The Impact of PSO based Dimension Reduction in EEG study

Adham Atyabi, Martin Luerssen, Sean Fitzgibbon, and David M. W. Powers

School of Computer Science, Engineering and Mathematics,
Flinders University, SA, Australia
{Adham.Atyabi,Martin.Luerssen,Sean.Fitzgibbon,
David.Powers}@flinders.edu.au

Abstract. High dimensionality nature of EEG data caused by the use of high number of electrodes and long periods of task time is one of the drawbacks in EEG study. Evolutionary based approaches are alternative methodologies to conventional dimension reduction methods with the advantage of not requiring the entire recording sessions for operation. Particle Swarm Optimization (PSO) is an Evolutionary method that achieves performance through evaluation of several generations of possible solutions. This study investigates the feasibility of a 2 layer PSO structure for synchronous reduction of both electrode and task period dimensions using 4 motor imagery EEG data. The results indicate the potential of the proposed PSO paradigm for dimension reduction with non significant lost in classification performance in addition to feasibility to be used in subject transfer applications.

Keywords: Particle Swarm Optimization, Electroencephalogram, Brain Computer Interface

1 Introduction

Electroencephalogram (EEG) is a non-invasive technique for signal acquisition that records variations of surface potential from scalp using some electrodes. The dimensions of the EEG data is measured as a factor of number of used electrode for signal acquisition and the number of extracted feature points during the task period (epoch). The high dimensional nature of EEG data caused by the use of high number of electrodes (up to 256 electrodes) with long task periods (from few seconds to several hours) makes it unsuitable for being used with on-line systems specially with EEG based Brain Computer Interface (BCI). Dimension Reduction (DR) is a preprocessing step that reduces the dimension of the EEG data through the use of some techniques. Conventional decomposition methods such as Principle Component Analysis (PCA), Singular Value Decomposition (SVD), and Common Spatial Pattern (CSP) reduces the dimensions of the data by isolating a set of features or electrodes that comply with some certain criteria regardless of their impact on classification. In addition, these techniques require the entire trials of a recording session of subject for the reduction process.
Evolutionary based approaches are common alternatives for conventional methods due to their capabilities in terms of addressing the short comings of such methods. This study investigates the potential of an Evolutionary based paradigm proposed in [1] to be used with a dataset containing EEG data of 3 healthy subjects performing 4 motor imagery tasks in multiple sessions. The outline of the study is as follows: Section 2 introduces the used Evolutionary based paradigm for synchronous feature and electrode reduction. Section 3 provides details about the used EEG data and the applied preprocessing techniques. Conducted experiments and the achieved results are presented in section 4. Conclusion is made in section 5.

2 Evolutionary based DR

several studies employed evolutionary methods such as Genetic Algorithm (GA) [2–6], Particle Swarm Optimization (PSO) [7–10], and Ant Colony Optimization (ACO) [11] for either electrode or feature reduction in EEG study. The used paradigm in majority of these studies is based on generating populations of possible solutions (subsets of indexes for either feature or electrode dimensions) and smoothly guide the optimization toward including features or electrodes that improve the overall classification performance. To do so, in most cases the EEG data is divided to two subsets of training and testing sets and the subset of indexes that results in the best discrimination of the performed task is chosen as the final solution. Despite the encouraging results achieved by such paradigm it is reasonable to expect data contamination through such paradigm due to the fact that the final product is tuned to perform well on the testing set and it is likely to have low generalizability. Atyabi et al., in [1] and [12], suggested the addition of an extra evaluation step in the paradigm that allows the evaluation of the final product on an unseen set of data that is not being used within the previous selection/evaluation stages. The results indicate the lack of generalizability in the final product. This issue is resolved by introducing three new index sets representing the best performing set on validation set, testing set, and most commonly used indexes. The results indicate the superiority of the set representing most commonly used indexes.

In [1], a new paradigm featuring a 2 layer PSO structure that allows over 90% DR through synchronous reduction of both feature and electrode dimensions is used. Although the study reports encouraging results through the use of most commonly used indexes, the fact that only one dataset containing 2 motor imagery tasks is used for assessment prevents the conclusion about feasibility of such paradigm in terms of being used as DR operator for motor imagery EEG data. The first objective of this study is to further analyze the proposed paradigm using a more complicated dataset featuring 4 motor imagery tasks.

In [13] the proposed paradigm in [1] is used in an inclusion with two frameworks to investigate Subject Specificity and Task Specificity in a subject transfer study. The results indicate the possibility of improving the classification performance through the use of combination of commonly used indexes and a frame-
work that provides Task Specificity. The second objective of this study is to further analyze this issue using a dataset of 4 motor imagery tasks. The proposed PSO paradigm and the used frameworks for Subject and Task Specificity are discussed in following section.

2.1 PSO Paradigm

Particle Swarm Optimization (PSO) is a population-based method that achieves performance through local and global interactions among particles (members of population). PSO uses parameters such as velocity (v), position in the search/solution space (x), acceleration coefficients (c₁ and c₂), and inertia weight (w) in its formulation and has limited memory containing the best found solution with each particle (member of the population) and the swarm denoted as Pbest and Gbest respectively. PSO achieves performance through updating the solutions in the population with respect to local and global found solutions iteratively using Eqs. 2 and 1 [14].

\[
\begin{align*}
V_{i,j}(t) &= w \times V_{i,j}(t-1) + C_{i,j} + S_{i,j} \\
C_{i,j} &= c_1 r_{1,j} \times (p_{i,j}(t-1) - x_{i,j}(t-1)) \\
S_{i,j} &= c_2 r_{2,j} \times (g_{i,j}(t-1) - x_{i,j}(t-1)) \\
x_{i,j}(t) &= x_{i,j}(t-1) + V_{i,j}(t)
\end{align*}
\]

Algorithm 1: PSO based Feature & Electrode reduction

**Initialization:** creates two instance of the population (P₁ and Gbest). Each sub-swarm Pᵢ represents a mask containing i) set of \( n \times k \) out of \( N \times K \) possible features (in PSO notation, this can be considered as \( x_{i,j} \) for \( i \in [1...n] \) and \( j \in [1...k] \)), ii) a set of \( n \) out of \( N \) electrodes, iii) best achieved mask denoted as \( P_{best} \), iv) a Velocity vector denoted as \( v_i \) for \( i \in [1...n] \).

**Gbest** represents a mask containing i) set of \( n \times k \) out of \( N \times K \) possible features, ii) a set of \( n \) out of \( N \) electrodes, and iii) the best achieved mask denoted as \( P_{best} \).

**Evaluation**

**repeat**

**Update the population:** Update the population using algorithm 2.

**Evaluation:** Evaluate all members of the population (sub-swarms) using a classifier.

**Update Bests:** Update Personal (Pbest) and Global (Gbest) Best.

**until** (Termination: the maximum iteration is achieved or the best member of the population (Global Best) has reached to the desired optimum)

**Final Evaluation:** Reevaluate the best mask of the swarm.
Algorithm 2 Pseudo-code for Updating the Population

Find the top 10 candidates: Sort particles based on their classification performance and only preserve the top 10.

for each particle $P_i$ do
  1) Generate a child particle with new set of electrodes that are positioned in nearby areas.
  2) Generate a child particle with new set of electrodes that are positioned in the same areas.
  3) Generate a child particle with new set of $n \times k$ features using velocity vector $v$ and position matrix $x$ and update equations 1 and 2.
  4) Generate a child particle with new set of randomly chosen electrodes that are positioned in the same areas and a new set of $n \times k$ features using velocity vector $v$ and position matrix $x$ and update equations 1 and 2.
end for

In [1] and [13], the proposed paradigm is used in a $10 \times 20$ cross validation (CV) resulting three sets of training, validation and testing with the ratio of 0.9, 0.05, and 0.05 among which the training and the validation are used for the production of final solution in each fold and the testing set is used within the final evaluation step to assess the generalizability of the suggested masks.

2.2 Subject and Task Specificity

The proposed PSO based DR paradigm in [1] is used to investigate its feasibility for subject transfer through two frameworks (Frame work 1 and 2) representing Subject Specificity and Task Specificity in [13]. These frameworks are demonstrated with diagrams in Figs. 1 and 2.

The achieved results in [13] indicate feasibility of the combination of subsets of most commonly used indexes in the applied $10 \times 20$ CV (denoted as Com-Mask) and Framework 2 and superiority of Task Specificity compared to Subject Specificity in a 2 motor imagery dataset.

3 Dataset

EEG data from the dataset IIIa of BCI Competition III is used [18]. The dataset contains EEG data of 3 healthy subjects ($k3b$, $k6b$, $l1b$) performing 4 motor imagery tasks (left hand, right hand, foot, tongue). The sample rate is 250Hz and band pass filter in the range of 1Hz and 50Hz is applied. 60 electrodes are used and the task period is set to 3s [18]. To be consistent with previous studies [1, 12, 12] the first and last 0.5s of each epoch iare considered as pre and post transition periods and omitted and the signal is sub-windows to 0.5s. Common Average Referencing (CAR) and Demeaning (D) are the applied and frequency features are extracted. Extreme Learning Machine (ELM) with sigmoid kernel and 80 nodes is used for internal evaluation of particles in the swarm and polynomial SVM and a modified single layer perceptron that incorporates early stopping are
Fig. 1. Diagram representing the appliance of Framework 1 on a dataset with 3 subjects. Meta dataset represent a repository that contains the extracted masks for each fold of $10 \times 20$ CV and their informedness results. ValMask and TesMask represents the best performing masks on validation and testing sets respectively. ComMask represents a subset of most commonly used indexes in $10 \times 20$ CV.

used for final evaluation on the testing set. All experiments follow $10 \times 20$ cross validation (CV) paradigm that creates sets with ratio of 90%, 5%, and 5% for training, validation and testing respectively. Bookmaker informedness is used to assess the classification performance. detail discussion about bookmaker can be found in [19–21].

4 Experimental Design and Achieved Results

This section introduced the conducted experiments for investigating the objectives of the study presents the achieved results. In all experiments, the PSO paradigm is parameterized to generate masks that contain 30 feature and 10 electrode indexes. In velocity equation of the PSO, EQ. 1, $c_1$ and $c_2$ are set to 0.5 and 2.5 respectively and $r_1$ and $r_2$ are random values in the range of 0 and 1. Linear Decreasing Inertia Weight (LDIW) is used to update the inertia weight ($w_1=0.2$ and $w_2=1$).

4.1 Experiment 1: The impact of PSO paradigm

This experiment investigates the impact of proposed PSO paradigm as a DR method. The results depicted in Figs. 3 and 4 illustrate the average achieved informedness with either of the used classifier (Polynomial SVM, Sigmoid ELM, and modified Perceptron) within each subject ($k3b$, $k6b$, $l1b$) with the masks that best represent the testing set, validation set and most commonly selected indexes. The procedure is based on first, applying PSO paradigm in a $10 \times 20$ CV to extract the masks and later, reapply the suggested masks to a second $10 \times 20$ CV to assess the performance. Given that the applied PSO paradigm
Fig. 2. Diagram representing the appliance of Framework 2 on a dataset with 3 subjects. Super Subject represent the EEG dataset resulted by the concatenation of preprocessed EEG signals of two other subjects that performed similar tasks. The preprocessing stage include demeaning, common average referencing, and extraction of frequency features. ValMask and TesMask represents the best performing masks on validation and testing sets respectively. ComMask represents a subset of most commonly used indexes in $10 \times 20$ CV.

reduces approximately 90% of the dimensions of the data, the lost of average 0.1 informedness with in subjects is likely to be acceptable. Among the used masks and classifiers, combination of common mask and polynomial SVM shows better performance across subjects. This is likely to be due to the fact that this mask represent provides better generalizability since it represent the indexes that appeared in most solutions while in each subject ValMask and TesMask represents the masks that are fine tuned on validation and testing sets in the first $10 \times 20$ CV. The results achieved by each subject without any dimension reduction is included in figures to help understanding the impact. This is illustrated as FullSet in Figs. 3 and 4.

4.2 Experiment 2: Subject Specificity

This experiment is designed to investigate the feasibility of the extracted masks through PSO paradigm for Subject Specificity using Framework 1. The procedure is to use the masks generated by PSO on each subject on others. It is noteworthy that in experiment 1 the extracted masks through the first layer $10 \times 20$ CV where fine tuned on 95% of the data within different folds of CV, contamination between training and testing samples is possible. This issue can be resolved through this experiment given that the used masks are generated based on the applied procedure on EEG data of other subjects.

The issue of Subject Specificity is investigated with in two experiments. Assuming three target subjects (k3b, k6b, l1b), in experiment 2 (a) the masks generated on the EEG data of another subject is used to reduce the dimensions of the target subject. As an instance, assuming subject k3b as target subject, two experiments are conducted to investigate the impact of masks originated from
subject k6b and subject l1b separately. In Fig. 5 this is denoted as $k6b > k3b$ and $l1b > k3b$ respectively.

In experiment 2b the masks extracted from two subjects (individually) are combined together in a meta dataset following the description of Framework 1 in Fig. 1. As an instant, assuming subject k3b as the target subject, this issue is illustrated as $k6b/l1b > k3b$ in Fig. 6.

### 4.3 Experiment 3: Task Specificity

This experiment is designed to investigate the feasibility of the extracted masks through PSO paradigm for Task Specificity using Framework 2 (Fig. 2). Given that the used dataset only contains 3 subjects, the experiment Super Subject is created based on the concatenation of EEG data of 2 subjects and the third subject is considered as the target subject. The results are illustrated in Fig. 7.
Fig. 4. The feasibility of the used masks on subject 11b (Exp. 1).

Fig. 5. (Exp. 2a).
Fig. 6. (Exp. 2b).
Fig. 7. Exp. 3
A comparison between the achieved performance with the conducted experiments indicate the potential of Framework 1 to be used for subject transfer. In addition, the results indicate the generalizability of the generated masks. Within the conducted experiments for subject transfer, the best average classification performance across subjects achieved in experiment 2a with the combination of ComMask (commonly selected indexes) and polynomial SVM.

5 Conclusion

This study examined the potential of a proposed PSO paradigm for dimension reduction of EEG data in addition to investigating its feasibility for subject transfer through the use of two Frameworks representing Subject Specificity and Task Specificity. The results illustrate the potential of the used paradigm for dimension reduction in terms of providing generalizable solutions in addition to demonstrating encouraging performance in subject transfer problem.

The illustrated generalizability within the generated solutions of PSO paradigm is consistent with our previous findings in [1]. The results achieved from Frameworks 1 and 2 that indicate Task Specificity and Subject Specificity contradicts with our previous findings in [13]. In [13], the best overall performance achieved through the use of a super subject representing EEG signal concatenation of 4 subjects with in Framework 2 paradigm (using 2 motor imagery dataset) while in this study the best overall performance across subject achieved through re-applying the ComMask (commonly selected indexes) generated from one subject on target subject with in Framework 1. The poor performance achieved with Framework 2 in this study is likely to be due to the lack of having enough number of subjects with proper variation of expertise for creation of Super Subject.

Further investigation with other datasets that contain higher number of subjects is required.

References