A zero-training algorithm for EEG single-trial classification applied to a face recognition ERP experiment

Agustín Lage-Castellanos, Juan I. Nieto, Ileana Quiñones and Eduardo Martínez-Montes.

Abstract—This paper proposes a machine learning based approach to discriminate between EEG single trials of two experimental conditions in a face recognition experiment. The algorithm works using a single-trial EEG database of multiple subjects and thus does not require subject-specific training data. This approach supports the idea that zero-training classification and on-line detection Brain Computer Interface (BCI) systems are areas with a significant amount of potential.

I. INTRODUCTION

EEG single trial classification has been extensively researched in the recent past. Previous works mainly concentrate on the development of subject-specific algorithms, based on the classification of single trials associated with a particular cognitive state [1], [2]. These algorithms require a training session where model parameters are optimized according to a training data set. BCI systems have primarily focused on problems related to motor imagery and the P300 speller [3]. However, BCI systems for face recognition problems have yet to be more thoroughly explored.

Recent works have proposed zero-training BCI algorithms, which eliminate the need for a training session and allow on-line detection of single trials [4], [5]. Due to the high inter-subject variability of EEG signals, this problem represents a great challenge to current methods.

In this paper we present a zero-training single-trial classifier applied to a face recognition ERP experiment. The objective of the experiment is the recognition of known faces embedded in a large sequence of unknown ones. The algorithm classifies the two different stimuli conditions using EEG single trials in a binary classification scheme.

To evaluate the performance of our method, we used a leave-one-out subject approach, where all the subjects in the database, except the tested one, were used to build the classification model. This process was done for each subject. We studied how inter-subject variability altered the algorithm results. The algorithm presented uses statistical analysis to select the most relevant signal features avoiding the use of predefined parameters that could be problem-specific. Thus, it can applied to a broad range of EEG classification problems.

Juan I. Nieto is with the Australian Centre for field Robotics, University of Sydney, Australia. (j.nieto@cas.edu.au)

Agustin Lage-Castellanos, Ileana Quiñones and Eduardo Martinez-Montes are with the Cuban Neuroscience Center, Havana Cuba. (agustin, ileana, eduardo)@cneuro.edu.cu

II. METHODS

A. Experimental Data and Electrophysiological recording

To support our study we present results with real data obtained by an oddball paradigm that consists of a face recognition task recorded in eight healthy subjects [6]. The experiment comprised of two experimental conditions: familiar faces, labeled as positive, and unknown faces, labeled as null conditions. Familiar faces were extracted from popular characters of a contemporary soap opera. Each familiar face was presented twice; comprising a total of 14 familiar faces randomly shuffled with 280 unfamiliar ones. The subjects were shown images in three runs, which amounted to a total of 42 familiar (5%) and 840 unfamiliar faces (95%). Subjects were required to discriminate familiar faces while responding to an inter-stimulus interval on a computer keyboard. Face stimuli were presented for 1000 ms, with an inter-stimulus interval of 1000 ms.

Data acquisition was carried out with 19 monopolar disk electrodes (Ag/AgCl) using a Medicid5 according to the 10-20 system, referenced to an electrode placed on the nose. Those epochs with generalized artifacts were eliminated, maintaining the relative ratio of familiar and unfamiliar trials in all subjects to 5/95%.

B. Algorithm Overview

We followed a machine learning approach based on the following steps: pre-processing, feature extraction and classification. For each test subject the algorithm was trained using the remaining seven subjects data.

Table I summarizes the algorithm scheme. First, the database and the target subject data were pre-processed using a low pass filter. Then, training was carried out using only database single-trials x_{Db}. Each electrode was used as an independent predictor and then the individual estimators were boosted into a single one through a linear classifier. This experiment contained a set of 19 electrodes. Each electrode contributes with one feature to the feature matrix f_{Db}. For each electrode j, the relevant time window $\{t\}_j$ was selected at the step called time segmentation. Furthermore, the relevant features were extracted for each trial in the database using the z-score statistic at the point of maximum amplitude. The *z-score* statistic is a dimensionless quantity that measures the distance between an observation and the population mean considering the population standard deviation. Thirdly, electrode weights $\{w\}$ for classification were assessed by linear discriminant analysis using the training classification vector z_{Db}.

During the test stage, the parameters estimated at the training stage were used for predictions. These parameters are: the relevant time window for each electrode $\{t\}_j$, the distribution mean and standard deviation for the training set in the positive condition, and the electrode weights $\{w\}$ computed via the linear classifier. Details are presented next.

Table I: Algorithm Scheme	
Input	x_{Db}, z_{Db}, x_{test}
Preprocessing	(x_{Db}, x_{test})=lowPassFilt(x_{Db},x_{test})
Training Stage	
	for each electrode $j = 1 : 19$
time segment	$\{t\}_j$=time_segment(x_{Db}^j, z_{Db})
feat extraction	$f_{Db}(:,j)$= z_score($x_{DB}^j, x_{Db}^j, \{t\}_j$)
	end
Classify	$\{w_{1..19}\}$ = linclass(f_{Db}, z_{Db})
Test Stage	
	for each electrode $j = 1 : 19$
	$f_{Test}(:,j)$ =z_score($x_{Test}^i, x_{Db}^j, \{t\}$)
	end
Output	$z_{Test} = $ linpred($f_{Test}, \{w\}$)

C. Preprocessing

The only preprocessing step included in our approach was a temporal low-pass filtering to remove high frequency noise. The EEG signals were filtered with a 10th order low-pass digital Butterworth filter with a cut-off frequency at 7 Hz. Experimentation proved this value to be the best. It is worth mentioning that similar classification results were obtained for cut-off frequencies in a broad range (5-20 Hz), showing that although this is an important parameter, fine tuning is not essential.

D. Time Segmentation

EEG signals were segmented by automatically selecting a time window for each electrode. The window was calculated using training data. A two-sample location T-test was performed between the means of EEG trials corresponding to the positive and null classes. The test assumes that the two samples come from normal distributions with unknown and unequal variances. Samples with p-value lower than 0.01 were selected. A sliding window was used to select the continuous data segment with the largest number of selected samples. In the implementations presented, we used windows of 60 samples (300 ms). The automatic time window selection is a very important feature of our approach. It avoids fixed predetermined time window locations that may not be optimal for particular subjects and/or sessions due to well known variations in latency that characterize ERPs.

E. Feature Extraction

Feature extraction reduces data dimensionality focusing in those dimensions that most contribute to accuracy in the classification. This is a central aspect of algorithms that deal with problems of high dimensionality. In particular, in this experiment, each trial is formed by the time series of 19 electrodes, each one having a length of 200 time points.

As result, we deals with 3800 dimensions for each trial. The features were selected based on the distinctive trait that single trials corresponding to the positive condition elicit ERP potentials of higher amplitudes within the selected time windows.

Once the relevant time window for each electrode was selected, we extracted features using a z-score test. The process is ilustrated in Fig 1. We applied the z-score statistic between the maximum of each trial and the maximum of the mean of all positive trials in the database, within the relevant time window. This strategy is useful for situations where the maximum of the mean and the particular maximum for each single trial could have differences in latencies.

Each data point is defined by three dimensions: i the trial index, j the electrode index and t the time index. The maximum for a particular trial i in a particular electrode j is the maximum amplitude for this trial, within the time window: $\{t\}_j$, selected at the time segmentation step.

$$x_{ij\,\max} = \max(x_{ij\{t\}_j})$$

The maximum computation equally holds for the mean of the positive condition μ_{jt} at electrode j at time t. The trial dimension i is absent since the algorithm uses the mean of all positive trials in the database.

$$\mu_{j\,\max} = \max(\mu_{j\{t\}_j})$$

The standard deviation $\sigma_{j\,\max}$ is computed across all trials at the point of occurrence of the maximum of the mean. Fig 1 describes the process for Pz ($j = 19$). As can be seen, the single trial and the database means have maximums at different latencies.

The *z-score* statistic calculates the difference between the two maximums, taking into account the variations of this maximum in the database for the positive condition (Fig 1):

$$f_{ij} = z_{score} = \frac{x_{ij\,\max} - \mu_{j\,\max}}{\sigma_{j\,\max}}$$

Using this feature extraction method we obtained a single scalar feature value f_{ij} for the electrode j, at trial i. Then the feature matrix has dimension: {number of trials · number of electrodes}. At this level individual electrodes can be used as individual predictors. We analyze this point further in the results section. Nevertheless, it will be shown that the best strategy is to combine all electrodes predictors using a linear classifier that weights them according to their classification accuracy.

F. Classification

Classification consists on taking an input vector of features f and assign it to one of k discrete classes $\mathbf{C}_k$. The training dataset consists on N samples $(\mathbf{f}_{1,...}\mathbf{f}_n)$ in a D dimensional space ($D = 19$), and N labels ($z_1...z_N$). This particular ERP paradigm has two experimental conditions: familiar faces and unknown faces, thus classification is a binary problem ($k = 2$). A classification algorithm consists mainly of two parts: (i) a cost function, and (ii) an optimization process over the cost function that finds a decision boundary between classes.
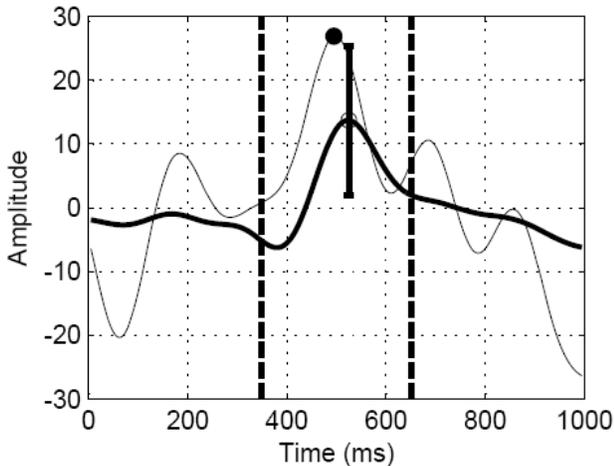
Fig. 1. Time segmentation and the feature extraction step for the electrode Pz ($j = 19$). The dashed vertical bars define the selected window of differences $\{t\}_{19}$ between the null and the positive conditions. The bold line is the mean of the positive condition in the database $\mu_{19,t}$; the vertical bar shows the standard deviation at the point of occurrence of the maximum $\sigma_{19,\max}$. The thick line is an individual trial selected at random $i = 12$; the black circle shows the maximum $x_{i=12\ j=19\,\max}$.

According to the cost function and the optimization process used, classification algorithms can be categorized in different types; e.g. linear, non-linear, generative, discriminative, etc. [7]. Since comparison of different classifiers is not the aim of this work, for the implementations presented here we chose from the literature the algorithms that have shown to be the most effective in EEG classification: Linear Discriminant Analysis (LDA), Logistic Regression (LR) and Support Vector Machines (SVM) [8], [9], [10].

III. RESULTS

A. Classification Algorithms

We explored three linear classifiers: linear discriminant analysis (LDA), support vector machines (SVM), and logistic regression (LR). These methods were implemented and compared using ROC areas. Table II illustrates the results obtained.

Table II: classifiers comparison. ROC areas			
subject	LDA	LR	SVM
1	0.739	0.751	0.747
2	0.925	0.900	0.915
3	0.685	0.661	0.734
4	0.847	0.829	0.816
5	0.933	0.916	0.918
6	0.939	0.932	0.934
7	0.924	0.928	0.924
8	0.939	0.934	0.932

As can be seen, similar results were obtained with the three of them, none of them presenting a clear improvement in performance respect to the other ones. Therefore, for the rest of the results presented in this work, we use LDA due to its simplicity and advantage in computational speed. Another important feature of LDA is that it yields a score that has a probabilistic interpretation, thus allowing assessment of the uncertainty of classification.

B. Subject-independent Models using LDA

Table III shows results for subject-independent models obtained with LDA. The first column illustrates ROC area relative to a cut-off point at 20% False Positive Rate (FPR). This is the region of ROC curves where the usual cut-off points are selected. The maximum possible value for this area is 0.2. The second column shows the total ROC area. The third column shows True Positive Rate (TPR) at 80% of True Negative Rate (TNR). We used TPR instead of classification accuracy due to the severe class imbalance (5%/95%). The expected TPR if classification is made by chance using this class proportions is 5%. Fig 2 shows ROC curves for all subjects. With exception of subjects 1 and 3, the algorithm yields TPR above 78.8%.

Table III: ROC results.			
subject	$A_{0.2}$	A	$TPR_{TNR=80\%}$
1	0.092	0.739	56.25
2	0.162	0.925	90.48
3	0.082	0.685	52.38
4	0.131	0.847	78.79
5	0.158	0.933	88.46
6	0.167	0.939	90.48
7	0.156	0.924	86.67
8	0.163	0.939	94.87

Fig 3 illustrates classification results for subject 8. It compares TPR versus TNR for all individual electrodes and the result obtained by combining them using LDA. For this plot we used the natural cut-off 0 for both: individual electrode z-scores f_{ij} and LDA classification scores. As can be seen, combining all the electrodes into one predictor improved the results obtained with the individual electrode models.

IV. DISCUSSION AND CONCLUSIONS

In this work we presented a zero-training algorithm for EEG single-trial classification. This algorithm handles the high inter-trial variations by the time segmentation that removes noisy regions of the signal and through the use of the maximum function at the feature extraction step. The maximum function deals with differences in latency between the database means and target single trials. The variations in amplitude of single trials are included in the algorithm by using *z-score* statistics. Results showed TPR above 78%. This value is even comparable with previously presented work in subject-specific experiments which shows the good potentials of our method for the more difficult subject-independent classification problem. Additionally, we explore a standarization approach to impose constant variance across trials for each subject. This strategy improves the TPR in one subject and worsens the TPR in other; the rest remains unchanged, resulting in a not useful approach.
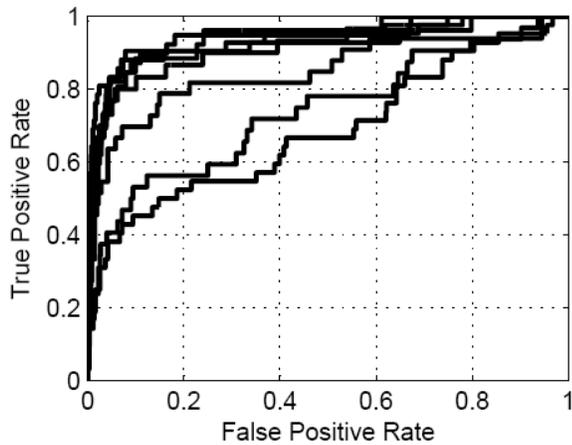
Fig. 2. ROC curves for the 8 subjects studied. ROC curves were computed using the classification scores obtained by LDA. Those curves with lower areas correspond to subjects 1 and 3.
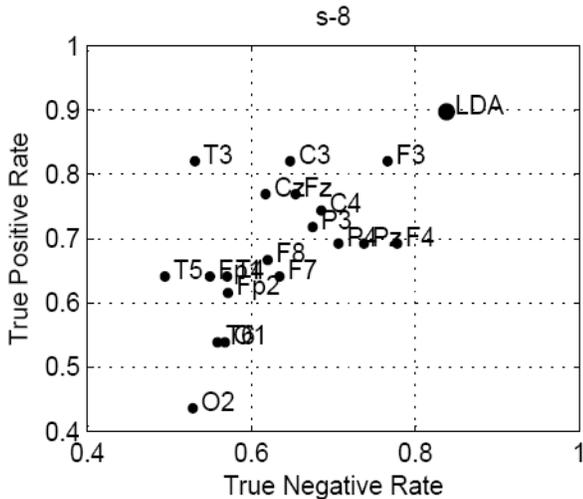

Fig. 3. True Positive Rate and True Negative Rate for all electrodes, when they are used as independent predictors, for subject 8. The biggest circle shows the result of combining all electrodes using linear discriminant analyisis (LDA).

An important attribute of this algorithm is the interpretability of electrodes relevance. It was shown that LDA can be used to combine individuals models, weighting electrodes according to their classification accuracy. This property can be used to include feedback, as short training sesions, that updates the algorithm with the objective of improving individual predictions.

As shown in the results section, subjects 1 and 3 exhibited lower classification accuracy than the remaining subjects. This is due mostly to inter-subject variability. The subjects correlation structure can be studied via hierarchical clustering using ERP data. Fig 4 shows a graph of the hierarchical clustering analysis across subjects using ERP signals corre-
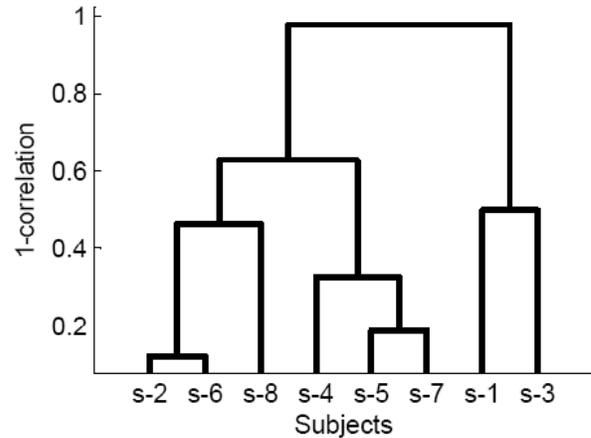

Fig. 4. Dendrogram of the hierarchical cluster tree of the ERP signals, corresponding to the positive condition. The cluster algorithm uses the correlation metric to link nodes

sponding to all electrodes and all time points of the positive condition concatenated. Correlation was used as similarity measure to form the clusters. This clustering structure reflects that single trial signals of subject 1 and 3 differ from the rest of the subjects. As a result, the single trial database is not a good predictor for these subjects. For the rest of the subjects, the single trial database works well to build subject-independent models.

REFERENCES

[1] C.Wang, C. Guan, H. Zhang, "P300 brain-computer interface design for communication and control applications", Engineering in Medicine and Biology Society, 2005, IEEE-EMBS 2005. 27th Annual International Conference of the of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference (2005), pp 5400-5403.

[2] F. Lotte, M. Congedo, A. L ecuyer, F. Lamarche, B. Arnaldi, A review of classi cation algorithm for eeg-based brain-computer interfaces, J Neural Eng 40 (4).

[3] E. Donchin, K. M. Spencer, R. Wijesinghe, "The mental prosthesis: Assessing the speed of a p300-based brain-computer interface", IEEE Transactions on Rehabilitation Engineering 8, 2000, pp 174-179.

[4] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Muller, V. Kunzmann, F. Losch, and G. Curio, The Berlin brain-computer interface: EEG-based communication without subject training, IEEE Trans, Neural Syst. Rehabil. Eng., vol. 14, no. 2, pp. 147-152, Jun. 2006.

[5] Shijian L, Guan C, Haihong Z, "Unsupervised brain computer interface based on inter-subject information", 30th Annual International IEEE EMBS Conference, 2008, pp 638-641.

[6] A. Lage, E. Martinez, J. Hernandez, L. Galan, False discovery rate and permutation test: an evaluation in erp data analysis, Statistics in Medicine, 2009, pp 1-29.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer; 2006.

[8] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. Mc-farland, T. M. Vaughan, J. R. Wolpaw, A comparison of classification techniques for the p300 speller, J. Neural Eng. 3 (4), 2006, pp 299 305.

[9] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, B. Arnaldi, A review of classification algorithms for eeg-based brain-computer interfaces, J Neural Eng 40(4).

[10] V. Abootalebi, M. H. Moradi, M. A. Khalilzadeh, A new approach for eeg feature extraction in p300-based lie detection, Computer Methods and Programs in Biomedicine 94 (1), 2009, pp 48 57.