Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2016, Article ID 1435321, 12 pages http://dx.doi.org/10.1155/2016/1435321



Research Article

User Adapted Motor-Imaginary Brain-Computer Interface by means of EEG Channel Selection Based on Estimation of Distributed Algorithms

Aitzol Astigarraga, ¹ Andoni Arruti, ² Javier Muguerza, ² Roberto Santana, ¹ Jose I. Martin, ² and Basilio Sierra ¹

¹Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, Computer Science Faculty, 20018 Donostia-San Sebastian, Spain

Correspondence should be addressed to Andoni Arruti; andoni.arruti@ehu.es

Received 4 June 2014; Accepted 19 November 2014

Academic Editor: Yudong Zhang

Copyright © 2016 Aitzol Astigarraga et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Brain-Computer Interfaces (BCIs) have become a research field with interesting applications, and it can be inferred from published papers that different persons activate different parts of the brain to perform the same action. This paper presents a personalized interface design method, for electroencephalogram- (EEG-) based BCIs, based on channel selection. We describe a novel two-step method in which firstly a computationally inexpensive greedy algorithm finds an adequate search range; and, then, an Estimation of Distribution Algorithm (EDA) is applied in the reduced range to obtain the optimal channel subset. The use of the EDA allows us to select the most interacting channels subset, removing the irrelevant and noisy ones, thus selecting the most discriminative subset of channels for each user improving accuracy. The method is tested on the IIIa dataset from the BCI competition III. Experimental results show that the resulting channel subset is consistent with motor-imaginary-related neurophysiological principles and, on the other hand, optimizes performance reducing the number of channels.

1. Introduction

Recent advances in Cognitive Neuroscience and Brain Imaging technologies provide us with the ability to interact directly with the human brain, offering thus an alternative to natural communication and control. In recent years, many Brain-Computer Interfaces (BCIs) have been built [1].

The aim of a BCI system is to establish a communication method that translates human intentions, mental tasks reflected by suitable brain signals (e.g., electric, chemical, and blood flow changes), into a control signal for an output device such as a computer application or a neuroprosthesis, not requiring any muscular response. The idea is to provide a new communication method to people who are paralysed but are cognitively intact. Motor-Imaginary BCI systems are based on the fact that imagination of movement changes brain

activity in the cortex. Therefore, the recognition of patterns associated with certain movements could be used to generate control signals.

Two general BCI systems can be found in literature: invasive technologies, in which sensors are implanted directly in the brain, and noninvasive technologies, which measure brain activity using external sensors. In most BCI applications, brain signals are measured by electroencephalography (EEG) sensors, recording electrical activity from the scalp with electrodes. They are inexpensive, portable and their temporal resolution is very good.

However, there are several challenges. For example, when developing new BCI systems, the selection of specific channels to be used for each subject is a crucial aspect. However, finding optimal channel number and their positions is still a challenging task [2]. In such cases, the most common way

²Department of Computer Architecture and Technology, University of the Basque Country UPV/EHU, Computer Science Faculty, 20018 Donostia-San Sebastian, Spain

is to use a large number of channels for signal classification. But applying a large number of channels is time consuming and may include noisy and redundant signals that degrade the BCI system performance.

Therefore, selecting the least number of channels that yield the best or required accuracy is crucial. Channel selection can be performed manually [3], based on neurophysiological knowledge, but this approach does not always guarantee optimal results [4]. Moreover, automated channel selection algorithms can provide optimal channel positions without any prior knowledge about the task at hand. Two types of automatic channel selection approaches can be found in literature: subject-independent methods, in which the electrode subset is shared between all subjects, and subjectdependent methods, in which the channel subset is customized for each subject in order to improve the individual BCI performances. This paper focuses on subject-dependent channel selection, and its main goal is the automatic selection of a reduced number of channels (and consequently their location) adapted to each subject, maintaining, or even improving, the classification accuracy in EEG-based BCI.

Our proposed method decomposes the channel selection problem into two steps: estimation of a suitable channel range by means of a backward greedy algorithm and optimal channel selection applying an evolutionary algorithm. In the first step, we propose to use a computationally inexpensive greedy algorithm to estimate a suitable channel range—this is made in a similar way as in [5]. In the second step, the optimal channels are selected applying an evolutionary algorithm. In particular, we investigate the adequateness of an Estimation of Distribution Algorithm (EDA) [6], an evolutionary algorithm based on probabilistic modelling of the search space, to exploit the channel's relationship and pick out the subset that yields the best performance. The ability of Estimation of Distribution Algorithms to maintain the quality of the joint probability distribution among the selected features is remarked by Yin and Wu [7]. In order to evaluate the performance of the proposed approach, public data available from BCI competition has been used. The main goal of this paper is to adapt BCI-EEG interfaces to each individual by selecting those channels which better discriminate among the different actions to be performed.

The paper is structured as follows: Section 2 discusses the related work. In Section 3 the two-step channel selection method (greedy algorithm + EDA) is described. Section 4 shows the experimental setup, the data acquisition procedure, and the characteristics of the EDA implementation. Experimental results are given in Section 5 and Section 6 presents the conclusions and future work.

2. Related Work

Several channel selection methods have been proposed in the literature [8–10]. An updated overview can be found in [2, 4]. Most relevant methods include common spatial patterns (CSP) [11], Support Vector Machines (SVM) [12], and methods based on the mutual information (MI).

The classical backward greedy strategy has been also applied in many research works [8, 13]. In this method the

subset is built iteratively. It starts with a full set of electrodes and, at each step, the electrode that maximizes the accuracy of the complementary subset is removed.

More closely related to our approach, previous researches have addressed the channel selection problem using evolutionary algorithms [4, 9, 14–17]. However, the application of EDAs has been constrained to the analysis of Magnetoencephalography (MEG) data in the context of multiobjective optimization [18], an approach with important differences with the one introduced in this paper. Nevertheless, we briefly review this related work to emphasize differences and similarities with our work.

Wei et al. [17, 19] apply genetic algorithms (GAs) for the analysis of multichannel electrocorticogram (ECoG) recordings in a ECoG-based BCI. They combine the application of the CSP method for feature extraction, Fisher discriminant analysis as classification method, with the use of a GA for feature selection. The authors acknowledge the ability of the GA-based approach to reduce the number of features without losing classification accuracy.

In [9], a GA is combined with a multilayer neural network (ML-NN) to find a subset of channels that maximize the ML-NN's classification accuracy. They use EEG and ECoG recordings. A particular characteristic of this work is that channel selection is not directly accomplished by the GA by optimizing the accuracy. Instead, the fitness function used in the GA is the training error when an MLP is trained for a limited number of epochs. Channels are selected a posteriori by analysing the best solutions of the GA, which includes a nonautomatic part that cannot be analysed.

In [14], a multiobjective evolutionary algorithm (EA) is applied to channel selection. The idea of simultaneously optimizing two or more objectives, including the selection of an optimal set of channels, is very promising for classification of brain recordings. However, a difficulty is that the multiobjective search can be much more complex than single-objective optimization, and this difficulty rapidly increases with the number of objectives involved. Also in [18], multiobjective EAs are applied for channel selection in MEG recordings. In this case, each objective corresponds to maximizing the classification accuracy for a different individual. A regularized logistic regression method is used as a classifier, and a posteriori analysis of the best solutions obtained serves to estimate the relevance of each channel. There is not a sound experiment concerning the real contribution of the evolutionary algorithm.

Kee et al. [4] propose the combination of a Bayesian linear discriminant analysis classifier with a GA for automatic channel selection for a P300 BCI. The authors recognize that a "GA exhibits great potential to study the correlation and the joint effect of various channel combination as the fitness value is not biased to the performance of individual channel." However, the limitations of classical EAs like GAs to capture and respect the dependencies between interacting variables are not discussed. EDAs are conceived to automatically identify and represent these types of dependencies.

More recently, Bhattacharyya et al. [20] suggest the use of a combination of a Learning Automata (LA) and Differential Evolution (DE) [21] for feature extraction in the analysis of EEG data in BCI experiments. Here, the LA-DA is not used for channel selection since features are extracted from only two channels (C3 and C4). Instead, it is applied to select the features that are passed to the classifier. The selection of those two channels is based on the authors experience, but the generalization of this approach is not clear.

There are other evolutionary computation algorithms to have in mind [22]. Nevertheless, the benefits of using EDAs that model the interactions between the variables over evolutionary algorithms that do not take correlations into account have been documented in previous work [6, 23].

To summarize, previous works have proved that channel selection by means of evolutionary algorithms can improve classification accuracy across a variety of brain signals, that is, EEG, ECoG, and MEG. However, these research works have been focused on the application of GAs and other standard EAs that do not explicitly capture and represent the dependencies between the channels. As recognized by some of the authors, the potential of the wrapper approach [24] is its capacity for considering the synergies between the features during the classification process. But in order to exploit these synergies, the EAs should be able to identify the most important interacting features and use this information at the time of generating new solutions. This is exactly what EDAs do.

While filter-based methods could in principle be applied for channel selection, these methods cannot cope with intricate and higher order dependencies between the channels, as our wrapper-type EDA approach does. In addition, the analysis of the probabilistic models generated by the EDA can provide information about what the relevant interactions for the problem are. However, in this paper, we use probabilistic models only to make a more efficient generation of solutions and leave the analysis of the potential information captured by the model as a topic of future research.

3. Proposed Approach

Having to personalize BCI-EEG interfaces to each user as main goal, this paper proposes a novel two-step method for personalized optimal channel selection. Firstly, the system selects an adequate range of number of channels to be explored in the second step. This selection is done by applying a simple greedy algorithm. Thereafter, a more computationally expensive EDA is applied to find the final channel selection. The main reason to choose EDA for the second step is its ability to capture the relevant relationships between the variables of the optimization problem, and, thus, the selection of the channels will be more accurate, reducing the complexity of the BCI system.

Both the greedy algorithm and the EDA will work with a population of candidate solutions to the problem. A solution $\mathbf{x}=(x_1,\ldots,x_{60})$ will be defined as a 60-tuple of binary 0, 1 values—the so-called Binary Encoding. Each position in the tuple refers to a concrete EEG channel, and the value indicates whether this channel is used (1 value) or not (0 value). Furthermore, the two algorithms require the use of a fitness function to evaluate the quality of the explored solutions. In this work we used the common spatial patterns (CSP)

method for feature extraction from the EEG raw signals, and the accuracy of a Support Vector Machine classifier as fitness function, because of its good performance in comparison with other paradigms [25] (see Section 4.2 for more details).

In the next two subsections we explain the main characteristics of the proposed greedy algorithm and the EDA.

3.1. Greedy Search Approach. For the first step we designed a simple greedy search algorithm to reduce the search space (see Algorithm 1). The aim is to ease the search to the EDA making a rough estimation of the adequate number of channels.

In this method, we start with a solution \mathbf{x}_{60} that uses the full set of electrodes (the 60 elements set to one). At each iteration l, a population D_l of solutions is generated by removing one electrode of the best solution selected in the previous iteration (60 – l + 1 possibilities). The solution that gives the best score when evaluated with the fitness function is chosen as the best distribution \mathbf{x}_{60-l} for that number of electrodes.

Simplicity of the greedy algorithm makes a good choice for a rough selection of the range of channels to be used in the more refined search.

3.2. Estimation of Distribution Algorithms. Estimations of distribution algorithms have successfully been developed for combinatorial optimization [6]. They combine statistical learning with population-based search in order to automatically identify and exploit certain structural properties of optimization problems.

EDAs typically work with a population of candidate solutions to the problem, starting with the population generated according to the uniform distribution over all admissible solutions.

The population is then scored using a fitness function. This fitness function gives a numerical ranking for each candidate, with the higher the number the better the solution. From this ranked population, a subset of the most promising solutions are selected by the selection operator. An example selection operator is truncation selection with threshold $\tau =$ 50%, which selects the 50% best solutions. The algorithm then constructs a probabilistic model which attempts to estimate the probability distribution of the selected solutions. Once the model is constructed, new solutions are generated by sampling the distribution encoded by this model. These new solutions are then incorporated back into the old population, possibly replacing it entirely. The process is repeated until some termination criteria are met (usually when a solution of sufficient quality is reached or when the number of iterations reaches some threshold), with each iteration of this procedure usually referred to as one generation of the EDA.

4. Experimental Setup

The performed experiment is explained in this section. First of all, we present the used dataset and how it has been obtained, then details about how the fitness function is

```
(1) \mathbf{x}_{60} \leftarrow solution with all the components (60) set to one.

(2) l=1

(3) \mathbf{do} {

(4) D_l \leftarrow Generate 60-l+1 different solutions removing one channel from \mathbf{x}_{60-l+1}

(5) Compute the fitness function for all the solutions in D_l

(6) \mathbf{x}_{60-l} \leftarrow Select the solution with maximum fitness function value

(7) l=l+1

(8) } \mathbf{until} \ l=60
```

ALGORITHM 1: Population-based greedy search.

calculated are given, and, finally, the EDA implementation is commented.

- 4.1. Data Acquisition. In this work we have selected the IIIa dataset from the BCI competition III [26], because it is publicly available and has been widely used for benchmark evaluation. It contains data from 3 subjects: K3b, K6b, and L1b, collected as follows [27]:
 - (1) Each subject, sitting in front of a computer, was asked to perform imaginary movements of the left hand, right hand, tongue, or foot during a specified time interval according to a cue. The order of cues was random.
 - (2) 60 electrodes were placed on the subject's scalp (see Figure 1(a)) recording a signal sampled at 250 Hz and filtered between 1 and 50 Hz using a Notch filter.
 - (3) As shown in Figure 1(b), each trial started with a blank screen. At t = 2 s, a beep was generated and a cross "+" was shown to attract the subject's attention. At t = 3 s an arrow pointing to the left, right, up, or down was shown for 1 s and the subject was asked to try one of four imaginary movements until the cross disappeared at t = 7 s. This was followed by a 2 s break, and then the next trial began.

The dataset contains 360 instances (cases) for subject K3b, 240 for K6b, and 240 for L1b. Each instance was labelled as belonging to one of the four classes. Each subject contains a balanced distribution of the classes. Two data files are available for each subject: training and testing. The number of instances in the training and testing datasets was equal for all subjects and was 180 for K3b, 120 for K6b, and 120 for L1b. The distribution of the classes was equal in both training and testing data.

The three subjects had different amounts of experience in BCI training. K3b was the most experienced, L1b had less experience, and K6b was a beginner. This has a great influence in classification results (K3b presents the highest accuracy and K6b the lowest) as can be seen in the work of AlZoubi et al. [25] or in [28].

In this work, being our goal the study of the channel selection problem, we will maintain the methodology used in one of these studies [25] as a reference.

- 4.2. Fitness Function. As said before, both algorithms (greedy and EDA) need to evaluate each individual in the population of possible solutions. For this, a classifier using the corresponding subset of channels is built and its mean accuracy is used as a fitness function. For comparison reasons, we have applied the same methodology used by AlZoubi et al. in [25]. These are the steps followed to compute the fitness function for a given channel selection:
 - (i) Feature Extraction. Firstly, the CSP method is applied to the channel-reduced raw EEG data. For each class, applying CSP for that class versus the others, a reduced set of 5 projection signals is obtained. Then 3 frequency band filters (for 8–12 Hz, 21-20 Hz, and 20–30 Hz) are applied, and, finally, 7 features are extracted: max, min, and mean voltage values, voltage range, number of samples above zero volts, zero voltage crossing rate, and average signal power. This process gives 420 (4 classes × 5 projections × 3 filters × 7 features) numeric features for each case of the dataset.
 - (ii) Training and Cross-Validation. In [25] twelve classification paradigms are compared, and for this research we have chosen Support Vector Machines (SVM) [29, 30] because their results are among the best. The value of the fitness function is the mean accuracy obtained with 10-fold cross-validation applied to the training dataset. This has been calculated by the data mining software package WEKA [31] (default parameters are used).
- 4.3. Characteristics of the EDA Implementation. One distinguished feature of EDAs is the type of probabilistic model they use. They should be able to capture the relevant relationships between the variables of the optimization problem. These relationships are encoded in terms of statistical dependencies. Complex models are able to represent higher order dependencies between the variables but they are also more computationally costly due to the requirements of the learning and sampling procedures. As the probabilistic model of choice for the EDA, we use a tree, a model that exhibits a fine balance between the power of representation and the computational cost.

Let $\mathbf{X} = (X_1, \dots, X_n)$ denote a vector of discrete random variables. We will use $\mathbf{x} = (x_1, \dots, x_n)$ to denote

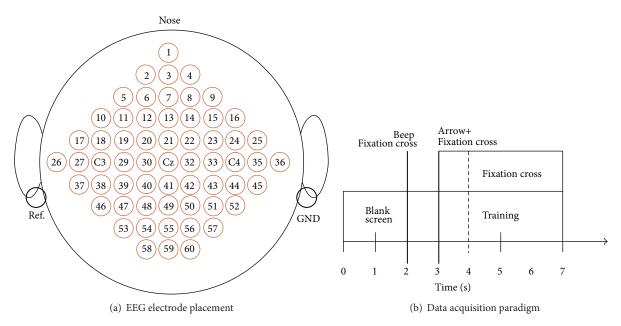


FIGURE 1: Data acquisition schemata (dataset IIIa, BCI competition III).

an assignment to the variables. A probability distribution $p_{\mathcal{T}}(\mathbf{x})$ that is conformal with a tree is defined as $p_{\mathcal{T}}(\mathbf{x}) = \prod_{i=1}^n p(x_i \mid \mathrm{pa}(x_i))$, where $\mathrm{pa}(x_i)$ is the parent of x_i in the tree, and $p(x_i \mid \mathrm{pa}(x_i)) = p(x_i)$ when $\mathrm{pa}(x_i) = \emptyset$; that is, x_i is the root of the tree. The distribution $p_{\mathcal{T}}(\mathbf{x})$ itself will be called a tree model when no confusion is possible.

Trees are able to represent bivariate marginal distributions and they can be learned from data with a computational complexity of n^2 . In order to learn the structure from a dataset, the univariate and bivariate probabilities are, respectively, calculated for every variable and pair of variables. Using the marginal probabilities, the mutual information between each pair of variables is computed. The tree structure itself is determined using the Chow-Liu algorithm [32]. It works by computing the maximum weight spanning tree from the matrix of mutual information between pairs of variables. We set a threshold on the minimal mutual information value required to connect two variables. This allows for representing disconnected trees, that is, a forest.

Algorithm 2 shows the different steps of Tree-EDA. The algorithm starts by generating the random population of solutions D_0 . In each iteration l, a set of solutions D_{l-1}^s is selected from population D_{l-1} and from this population the probabilistic model is learned.

The Tree-EDA shown in Algorithm 2 incorporates the method for learning the tree structure. To sample new solutions from a tree, probabilistic logic sampling [33] is applied to sample. In this method, the roots of each tree are initially sampled according to their univariate probabilities. The rest of the variables are sampled following the order determined by the trees structures and using the conditional probability distributions. As a selection method, Tree-EDA uses truncation selection in which the T=50% best percentages of the population (highest objective values) are selected. The new sampled solutions are combined with the set of

best solutions (elitist solutions) selected from the previous iteration.

EDAs that used trees were originally introduced in [34] and those that employ forests were introduced in [35]. The scheme of Algorithm 2 corresponds to the algorithm introduced in [36]. It has been implemented in MATLAB using the MATEDA software [37], a highly modular implementation in which each EDA component (either added by the user or already included in the package) is implemented as an independent program.

In this paper we use as stop criterion a fixed number of generations. The Tree-EDA parameters used in our experiments are presented in Section 5.1.

4.4. Limiting the Number of Channels of EDA Population. EDA is an evolutionary algorithm that can change the number of channels in any direction while trying to optimize the fitness function. In this work, being our goal to reduce the number of channels, we will need to fix an upper limit for number of channels at each experiment. We will achieve this with two actions:

- (i) setting a unitation constraint when generating the first population: we will generate solutions uniformly distributed but with a fix number of channels;
- (ii) repairing the sampled population at each generation: solutions with a number of channels greater than limit are truncated randomly before evaluating them.

5. Experimental Results

In this section we analyse the overall performance of the proposed approach in the channel selection problem. We evaluate the behaviour of the system in terms of the accuracy

```
(1) D_0 \leftarrow \text{Generate } M \text{ solutions randomly}
    l = 1
(2)
(3)
     do {
(4)
           D_{l-1}^s \leftarrow \text{Select } N \leq M \text{ solutions from } D_{l-1} \text{ according to a selection method}
(5)
           Compute the univariate and bivariate marginal frequencies p_i^s(x_i \mid D_{l-1}^s) and p_{i,j}^s(x_i, x_i \mid D_{l-1}^s) of D_{l-1}^s
(6)
           Calculate the matrix of mutual information using bivariate and univariate marginals.
(7)
           Calculate the maximum weight spanning tree from the matrix of mutual information.
(8)
           Compute the parameters of the model.
(9)
           D_1 \leftarrow Sample M solutions (the new population) from the tree and add elitist solutions.
(10)
           l = l + 1
(11) } until A stop criterion is met
```

ALGORITHM 2: Tree-EDA.

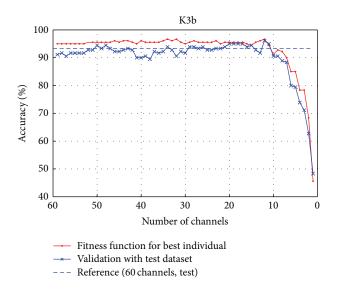


FIGURE 2: Results obtained with greedy search algorithm for subject K3b.

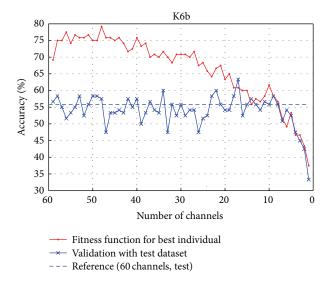


FIGURE 3: Results obtained with greedy search algorithm for subject K6b.

obtained using a standard classification paradigm. Furthermore, we compare the results of the complete system with those obtained by the greedy algorithm.

Obtained results applying the greedy algorithm to the three subjects are shown in Figures 2, 3, and 4. The figures represent the fitness function for the best solution at each iteration, the validation with testing dataset also for the best solution at each iteration; the results obtained with the testing dataset when all channels are used for classification are also shown as baseline reference.

As it can be seen, for the three subjects, the results obtained with 60 electrodes are similar to those obtained after the channel number has been decreased, and the obtained accuracy only decreases when less than 10 electrodes are used.

It can be seen, as said before, that the performance of the classifier depends strongly on the subject, very good for K3b, and poor for the untrained K6b and L1b. Surprisingly, this simple search algorithm shows clearly that, for the three subjects, the number of channels can be reduced drastically without losing accuracy.

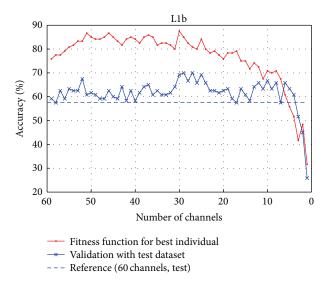


FIGURE 4: Results obtained with greedy search algorithm for subject L1b.

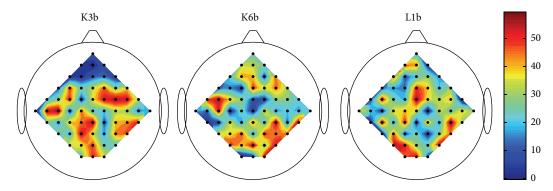


FIGURE 5: Channels position and frequency using greedy search algorithm.

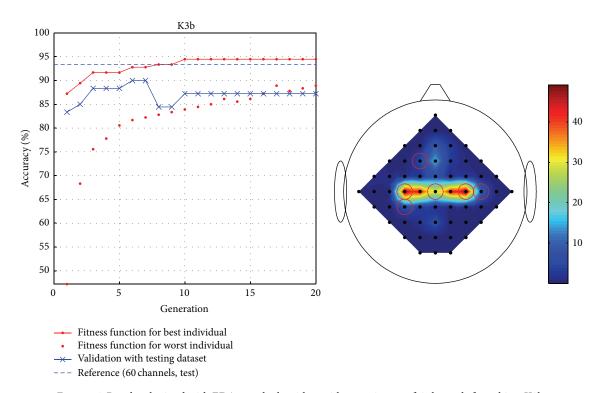


FIGURE 6: Results obtained with EDA search algorithm with a maximum of 6 channels for subject K3b.

In Figure 5 a topographical map for each subject shows the frequency of occurrence of the channels in the 60 best solutions (one for each iteration). The low values correspond to first discarded channels and the highest to the most significant for the classification task. It can be seen that the distribution is subject-dependant, thus justifying the development of an automatic method for searching the optimal subset of channels, making the adaptation of the system to each subject automatic.

5.1. EDA Approach. Seeing the results obtained with the greedy algorithm, we decided to carry out a sequence of experiments with the EDA, changing the top channel number limit in the interval from 16 to 1, wider than the minimum range (from 10 to 1) observed with the greedy algorithm. This

focuses on the work of the EDA to a range where doing a more refined search could be interesting.

With respect to EDA parameters, and based on preliminary experiments, the population size is 200 and the number of generations 20.

To show a sample of the evolution of the EDA, Figures 6, 7, and 8 show intermediate results for the 3 subjects using 6 channels. The figures show the evolution of fitness function. A validation value obtained for the best solution of each generation, with the testing dataset, is also shown, but these values are not used by EDA. A topographic map shows the presence of channels in 50 solutions (the best ten in the last 5 generations). The results show that, for the subject L1b, with only 6 channels the EDA is able to find a solution even better than using all the channels. For subjects K3b and K6b the found solutions are near to the 60-channel based solution.

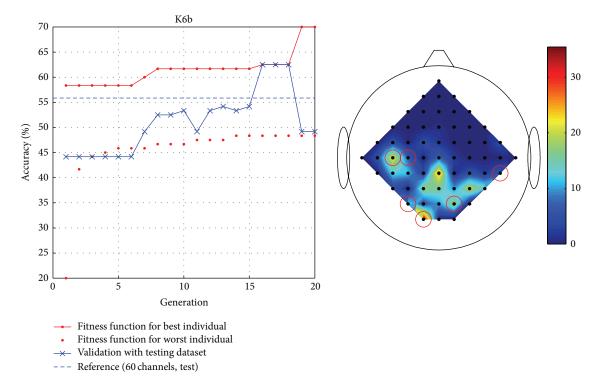


FIGURE 7: Results obtained with EDA search algorithm with a maximum of 6 channels for subject K6b.

Figures 9, 10, and 11 show results obtained with EDA in the last generation with different number of channels for the 3 subjects. Values obtained with the greedy search algorithm are also shown. As it can be seen, the EDA finds a better solution for the subjects K6b and L1b than the solution based on all the channels (with only 4 channels) and finds a similar solution with 10 channels for the K3b subject.

In Figure 12 channel position and frequency as a percentage, in best solutions (the best ten in the last 5 generations of 13 experiments), are shown for the tree subjects. This figure shows again the different distribution of the selected channels for the three subjects, making a specific selection of channels for each subject necessary.

Finally, the results obtained with the EDA have been compared with the results obtained with the greedy algorithm. We applied a Wilcoxon signed ranks test (0.05 confidence level) to determine if statistically significant differences are between the results of both algorithms. The test shows that the differences are statistically significant (p value = 0.037).

Therefore, the combination of the two algorithms in a two-step system is a good option to achieve a simpler user adapted interface maintaining or, even, improving the accuracy of the system.

5.2. Final Discussion. Experimental results show the adequacy of the proposed approach. On the first step, a greedy approach has been used in order to show that channel number reduction is appropriate in a 60-electrode EEG-BCI model. It can be seen in the obtained results that electrode number is decreased to about ten/sixteen channels without losing accuracy for the three subjects, focusing on the search

space in the second step. On the second step, the evolutionary algorithm is used to select a fixed number of electrodes (from 1 to 16). The obtained results of the EDA based approach outperform those achieved by using both the 60 channels and the greedy approach alone as well, with statistically significant differences. In addition, our experiments show that the combination of the two algorithms in a two-step system is a good option to achieve a simpler user adapted interface maintaining or, even, improving the accuracy of the system.

It has to be noticed that, among the different brain zones, selected channels appear to be consistent with MI-related neurophysiological principles [3]. In this sense, the contribution of the performed research can be seen as a step towards a personalized EEG-based BCI interface, in which a person is first trained with a general 60-electrode system, and then the relevant ones are selected to improve the human-machine communication process.

6. Conclusions and Future Works

In this paper a two-step system for channel selection by means of EDA has been presented, aiming at maintaining or even improving the classification accuracy with a few EEG channels automatically selected for each subject. The motivation behind our research work is that electrode signals received by an EEG-BCI interface are not independent among them and that selecting an optimal subset of the electrodes can improve the results when the goal is to identify motor imagery states. In this way, the fine selection

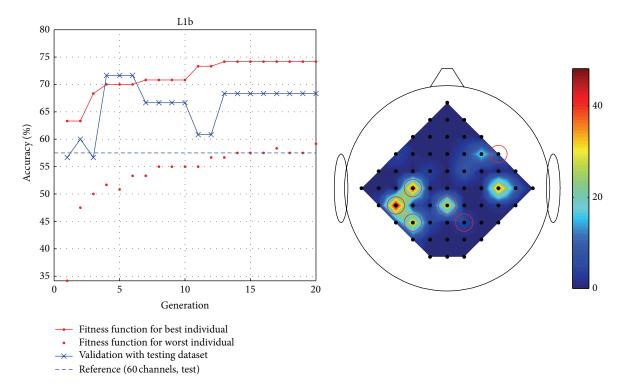


FIGURE 8: Results obtained with EDA search algorithm with a maximum of 6 channels for subject L1b.

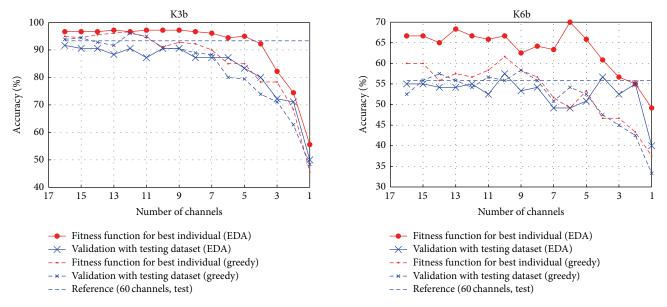


FIGURE 9: Results obtained with EDA search algorithm in the last generation with different number of channels for subject K3b.

FIGURE 10: Results obtained with EDA search algorithm in the last generation with different number of channels for subject K6b.

is performed using an evolutionary computing paradigm—EDA—which looks for, and maintains, a relationship among the selected input channels, thus reducing or suppressing existing redundancies among the channels. The obtained results show that the combination of the greedy and EDA approaches, in a two-step system, is a good option to achieve a simpler and more accurate personalized system.

As future work, along with a comparison with other methods, we are planning to apply the presented approach taking more base classifiers or multiclassifier systems [38] in the channel selection process. Other joint probability models such as Bayesian Networks could be used to extract the relationships of the channels [39]. Fitness-scaling methods will be explored to improve the EDA searching process [40].

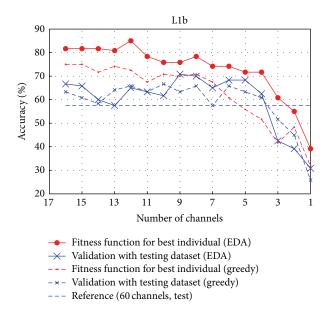


FIGURE 11: Results obtained with EDA search algorithm in the last generation with different number of channels for subject L1b.

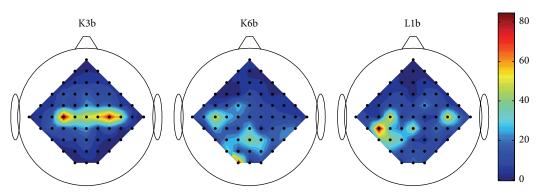


FIGURE 12: Channels position and frequency using EDA search algorithm.

On the other hand, in order to avoid high dimensionality, certain features must be selected before classification starts [23]. A more in-depth analysis of the channel signal is to be done, in order to detect parts of the brain activity that have more influence. A (temporal) clustering among the received signals [41] is also a line for future research which is being considered by the authors. Finally, the inclusion of nonintentional control state patterns—as fifth class—in the experiments could be a first step towards an asynchronous system.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The work described in this paper was partially supported by the Basque Government under Grants IT395-10, IT313-10, and IT609-13, the University of the Basque Country UPV/EHU under Grant UFI11/45, and the Ministry of Science and Technology of Spain under Grant TIN-41272P.

References

- [1] C. Guger, B. Z. Allison, and G. Edlinger, *Brain-Computer Interface Research: A State-of-the-Art Summary 2*, Springer, New York, NY, USA, 2014.
- [2] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1865–1873, 2011.
- [3] M. Thulasidas, C. Guan, and J. Wu, "Robust classification of EEG signal for brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 1, pp. 24–29, 2006.
- [4] C. Y. Kee, R. K. Chetty, B. H. Khoo, and S. Ponnambalam, "Genetic algorithm and bayesian linear discriminant analysis based channel Selection method for P300 BCI," in *Trends in Intelligent Robotics, Automation, and Manufacturing*, vol. 330 of Communications in Computer and Information Science, pp. 226–235, Springer, Berlin, Germany, 2012.

- [5] T. Shibanoki, K. Shima, T. Tsuji, A. Otsuka, and T. Chin, "A quasi-optimal channel selection method for bioelectric signal classification using a partial Kullback-Leibler information measure," *IEEE Transactions on Bio-Medical Engineering*, vol. 60, no. 3, pp. 853–861, 2013.
- [6] P. Larrañaga and J. A. Lozano, Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation, Kluwer Academic Press, 2001.
- [7] P.-Y. Yin and H.-L. Wu, "Cyber-EDA: estimation of distribution algorithms with adaptive memory programming," *Mathematical Problems in Engineering*, vol. 2013, Article ID 132697, 11 pages, 2013.
- [8] A. Barachant and S. Bonnet, "Channel selection procedure using riemannian distance for BCI applications," in *Proceedings* of the 5th International IEEE/EMBS Conference on Neural Engineering (NER '11), pp. 348–351, Cancun, Mexico, May 2011.
- [9] J. Yang, H. Singh, E. L. Hines et al., "Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach," *Artificial Intelligence in Medicine*, vol. 55, no. 2, pp. 117–126, 2012.
- [10] T. N. Lal, M. Schröder, T. Hinterberger et al., "Support vector channel selection in BCI," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1003–1010, 2004.
- [11] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.
- [12] V. N. Vapnik, The Nature of Statistical Learning Theory, Statistics for Engineering and Information Science, Springer, New York, NY, USA, 2nd edition, 2000.
- [13] M. Xu, H. Qi, L. Ma et al., "Channel selection based on phase measurement in P300-based brain-computer interface," PLoS ONE, vol. 8, no. 4, Article ID e60608, 2013.
- [14] B. A. S. Hasan, J. Q. Gan, and Q. Zhang, "Multi-objective evolutionary methods for channel selection in brain-computer interfaces: some preliminary experimental results," in *Proceedings of the IEEE Congress on Evolutionary Computation (CEC '10)*, pp. 1–6, Barcelona, Spain, July 2010.
- [15] J.-Y. Kim, S.-M. Park, K.-E. Ko, and K.-B. Sim, "Optimal EEG channel selection for motor imagery BCI system using BPSO and GA," in *Robot Intelligence Technology and Applications 2012*, vol. 208 of *Advances in Intelligent Systems and Computing*, pp. 231–239, Springer, Berlin, Germany, 2013.
- [16] M. Schröder, M. Bogdan, T. Hinterberger, and N. Birbaumer, "Automated EEG feature selection for brain computer interfaces," in *Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering*, pp. 626–629, IEEE, March 2003
- [17] Q. Wei, Z. Lu, K. Chen, and Y. Ma, "Channel selection for optimizing feature extraction in an electrocorticogram-based brain-computer interface," *Journal of Clinical Neurophysiology*, vol. 27, no. 5, pp. 321–327, 2010.
- [18] R. Santana, C. Bielza, and P. Larrañaga, "Regularized logistic regression and multiobjective variable selection for classifying MEG data," *Biological Cybernetics*, vol. 106, no. 6-7, pp. 389–405, 2012
- [19] Q. Wei and W. Tu, "Channel selection by genetic algorithms for classifying single-trial ECoG during motor imagery," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '08*), pp. 624–627, IEEE, Vancouver, Canada, August 2008.

- [20] S. Bhattacharyya, A. Sengupta, T. Chakraborti, A. Konar, and D. N. Tibarewala, "Automatic feature selection of motor imagery EEG signals using differential evolution and learning automata," *Medical & Biological Engineering & Computing*, vol. 52, no. 2, pp. 131–139, 2014.
- [21] K. Price, R. Storn, and J. Lampinen, Differential Evolution: A Practical Approach to Global Optimization, Natural Computing Series, Springer, Berlin, Germany, 2005.
- [22] X. Zhang, L. Jiao, A. Paul, Y. Yuan, Z. Wei, and Q. Song, "Semisupervised particle swarm optimization for classification," *Mathematical Problems in Engineering*, vol. 2014, Article ID 832135, 11 pages, 2014.
- [23] I. Inza, P. Larrañaga, R. Etxeberria, and B. Sierra, "Feature subset selection by Bayesian network-based optimization," *Artificial Intelligence*, vol. 123, no. 1-2, pp. 157–184, 2000.
- [24] Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection," *Knowledge-Based Systems*, vol. 64, pp. 22–31, 2014.
- [25] O. AlZoubi, I. Koprinska, and R. A. Calvo, "Classification of brain-computer interface data," in *Proceedings of the 7th Australasian Data Mining Conference (AusDM '08)*, J. F. Roddick, J. Li, P. Christen, and P. J. Kennedy, Eds., vol. 87 of *CRPIT*, pp. 123–131, Australian Computer Society, Adelaide, Australia, November 2008, http://dblp.uni-trier.de/db/conf/ausdm/ausdm2008.html #AlZoubiKC08
- [26] B. Blankertz, K.-R. Müller, D. J. Krusienski et al., "The BCI competition III: validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.
- [27] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *Journal of Neural Engineering*, vol. 2, no. 4, pp. 14–22, 2005.
- [28] F. Lee, R. Scherer, R. Leeb, C. Neuper, H. Bischof, and G. Pfurtscheller, "A comparative analysis of multi-class EEG classification for brain-computer interface," in *Proceedings of the 10th Computer Vision Winter Workshop (CVWW '05)*, Zell an der Pram, Austria, February 2005.
- [29] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [30] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, no. 1-2, pp. 169– 186, 2003.
- [31] M. Hall, E. Frank, G. Holmes, B. P. P. Reutemann, and I. Witten, "The weka data mining software: an update," *SIGKDD Exploration Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [32] C. K. Chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.
- [33] M. Henrion, "Propagating uncertainty in bayesian networks by probabilistic logic sampling," in *Proceedings of the 2nd Annual Conference on Uncertainty in Artificial Intelligence*, J. F. Lemmer and L. N. Kanal, Eds., pp. 149–164, Elsevier, 1988.
- [34] S. Baluja and S. Davies, "Using optimal dependency-trees for combinatorial optimization: learning the structure of the search space," in *Proceedings of the 14th International Conference on Machine Learning (ICML '97)*, D. H. Fisher, Ed., pp. 30–38, Morgan Kaufmann, San Francisco, Calif, USA, July 1997.
- [35] M. Pelikan and H. Mühlenbein, "The bivariate marginal distribution algorithm," in *Advances in Soft Computing—Engineering*

- *Design and Manufacturing*, R. Roy, T. Furuhashi, and P. Chawdhry, Eds., pp. 521–535, Springer, London, UK, 1999.
- [36] R. Santana, A. Ochoa, and M. R. Soto, "The mixture of trees factorized distribution algorithm," in *Proceedings of the Genetic* and Evolutionary Computation Conference (GECCO '01), L. Spector, E. Goodman, A. Wu et al., Eds., pp. 543–550, Morgan Kaufmann, San Francisco, Calif, USA, July 2001.
- [37] R. Santana, C. Bielza, P. Larrañaga et al., "Mateda-2.0: a MAT-LAB package for the implementation and analysis of estimation of distribution algorithms," *Journal of Statistical Software*, vol. 35, no. 7, pp. 1–30, 2010, http://www.jstatsoft.org/v35/i07.
- [38] L. Susperregi, A. Arruti, E. Jauregi et al., "Fusing multiple image transformations and a thermal sensor with kinect to improve person detection ability," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 8, pp. 1980–1991, 2013.
- [39] B. Sierra, E. Lazkano, E. Jauregi, and I. Irigoien, "Histogram distance-based Bayesian Network structure learning: a supervised classification specific approach," *Decision Support Systems*, vol. 48, no. 1, pp. 180–190, 2009.
- [40] Y. Zhang, L. Wu, and S. Wang, "UCAV path planning by fitness-scaling adaptive chaotic particle swarm optimization," *Mathematical Problems in Engineering*, vol. 2013, Article ID 705238, 9 pages, 2013.
- [41] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognition*, vol. 46, no. 1, pp. 243–256, 2013.



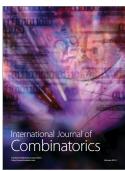








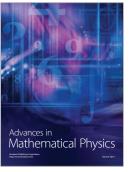






Submit your manuscripts at http://www.hindawi.com











Journal of Discrete Mathematics

