Biomedical Signal Processing and Control 46 (2018) 166–173

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Elastic net ensemble classifier for event-related potential based automatic spelling

Sunghan Kim [a,*], Austin White [a], Fabien Scalzo [b], David Collier [c]

[a] Biomedical Instrumentation & Data Analysis Laboratory, College of Engineering and Technology, East Carolina University, Greenville, NC, USA
[b] Department of Neurology, UC Los Angeles, CA, USA
[c] Department of Pediatrics, Brody School of Medicine, East Carolina University, Greenville, NC, USA

ARTICLE INFO

Article history:
Received 12 May 2017
Received in revised form 2 April 2018
Accepted 23 June 2018

Keywords:
Brain computer interface
Elastic net regression method
Ensemble classifiers
Event-related potentials
Data sparsity
Oddball paradigm
P300

ABSTRACT

Background: Brain–computer interface (BCI) refers to a direct communication between the brain and an object to be controlled by information embedded in various signals originated from the brain. Electroencephalogram (EEG) is the most common brain signal modality due to its noninvasiveness, ease-of-use, and high time-resolution. P300-based automatic spelling is the most celebrated EEG-based BCI system, which allows its user to relay a message by just staying focused on letters and numbers displayed on a screen. The objectives of this work are (1) to introduce a novel P300 detection algorithm and (2) to compare its performance against the best current practice algorithm for P300 detection.
Results: Four volunteers tried to spell letters, "PIRATES", using our automatic spelling system after proper training. Our results indicate that the proposed P300 detection algorithm can speed up the automatic spelling process since it requires a smaller number of flashing sequences than the best current practice algorithm does to recognize target event-related potentials. While the proposed P300 detection algorithm performs better than the best current practice algorithm, it does not necessarily require a heavy computational burden.
Conclusion: We designed a novel P300 detection algorithm assuming that the sparsity of EEG signals could be effectively utilized to detect target event-related potentials such as P300. Our pilot study results indicate that utilizing the sparsity of EEG signals can improve the automatic spelling experience.

1. Background

The feasibility of controlling external devices via brain signals was first demonstrated by Vidal in 1973 [1]. He first introduced a term, brain–computer interface (BCI) referring to a direct communication between the brain and an object to be controlled by relaying information embedded in the brain's signals to the object. The most common brain signal modality for BCI is electroencephalogram (EEG), which measures small voltage fluctuations on the scalp resulting from the cortical neurons' synchronized postsynaptic activities. Due to its noninvasiveness, ease-of-use, and high time-resolution, EEG is often preferred for practical BCI applications over other signal modalities such as electrocorticogram (ECoG), functional magnetic resonance imaging (fMRI), and positron emission tomography (PET). Most EEG-based BCI systems utilize one of four types of EEG signals including slow cortical potentials, μ- or β-rhythms, sensory evoked potentials (SEPs) and event-related potentials (ERPs) [2–6].

The advent of powerful low-cost computer equipment over the last two decades has led to successful development of several BCI systems. Arguably the most celebrated EEG-based BCI system is P300-based automatic spelling [7]. The P300 wave was discovered half a century ago and has since been the major research component in the field of ERP studies [8]. For visual/auditory stimuli, the latency of the P300 wave ranges from 250 ms to 400 ms for most adult subjects between 20 and 70 years of age. The oddball paradigm is the most popular method to elicit P300 ERPs [9,10]. In this paradigm two (or three) types of stimuli are presented in a series such that one type of stimuli occurs more frequently than the other. The less

Abbreviation: AUC, area under the curve; BCI, brain–computer interface; CAR, common average reference; EEG, electroencephalogram; ENEC, elastic-net ensemble classifier; ERP, event-related potential; GLMNET, general linear models elastic-net regularization; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; SWLDA, step-wise linear discriminant analysis.

* Corresponding author.
E-mail addresses: kims@ecu.edu (S. Kim), whiteau12@students.ecu.edu (A. White), fscalzo@mednet.ucla.edu (F. Scalzo), collierd@ecu.edu (D. Collier).

https://doi.org/10.1016/j.bspc.2018.06.005
1746-8094/

frequent stimuli are referred to as “oddballs”, which actually elicit P300 ERPs.

Donchin et al. introduced an automatic spelling system, which enabled a user to relay a message by sequentially selecting from among the 36 letters and numbers displayed on a screen by utilizing P300 ERPs elicited by rare visual stimuli [5,7]. The standard automatic spelling system utilizes a 6-by-6 matrix containing the letters of the alphabet and numbers. The rows and columns of the matrix are intensified (or flashed) one at a time in a random order. The user is to focus his/her attention on one cell of the matrix that contains the intended letter to be spelled. A P300 ERP is elicited whenever the row or column containing the cell of the user’s choice is flashed. Since the automatic spelling system’s performance is directly associated with discriminating ERPs with the P300 component from those without it, it is critical to incorporate an effective and reliable discrimination or classification algorithm into the spelling system.

Numerous discrimination/classification algorithms have been proposed for automatic P300 spelling including Fisher’s linear discriminant analysis (FLDA), Bayesian linear discriminant analysis (BLDA), support vector machines (SVM), Pearson’s correlation method and stepwise linear discriminant analysis (SWLDA) [11–16]. Among those algorithms, SWLDA has been deemed as the most widely accepted one due to its ease-of-use, small number of parameters to tune, and high classification accuracy [13]. When tested against other P300 detection algorithms, SWLDA has regularly outperformed with regards to classification accuracy and feature selection [11,12,16]. Having demonstrated consistently superior performance in previous studies, SWLDA was selected as the best current practice for P300 detection, against which our proposed algorithm was compared.

The objectives of this paper are (1) to introduce a novel P300 detection algorithm, i.e. Elastic Net Ensemble Classifier (ENEC), which exploits the sparsity of EEG data, and (2) to compare its performance against SWLDA utilizing real EEG data collected from four healthy volunteers who performed P300 automatic spelling. The proposed algorithm (ENEC) is an enhanced and specialized version of the ensemble sparse classifier (ESC) that was previously introduced by the authors as a general classification technique for high-dimensional biological data sets. More details about the ESC technique can be found in [17,18]. Among those four volunteers, three of them had previous experience with P300 spelling while the other one was a first-time user.

2. Methodology

2.1. Elastic net ensemble classifier

Elastic net ensemble classifier (ENEC) is a modified bagging technique that generates an ensemble of weak classifiers. Each weak classifier of ENEC is a linear sparse solver (i.e. elastic net regressor) that exploits the sparsity of a given data set. P300 EEG data are “sparse” in a sense that only a small number of features or variables (signal sample points in our application) are relevant to the occurrence of P300 ERPs. Given an underdetermined linear system, $y = X\beta$, the sparsest solution is the one containing the smallest number of non-zero elements. It can be expressed as,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \|\beta\|_0 \quad \text{subjectto} \quad y = X\beta \tag{1}$$

where $\|\beta\|_0$ is the ℓ_0-norm of β, y a $n \times 1$ outcome vector, X a $n \times m$ matrix, and β a $m \times 1$ linear coefficient vector to be estimated. However, minimizing the ℓ_0-norm of β is mathematically intractable as the number of variables, m, increases because it requires a combinatorial search.

Assuming certain conditions on the measurement matrix, X, the sparsest solution can be well approximated by minimizing surrogate measures such as $\|\beta\|_p$-norms where p is not equal to 0. For example, the LASSO (least absolute shrinkage and selection operator) method seeks for a sparse solution by minimizing the absolute square error while regularizing the ℓ_1-norm of β [19]. It can be expressed as,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \|y - X\beta\|_2^2 \quad \text{subjectto } \|\beta\|_1 \le \lambda \tag{2}$$

where λ is a regularization coefficient. However, the LASSO method has several limitations [20]. For example, the LASSO method cannot handle highly correlated variables properly and performs poorly when the number of variables, m, is much greater than that of observations, n.

Zou and Hastie proposed a new regularization and variable selection method called the elastic net, which treats strongly correlated variables as a group [20]. Their simulation results demonstrated the superior performance of the elastic net over the LASSO and its usefulness when the number of variables, m, is greater than that of observations, n. The solution from the elastic net method is defined as follows,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \|y - X\beta\|_2^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2 \tag{3}$$

where λ_1 and λ_2 are two regularization coefficients. When $\lambda_1 = \lambda$ and $\lambda_2 = 0$, the elastic net becomes equivalent to the LASSO in Eq. (2). If $\lambda_1 = 0$ and $\lambda_2 = \lambda$, it becomes a ridge regression. In general, the elastic net is a regularized regression method that combines the ℓ_1-norm penalty (i.e. LASSO) and ℓ_2-norm penalty (i.e. ridge) [19–21].

Algorithm 1. Elastic net ensemble classifier.

Training phase
Divide the training data set $\{X, y\}$ into positive &negative sample sets
$\{X^+, y^+\} \bigcup \{X^-, y^-\} = \{X, y\}$
$\{X^+, y^+\} \bigcap \{X^-, y^-\} = \emptyset$
Build N_{wc} weak classifiers
for $i = 1, \ldots, N_{wc}$ **do**
Draw N_p random samples from $\{X^-, y^-\}$ w/o replacement
$\{X_i^-, y_i^-\} = \{X(1)^-, y(1)^-, \cdots, X(N_p)^-, y(N_p)^-\}$
Create the ith training data set $\{X_i, y_i\}$
$\{X_i, y_i\} = \{X^+, y^+\} \bigcup \{X_i^-, y_i^-\}$
Identify the weight of critical (non-zero) features
$\beta_i \longleftarrow$ **ElasticNet**(X_i, y_i)
Compute the performance weight ω_i
$\{X_i^c, y_i^c\} = \{X, y\}$ -{Xi, yi}
$\omega_i = 1 - \dfrac{\|y_i^c - \operatorname{sgn}\left(X_i^c\beta_i\right)\|_1}{2(N_s - 2N_p)}$
end for
Normalize the performance weight ω_i
for $i = 1, \ldots, N_{wc}$ **do**
$\tilde{\omega}_i = \dfrac{\omega_i}{\sum_{i=1}^{N_{wc}} \omega_i}$
end for
Prediction Phase
Predict the label of a new sample x_{new}
$\hat{y}_{new} = \operatorname{sgn}\left(\sum_{i=1}^{N_{wc}} \tilde{\omega}_i x_{new}^{T} \beta_i\right)$
where $\{X^+, y^+\}$ is the set of all positive samples in the training data set $\{X, y\}$, $\{X^-, y^-\}$ the set of all negative samples, N_{wc} the total number of weak classifiers, N_p the number of all positive samples, $\{X_i, y_i\}$ the training data set for the ith weak classifier, $\{X_i^c, y_i^c\}$ the set of samples not included in $\{X_i, y_i\}$, and N_s the total number of samples in $\{X, y\}$.

Algorithm 1 represents the pseudo code for ENEC in the training and prediction phases. The operator **ElasticNet** was implemented by utilizing a publicly available GLMNET (General Linear Models Elastic-Net Regularization) package for MATLAB, which was originally developed by Friedman and his colleagues [22,23]. Each of the N_{wc} weak classifiers in the training phase is a linear regression model with the elastic net penalty utilizing a set of $2N_p$ samples