Boosting Linear Logistic Regression for Single Trial ERP Detection in Rapid Serial Visual Presentation Tasks

Yonghong Huang, Student Member, IEEE, Deniz Erdogmus, Member, IEEE
Santosh Mathan, Member, IEEE, Misha Pavel, Member, IEEE

Abstract— **In this paper, we employ the AdaBoost algorithm to the linear logistic regression model to detect encephalography (EEG) signatures, called evoked response potentials of visual recognition events in a single trial. In the experiments, a large amount of images were displayed at a very high presentation rate, named rapid serial visual presentation. The EEG was recorded using 32 electrodes during the rapid image presentation. Subjects were instructed to click the mouse when they recognize a target image. The results demonstrated that the boosting method improves the detection performance compared with the base classifier by approximately 3% as measured by area under the ROC curve.**

I. Introduction

In a variety of fields, from medical diagnosis to intelligence analysis, effective image search through large volumes of imagery for important information has become a crucial problem for practitioners in recent years. In general, computer vision based systems have proven to perform poorly relative to human image analysts due to the problem of optimizing image throughput. In an attempt to raise the efficiency associated with the search process, an effective triage system may be developed to leverage human perceptual capability. An effective triage mechanism would rapidly process high volumes of imagery and identify a subset of images that merit careful scrutiny by an image analyst. An ideal triage system might be one that leverages human visual processing capabilities in the role of a target detector, while dramatically raising the efficiency associated with the search process. One key issue to build an efficient triage system to exploit human visual processing capabilities is to utilize rapid serial visual presentation (RSVP) of images and encephalography (EEG).

Evoked response potentials (ERP) arise from some morphological changes in EEG waveforms in response to important task-relevant stimuli [1]. Prior research demonstrates that ERP in EEG signals, which reflect the activity of underlying cognitive processes associated with perceptual decision making, may be used to identify targets within image sequences presented at very high presentation rates [2-4]. During an RSVP task, a continuous sequence of images is rapidly presented. The neurophysiological response to target images during RSVP is recorded and documented. A target image in a sequence of nontarget distractor images elicits a stereotypical spatiotemporal response in the EEG. ERP could be used in conjunction with RSVP of images to dramatically raise the efficiency of searching through high volumes of imagery.

The challenge of ERP detection is that the system would require the capability to detect ERP reliably and quickly. ERP are difficult to detect due to its low signal to noise ratio. The traditional approaches rely on a strategy of trial averaging [5] in which an experimental stimulus is presented to a subject many times and the waveforms elicited by each stimulus are averaged. Repeated presentation of stimuli compromises the efficiency of the search process. In domains where efficient ERP detection is critical, accurate detection of ERP within a single trial becomes necessary. Therefore more efficient signal processing and classification approaches are needed for single trail detection of ERP.

There are various multivariate signal processing algorithms applied for EEG detection [6-8]. Linear techniques are commonly employed in ERP detection. The linear logistic regression model and other two nonlinear classifiers had been proved to be effective in single trial ERP detection for the RSVP task in our recent work [9]. Boosting techniques have undergone intense theoretical study and empirical testing. Boosting has been shown to give a significant improvement in performance in many applications, such as text and speech categorization [10][11]. In this work, we applied AdaBoost technique based on the linear logistic regression classifier on single trail detection of ERP in the context of a triage platform. The objective is to investigate efficiency of the boosting technique on the performance improvement for ERP detection.

II. Methods

A. Data Collection

1) Data Acquisition

Subjects were instructed to perform visual target detection amongst distractors. Objects of interest, referred to as

Yonghong Huang is with the Computer Science and Electrical Engineering Department, Oregon Health and Science University, Portland, OR 97239 USA, (phone: 503-748-1188; fax: 503-748-1548; e-mail: huang@csee.ogi.edu).
Deniz Erdogmus is with the Computer Science and Electrical Engineering Department, Oregon Health and Science University, Portland, OR 97239 USA, (e-mail: derdogmus@ieee.org).
Santosh Mathan is with the Human Centered Systems Group, Honeywell Laboratories, Minneapolis, MN 55418 USA, (e-mail: Santosh.Mathan@honeywell.com).
Misha Pavel is with the Biomedical Engineering Department, Oregon Health and Science University, Portland, OR 97239 USA, (e-mail: pavel@bme.ogi.edu).

targets, consisted of satellite photographs of ships or boats in the midst of a pool of satellite images around a port/coast scene. Both target and distractor images were drawn from a common high-resolution, broad-area, satellite image. All imagery was presented using the RSVP paradigm as shown in Figure 1. Images were presented in rapid succession for durations of 50 or 100 milliseconds per image.

EEG data was collected over the course of little over an hour. Each session lasted approximately 20 minutes with a 5-minute rest between sessions. A fixation screen, which lasted several seconds, was used to separate trails. Each trial contained a sequence of approximately 50 images. Of the trials, 50% contained targets while 50% did not. Each trial consisted of a sequence of images in which, if existed, a target image was positioned randomly (except at the first and last 10 images in the sequence).

The data were collected using a 32 channel BioSemi Active Two system. All channels were referenced to a common mean reference. Data was sampled at 256 Hz. Image presentation triggers were received by the BioSemi system over a parallel port and recorded concurrently with EEG signals. A variety of signal processing components were implemented for reducing the impact of noise artifacts that could compromise ERP detection.

2) Data Description

EEG data was segmented into *epochs* (shown as a trial in Figure 1). In the case of target trials, each epoch consisted of a two second segment of EEG, one second before, and one second after the onset of target stimuli. For the distractor trials (no target trials), epochs were extracted around the trigger associated with the middle image of each trial block. Data associated with each epoch were stored in a 32*512 matrix (number of channels times EEG samples ranging from 1s before to 1s after each target or distractor stimulus). Each session yielded 80 to 90 target epochs and 80 to 90 distractor epochs each. These sets of datasets are referred to as one session (collected in one continuous session in real time).

The pilot data we used in this work was collected from two subjects. The subject was instructed to indicate presence of targets by clicking the mouse at the end of each trial containing a target. There were three sessions for the each subject and the middle session was intended for classifier validation. In each session, there are one set data with targets and one set of data without targets. The features used for classification are simply the temporal EEG measurements from 32 channels at 512 time instances centered on the target stimuli.

B. Linear Logistic Regression Classifier

This is a state-of–the-art linear discrimination approach in ERP detection based on logistic regression [6-9]. The linear approach relies on the assumption that the EEG signals are a linear combination of distributed source activity and zero-mean white Gaussian measurement noise. Consequently, the optimal ERP detection strategy under this assumption is to determine optimal linear projections of the sensor measurements to maximize discrimination ability.

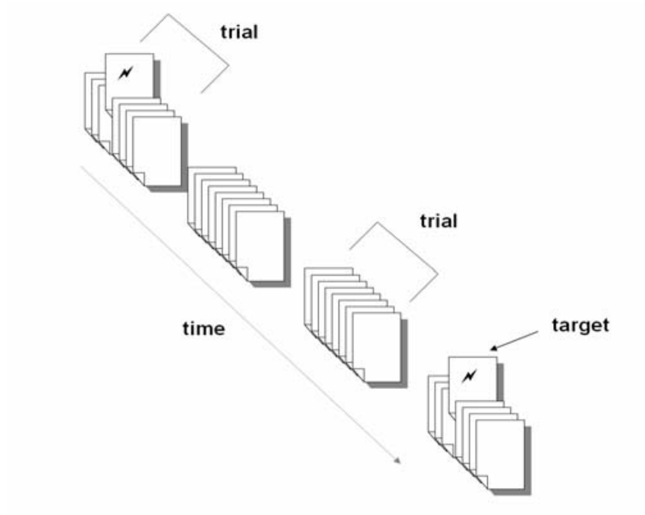

Figure 1. Experimental design. Subjects viewed trials with or without targets. 50% of trial blocks contained targets. Fixation screen separated trial blocks.

A linear discriminant function is defined as linear combinations of the components of $\mathbf{x} = [\mathbf{x}_{1\ldots}\mathbf{x}_n]^T$,

$$\mathbf{y} = \mathbf{w}^T \mathbf{x} + b \tag{1}$$

where $\mathbf{w}$ is the weight vector, b is the bias and n is the number of samples [12]. The linear projections are optimized using the logistic regression technique that assumes the conditional class probability given the projection will follow a logistic model,

$$f = p(c \mid \mathbf{x}) = \frac{e^y}{1+e^y} = \frac{e^{w^T \mathbf{x}+b}}{1+e^{w^T \mathbf{x}+b}} \tag{2}$$

which is consistent with the Gaussianity assumption. This likelihood is parameterized by the weight vector $\mathbf{w}$ and bias b. The parameters are adjusted by maximizing the likelihood of the data so that the data matches the logistic model distribution in (2). In order to compute the optimal coefficients efficiently, weighted least-squares was used as the objective function and batch gradient descent algorithm was used to optimize the parameters [13]. The objective function is defined as

$$J = \sum_{i=1}^{N} (c_i - f(\mathbf{x}_i))^2 \tag{3}$$

where c_i is the class label corresponding to input sample $\mathbf{x}_i$. The gradient for weight updates is simply

$$\nabla J = -2\sum_{i=1}^{N} (c_i - f(\mathbf{x}_i)) f_i'(\mathbf{x}_i)\mathbf{x}_i \,. \tag{4}$$

The weight vector is updated as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu \nabla J \tag{5}$$

where μ is a constant leaning rate.

C. Boosting Algorithm: AdaBoost

Boosting refers to a general and provably effective method for improving the accuracy of any given classification algorithm. The AdaBoost algorithm, introduced by Freund and Schapire has undergone intense theoretical study and empirical testing [10][11]. In this work, we applied AdaBoost based on the linear logistic regression method.

Given a training set $(\mathbf{x}_1,c_1),\ldots,(\mathbf{x}_n,c_n)$ where $\mathbf{x}_i$ is the input samples, n is sample numbers and $c_i \in \{-1,+1\}$ is the class label for each sample in a detection problem. The idea of AdaBoost is to create an ensemble of classifiers with identical topology sequentially, such that each classifier is trained emphasizing the samples incorrectly classified by the previous one in the sequence. This is achieved by iteratively training classifiers and reweighting the training samples, where the weights of incorrectly classified samples are increased in the training of the following classifier. The ensemble is created by a preselected T iterations/classifiers.

The algorithm is initialized by assigning equal (importance) weights $D_{1i}=1/n$ to each training sample. For $t=1,\ldots,T$, the sequence of classifiers $h_t(\mathbf{x})$ are trained using the weights D_t on the training samples, the objective function (3) and the weighted gradient

$$\nabla J = -2\sum_{i=1}^{n} D_{ti}(c_i - f(\mathbf{x}_i))f'(\mathbf{x}_i)\mathbf{x}_i \tag{6}$$

Then, the empirical error probability of the current classifier hypothesis $h_t(\mathbf{x})$ is evaluated using

$$\varepsilon_t = \Pr_{i\sim D_t}[h_t(\mathbf{x}_i) \neq c_i] = \sum_{i:h_t(\mathbf{x}_i)\neq c_i} D_{ti} \tag{7}$$

A positive weight α_t that is a monotonically decreasing function of the empirical error in the domain $\varepsilon_t \in [0,1/2]$ is assigned to the current hypothesis. Specifically, the function

$$\alpha_t = \frac{1}{2}\ln(\frac{1-\varepsilon_t}{\varepsilon_t}) \tag{8}$$

is utilized. Note that $\alpha_t \geq 0$ if $\varepsilon_t \leq 1/2$, and is monotonically decreasing. Furthermore, to emphasize the samples for which the current hypothesis is unsuccessful the training sample importance weights are updated as follows:

$$D_{(t+1)i} = D_{ti} \times \begin{cases} e^{-\alpha t}, if\ h_t(\mathbf{x}_i) = c_i \\ e^{\alpha t}, if\ h_t(\mathbf{x}_i) \neq c_i \end{cases} \tag{9}$$
$$= D_{ti}\exp(-\alpha_t c_i h_t(\mathbf{x}_i))$$

Before proceedings to the next iteration, D_{t+1} is normalized to become a distribution. The final hypothesis $H(\mathbf{x})$ is a weighted linear combination of the T hypotheses:

$$H(\mathbf{x}) = sign\left(\sum_{t=1}^{T}\alpha_t h_t(\mathbf{x})\right). \tag{10}$$

D. Evaluation

Receiver operating characteristic (ROC) analysis [14] is used to quantify a detection system's performance in the lack of definite risk assignment for the two types of errors (miss and false detection). An ROC curve shows the relationship between false positives and true positives. In the ROC curve the horizontal axis has the percentage of false positives and vertical axis has the percentage of true positives for a database sample. The final performance of this work is assessed using the area under the ROC curve. The actual probability of detection error depends on the frequency of targets, a factor that is determined by specific operational details.

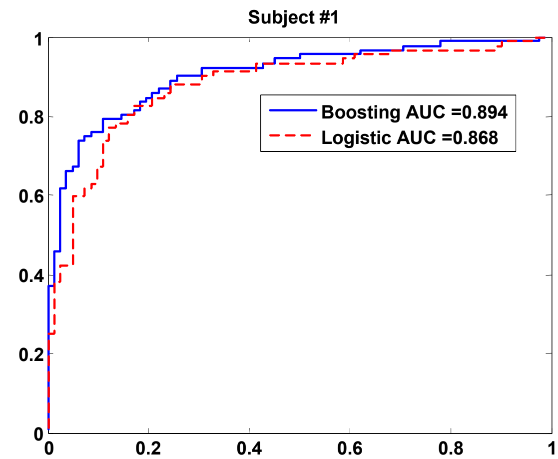

Figure 2. Prob(Detect) vs Prob(FalseAlarm) for subject #1.

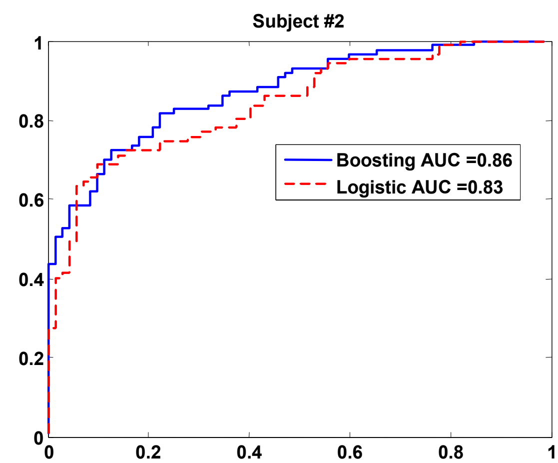

Figure 3. Prob(Detect) vs Prob(FalseAlarm) for subject #2.

III. Results

The goal of the experiments is to determine the effect of boosting algorithm on classification performance, as well as to assess the feasibility of rapid image search using the RSVP paradigm. The dataset was collected from two subjects. There were three sessions for the each subject and the middle session was intended for classifier validation. Each sample is a paired 512-point EEG measurement at 32 channels centered at the stimulus image and the true label (target/distractor) for this image. The data for distractors in an epoch other than the selected target image index are discarded in this study. For subject #1, there are 166 samples for training and 174 samples for test. For subject #2, there are 168 samples for training and 159 samples for test. The study evaluated the session-to-session transfer of classification performance. The aim is to compare the performance between the base classifier (linear logistic regression) with and without the boosting algorithm.[1]

In this experiment, we examine the performances of the base classifier with and with the AdaBoost technique across sessions on subject #1. We investigated training on session #1 and test on session #3. Figure 2 depicts the discrimination performance for subject #1. It can be observed that the base classifier has an area under the ROC curve of around 0.87 while the boosting classifier achieves an ROC area more than 0.89. The improvement is around 0.03. The result demonstrates that the AdaBoost performed better than the based method.

We conducted the same experiment on subject #2. The discrimination performance for subject #2 with training on session #1 and test on session #3 is illustrated in Figure3. We can observe that the based method only has an ROC area of 0.83 while the boosting method obtained 0.86, which shows a 0.03 improvement. The results indicate that the AdaBoost technique does boost the ERP detection performance.

From these results, it is clear that a discriminator trained on data from one session generalizes well to data from two test sessions, which were separated by over a twenty-minute gap. The results demonstrate that the AdaBoost algorithm is feasible and useful on across-session ERP detection for both subjects. However, the experiment results also suggest that more training data are needed to get even better performance.

IV. Conclusion

We studied the effectiveness of the AdaBoost algorithm based on the linear logistic regression classifier on single trial ERP detection in the context of RSVP target search in massive imagery databases. The results confirm that reliable visual target detection in large image databases is feasible with the RSVP paradigm and classification based on EEG measurements. The preliminary results presented here demonstrate that the boosting method outperformed the based classifier.

The raw temporal signal-based features coupled with dense EEG arrays yield very high dimensional feature vectors that make it infeasible to expect good generalization given the low sample to parameter ratio. Theoretically the performance of boosting will improve the performance for given sufficient data. However, in our particular problem, given the low number of instances in the datasets we have collected, it is reasonable the boosting does not bring huge improvements. The future challenges for this field include extraction of fewer more reliable features (perhaps wavelet-based).

[1] Previously, we compared various linear and nonlinear classifiers without boosting for this data and determined that linear logistic regression provides the best complexity-performance tradeoff [9].

Acknowledgment

This work was supported by DARPA under contract HM1582-05-C-0046 and by NSF under grant ECS-0524835. It has been approved for Public Release, Distribution Unlimited. The EEG data used in the experiments were collected at the Honeywell Human-Centered Systems Laboratory (Minneapolis, MN).

References

[1] S. Makeig, M. Westerfield, T-P. Jung, .S Enghoff, J. Townsend, “Dynamic brain sources of visual evoked responses”. *Science*, vol. 295, pp. 690–693, 2002.

[2] S. Thorpe, D. Fize, C. Marlot, “Speed of processing in the human visual system”, *Nature*, vol. 381, pp. 520-522, 1996.

[3] A.D. Gerson, L.C. Parra, P. Sajda, “Cortical origins of response time variability during rapid discrimination of visual objects”, *NeuroImage,* vol. 28, no. 2, pp. 326-341, 2005.

[4] A.D. Gerson, L.C. Parra, P. Sajda, “cortically-coupled computer vision for rapid image search”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, in press, 2006.

[5] R. Jones, “Brain waves in phase”. *Nature Reviews Neuroscience*, vol. 3, pp. 167, 2002.

[6] S. Makeig, A.J. Bell, T.Jung, T.J. Seinowski, “Indpendent component analysis of electroencephalographic data”, *Advances in Neural Information Processing Systems,* pp. 145-151, 1996.

[7] A. Hyvarianen, P.O.Hoyer, J. Hurri, and M. Gutmann, “Statistical Models of Images and early vision,” *Proceedings of International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning*, pages 1-14, June 2005.

[8] L.C. Parra, C.D. Spence, A.D. Gerson, P. Sajda, “Recipes for the linear analysis of EEG”, *Neuroimage*, vol. 28, pp. 326-341, 2005.

[9] Y. Huang, D. Erdogmus, S. Mathan, M. Pavel, “Comparison of linear and nonlinear approaches in single trial ERP detection in rapid serial visual presentation tasks,” *2006 IEEE International Joint Conference on Neural Networks*, July 16-21, 2006.

[10] Y. Freund and R. E. Schapire, “ A short introduction to boosting”, *Journal of Janpanse Society for Artificial Intelligence*, 14(5):771-780, 1999.

[11] R. E. Schapire, “ The boosting approach to machine learning an overview”, *MSRI Workshop on Nonlinear Estimation and Classification*, 2002.

[12] R. Duda, P. Hart, D. Stork, *Pattern Classification*, Wiley, New York, 2001.

[13] P. McCullagh, J. A. Nelder, *Generalized Linear Model*, 2nd ed., Chapman and Hall, London, 1999.

[14] T. Fawcett, “ROC Graphs: Notes and Practical Considerations for Data Mining Researchers”, *Technical Report, Copyright Hewlett-Packard Company*, 2003.