Neurocomputing 74 (2011) 2041–2051

Contents lists available at ScienceDirect

Neurocomputing

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

A framework for rapid visual image search using single-trial brain evoked responses

Yonghong Huang [a,*], Deniz Erdogmus [a,b], Misha Pavel [a], Santosh Mathan [c], Kenneth E. Hild II[a]

[a] Biomedical Engineering Department, Oregon Health & Science University, Portland, OR 97239, USA
[b] Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA
[c] Human Centered Systems Group, Honeywell Laboratories, Redmond, WA 98052, USA

ARTICLE INFO

Article history:
Received 7 December 2009
Received in revised form
1 November 2010
Accepted 9 December 2010
Communicated by S. Choi
Available online 21 March 2011

Keywords:
Brain computer interface (BCI)
Single-trial event-related potential (ERP)
Electroencephalography (EEG)
Image retrieval
Incremental learning

ABSTRACT

We report the design and performance of a brain computer interface for single-trial detection of viewed images based on human dynamic brain response signatures in 32-channel electroencephalography (EEG) acquired during a rapid serial visual presentation. The system explores the feasibility of speeding up image analysis by tapping into split-second perceptual judgments of humans. We present an incremental learning system with less memory storage and computational cost for single-trial event-related potential (ERP) detection, which is trained using cross-session data. We demonstrate the efficacy of the method on the task of target image detection. We apply linear and nonlinear support vector machines (SVMs) and a linear logistic classifier (LLC) for single-trial ERP detection using data collected from image analysts and naive subjects. For our data the detection performance of the nonlinear SVM is better than the linear SVM and the LLC. We also show that our ERP-based target detection system is five-fold faster than the traditional image viewing paradigm.

1. Introduction

Brain computer interface (BCI) provides a non-muscular avenue for the user to communicate with others and to control external devices. In the past decade, there has been a tremendous amount of research performed in the highly multidisciplinary field of BCI [18]. The first convincing demonstration of a direct functional interface between a brain and a robotic arm was documented in 1999 [5]. BCI is primarily applied to restore motor control for severely disabled people, particularly those suffering from spinal cord injury, amyotrophic lateral sclerosis, stroke, or cerebral palsy. The goal of BCI is to decode a user's intents, using only brain signals, in order to control an external device. There are a variety of methods used to record brain signals that might be used in an BCI. For example, electroencephalography, electrocorticogram, magnetoencephalography, functional magnetic resonance imaging, positron emission tomography. In reality, however, electrical field recording is more practical at present and in the near future [34]. The electrophysiological recording methods include: (1) electroencephalogram (EEG), which is recorded by electrodes on the scalp, (2) electrocorticogram, which is recorded by electrodes on the cortical surface, and (3) action potentials (from a single neuron or local field potentials), which are recorded by inserting electrodes into the cortex. Notice that EEG, unlike the last two invasive options, avoids the risks of brain surgery and is one of the most popular BCI approaches.

One way to use EEG for BCI involves extracting neural signatures from the data. Neural signatures, which are called event-related potentials (ERPs), are associated with perceptual and cognitive events. ERPs have drawn a lot of attention in the field of cognitive neuroscience [21,1,25]. A successful ERP-based BCI system depends on robust ERP detection, which can be very challenging due to the low signal-to-noise ratio of an ERP (the amplitude of a typical ERP is on the order of $1-10\,\mu V$, whereas the background EEG amplitude is on the order of $100\,\mu V$). The conventional strategy is to average across trials with identical stimuli, which increases signal-to-noise ratio and makes the ERP more detectable [14,20,32]. However, the trial-averaged approach filters out much of the information about cortical dynamics and requires that each stimulus be presented multiple times, which may not feasible for real-time systems. In order to avoid this limitation, single-trial methods have been recently developed. Instead of averaging across trials, one solution for single-trial approaches is to integrate information over sensors.

Object detection is one of the many possible applications for ERP-based BCI. A number of investigators have gained valuable insights into the mechanisms of object recognition and limits of

* Corresponding author.
E-mail addresses: huang@csee.ogi.edu (Y. Huang), erdogmus@ece.neu.edu (D. Erdogmus), pavel@bme.ogi.edu (M. Pavel), Santosh.Mathan@honeywell.com (S. Mathan), k.hild@ieee.org (K.E. Hild II).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.12.025

visual temporal processing [14,31,15]. Sajda et al. [27] have recently demonstrated an object (referred to as *target*) detection system, named cortically coupled computer vision, which uses single-trial ERP detection. They proposed to detect a subject's ERPs, which are elicited by rare, attended targets, while the subject views a stream of images presented at a high rate. This presentation paradigm is known as rapid serial visual presentation (RSVP). They then use a weighted linear logistic classifier (LLC) [23] for ERP detection. As an alternative to a brain evoked response, RSVP could be combined with a behavioral response, such as a button press, for a non-BCI approach. Moreover, a button press could be fused with a ERP to bring performance benefits [10,12]. However, comparing with a button press, the ERP-based BCI approach can be used by people with motor disorders, reduces fatigue, provides a continuous (as opposed to binary) measure of confidence, has a lower latency, and has lower variance in the timing. Therefore in this work we use ERPs as the only source. Fig. 1 shows example ERP signals associated with targets and distractors for one subject at channel Cz. The bottom traces are the EEG signals averaged across trials (trial averaging is shown here for visualization purposes only). When targets are presented, one can observe a perturbation in the EEG signal with a peak at 300 ms. Furthermore, there is no such discernable pattern when distractors are presented.

The main challenges of single-trial ERP detection are the high dimensionality and the scarcity of labeled EEG data. Although the existing methods [27,19,14,2] have been successful, they have several shortcomings. In particular, the algorithm for each observer must be trained anew for each session, and the system does not benefit from adding other observers. Most of the existing methods for single-trial ERP detection are trained using within-session data. The problem with using within-session data is that we may not record enough ERPs from a single subject in one sitting to sufficiently train the classifiers, at least in part because the amplitude of the ERP reduces for closely spaced targets. The natural tradeoff is that cross-session ERPs are expected to have considerably higher variation than within-session ERPs. Recently researchers start to explore cross-session training [16] and cross-subject training [28].

Here we develop an ERP-based BCI system for visual target search using RSVP based on incremental learning and cross-session training. We propose to use incremental learning as an alternative to batch learning [13] for ERP detection. The impetus for using incremental learning is to combine additional available training examples without having to retrain classifiers from scratch to reduce the computational load and memory storage, which is critical for real-time implementations. We also describe an adaptive training method that uses cross-session data. Fig. 2 shows the framework of our ERP-based BCI system. The experimental paradigm relies on the generation of ERP(s) in the frontal cortex of the subject who is instructed to look for specific kinds of scenes or target objects. The ERP is the brain response associated with the detection of a pattern matching a predefined target. When the subjects search for the objects in the image sequences in RSVP mode, their brain responds to the presentation of a relevant target scene/object and the EEG signals can be used to detect this cognitive process. The upper portion of the figure is the classification scheme. The stages include data collection, data extraction, ERP detection and image triage based on the ERPs associated with targets or distractors. The lower portion of the figure is the training schemes, which include the support vector machine (SVM) naive training using single-session data, the SVM batch learning using cross-session data and the SVM incremental learning using only support vectors.

This paper is organized as follows. First, we show that nonlinear SVM has a better single-trial ERP detection performance than the linear SVM and the LLC on our data. Second, we demonstrate that our ERP-based target search system using within-session training has a throughput that is five times higher, in terms of square meters per unit time, than the traditional image viewing approach currently used by image analysts. Third, we demonstrate that the cross-session training approach is feasible on single-trial ERP detection problem, which could have large inter-session variances. Fourth, we introduce a novel incremental approach with less storage space and computational cost. We show that even though the incremental learning is as effective as batch learning, the memory storage is only 1/3 that of the batch learning (measured in terms of number of training samples) and the computational complexity is liner growth compared to the exponential growth of the batch learning.

2. Empirical data collection

The subjects performed target detection by clicking on a button (button presses in our experimental protocol were used to confirm targets explicitly recognized by subjects for proper data labeling) as soon as they saw a target. At the same time, we recorded their EEG signals. We used two computers to acquire data, one for image display and one for data collection.

2.1. RSVP image display paradigm

A large-scale satellite image ($27\,000 \times 6500$ pixels, representing an area of over 200 km^2) was decomposed into hundreds of smaller chips, which were labeled according to whether or not they contained a target. Each chip (500×500 pixels) represented an area of 0.09 km^2. The targets were surface-to-air missile sites found in the gray scale satellite imagery, as shown in Fig. 3. While very low level features like local textures are similar between

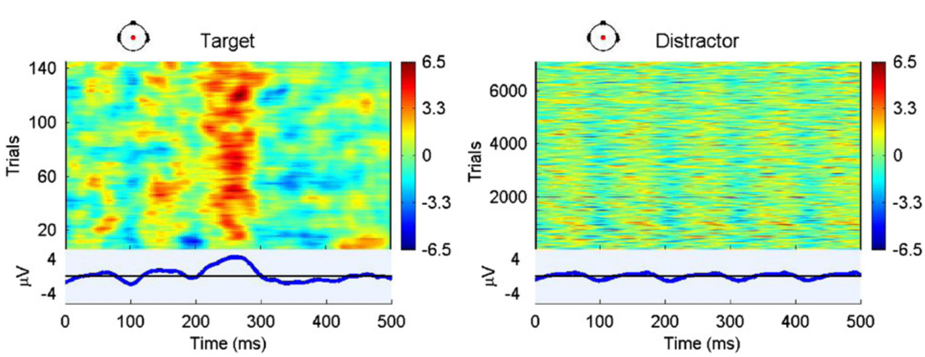

Fig. 1. Images of signals associated with targets (left) and distractors (right) respectively for one subject at channel Cz. The stimulus onset in each trial corresponds to 0 ms. The bottom traces are the EEG signals averaged over trials.