, pages 1–6, 2010.

[22] P L Nunez, R Srinivasan, A F Westdorp, R S Wijesinghe, D M Tucker, R B Silberstein, and P J Cadusch. Eeg coherency. i: Statistics, reference electrode, volume conduction, laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and Clinical Neurophysiology*, 103(5):499–515, 1997.

[23] Bernhard Obermaier, Gernot R. Müller, and Gert Pfurtscheller. 'virtual keyboard' controlled by spontaneous eeg activity. In *ICANN*, pages 636–641, 2001.

[24] David A Van Veldhuizen and Gary B Lamont. Multiobjective evolutionary algorithms: Analyzing the state-of-the-art. *Evolutionary computation*, 8(2):125–147, 2000.

[25] Qingguo Wei and Yanmei Wang. Binary multi-objective particle swarm optimization for channel selection in motor imagery based brain-computer interfaces. In Yongsheng Ding, Yonghong Peng, Riyi Shi, Kuangrong Hao, and Lipo Wang, editors, *BMEI*, pages 667–670. IEEE, 2011.

[26] Shang-Ming Zhou, John Q. Gan, and Francisco Sepulveda. Classifying mental tasks based on features of higher-order statistics from eeg signals in braincomputer interface. *Information Sciences*, 178(6):1629 – 1640, 2008.