TARGET DETECTION USING INCREMENTAL LEARNING ON SINGLE-TRIAL EVOKED RESPONSE

Yonghong Huang, Deniz Erdogmus, Misha Pavel, Kenneth E. Hild *Santosh Mathan*

Oregon Health & Science University
Department of Science & Engineering
huang@csee.ogi.edu, erdogmus@ece.neu.edu,
pavel@bme.ogi.edu, k.hild@ieee.org

Honeywell Laboratories
Human Centered Systems Group
Redmond, WA 98052
Santosh.Mathan@honeywell.com

ABSTRACT

The human neural responses associated with cognitive events, referred as event related potentials (ERPs), can provide reliable inference for target image detection. Incremental learning has been widely investigated to deal with large datasets. To solve the problem of data growing over time in cross session studies, we apply an incremental learning support vector machines (SVM) method on single-trial ERP detection for identifying targets in satellite images. We implement the incremental learning SVM by keeping only the support vectors, instead of all the data, from the previous sessions and incorporating them with the data of the current session. Thus the incremental learning dramatically reduces the computational load. The results demonstrate that the incremental learning ERP detection system performs as well as the naive method, which uses only the current training session, and the batch mode, which uses all training data. Furthermore, it is more computationally efficient, which allows it to better cope with a continuous stream of EEG data.

***Index Terms*—** Brain computer interface, Event-related potential, Incremental learning, Support vector machine, Target detection

1. INTRODUCTION

Incremental learning paradigm, as opposed to the batch learning paradigm in which all training examples are provided at once for optimization, is a training mode where only a few training examples are added at a time to update model parameters. The naive learning paradigm simply uses examples from a single session to train a classifier. The motivation of incremental learning is to deal with very large training sets or non-stationary data. An important advantage of incremental learning is that it allows the algorithm to combine additional available training examples without having to retrain classifiers from scratch. This has numerous benefits, including saving a substantial amount of storage space and speeding the computation up. Therefore incremental learning algorithms have been investigated in many applications. Since the emergence of support vector machines (SVM) in the 1990s, incremental learning of SVMs has been investigated intensively [1, 2]. In [1], Syed proposed an approximate solution to the problem of incremental SVM learning. An SVM was trained on new data by discarding all previous data except the support vectors, which were combined with the new subset of data. Cauwenberghs and Poggio first proposed an exact SVM incremental algorithm [2]. It used an online recursive algorithm for SVM training and updated an optimal solution of the training one vector at a time. Compared with Cauwengerghs's work, Syed's method is more straightforward and easier to implement. Here we apply Syed's version of the incremental leaning algorithm [1] to very large training sets in the field of electroencephalography (EEG) evoked potential detection for identifying targets in satellite images. EEG evoked responses have drawn a lot of attention in the field of cognitive neuroscience. The neural responses associated with specific cognitive events are referred to as event-related potentials (ERPs). Recently researchers began to exploit human brain signals elicited by perceptual judgments as the basis for target image detection [3, 4]. They try to detect the ERP

This work was supported by DARPA under contract HM1582-05-C-0046 and by NSF under grants ECS-0524835, ECS-0622239, and IIS-0713690. The authors thank Tian Lan for helping with data collection.

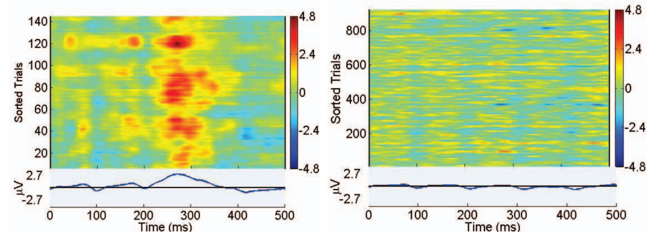

Fig. 1. Images of ERP associated with targets (left) and non-ERP associated with distractors (right) on channel Pz. Time-zero corresponds to the stimulus onset in each trial. The bottom traces are the EEG signals averaged over trials.

patterns from the EEG data that indicate whether or not an image seen briefly by a human subject contains a target. An example of ERP detection is shown in Figure 1. A more recent study indicated that the task of distinguishing the target images from distractor images (images without targets) could be achieved via single-trial ERP detection [5]. Our efforts have focused on the development of an image triage platform that uses single trial ERP detection to flag prospective targets within large image sets presented to users at rates of several images per second.

The main challenges of single-trial ERP detection are the high data dimensionality and the scarcity of labeled EEG data. Ideally, we would collect large amounts of data from each subject during a single protracted session. However, this is both monotonous and time consuming. When multiple EEG measurements are obtained from each individual at different times and possibly under changing experimental conditions, we cannot perfectly duplicate the conditions under which previous measurements were taken. Hence, there are considerable variations of the measurements from session to session. To capture the range of variation that can be expected in EEG data, we train classifiers based on the data aggregated across multiple sessions. Our previous cross-session experiments showed promising results for training the SVM in batch mode [6]. However, such batch training is computationally intensive. Hence, it is infeasible for real-time systems.

Here we apply an incremental learning SVM for cross session ERP detection. Our motivation is based on the fact that the SVM algorithm is able to summarize the data in the compact form of support vectors (SVs). The incremental learning approach (which propagates only SVs to the next training session) is compared with the naive learning method (using the current training session of data for training) and the batch learning approach (using all training data). The results show that the incremental learning approach performs better than the naive method and performs as well as the batch mode, but requires substantially less computational load than the batch method.

2. METHODS

2.1. Data Preparation

Four male graduate students ranging in age from 27 to 35 were recruited for the study. The subjects performed target detection by clicking on a button as soon as they saw a target. At the same time we recorded their EEG signals with a 32-channel Biosemi system. The sampling rate was 256Hz. The images were presented at the very high rate of 100 ms/image using the rapid serial visual presentation (RSVP) paradigm [4]. The RSVP paradigm consists of showing subjects a sequence of images. In our study each sequence contains dozens of images and at most one of these images contains a target We conducted the same data preprocessing as in [7] to extract the EEG data. The raw EEG data were segmented into task-relevant epochs. Each epoch consisted of a short segment of EEG (a window from the stimulus onset to $500ms$ after the stimulus onset). Each epoch represented the spatiotemporal electrical activity elicited in repones to a single image. The 32-channel data in each epoch were eventually concatenated to form a feature vector and the processed EEG measurements were subjected to the classifier.

To assess cross session performance, data collected at different times and under different experimental conditions were utilized. The subjects performed target detection tasks in the RSVP paradigm. Data were collected from each subject during one morning session and one afternoon session each day on five consecutive days. Each session contained 200 trials. Each trial contained around 45 images and was about 5 seconds. There were 75% of the trials containing a single target instance. Images were drawn with replacement and shown in a random order. We simulated a realistic scenario. We used only the current session as the testset and used all previous sessions as the sessions included in the current training set for the classifier. For instance, for batch learning we trained on $session\ 1(S_1)$and tested on S_2; then we trained on S_{1+2} and tested on S_3 and so on until we trained on $S_{1+2+...+9}$ and tested on S_{10}. The aggregated data were subjected to the classifier to evaluate the cross session performance.

2.2. The SVM algorithm

The SVM [8, 9] is a widely used statistical learning algorithm. The main idea of the SVM algorithm is to map input observations to a high dimensional space via kernel tricks and then optimize the decision boundary by constructing a maximum-margin hyperplane. For a classification problem, given n data samples $\mathbf{x}_i$ and class lables y_i, where $i = 1, ..., n$, the hyperplane is defined as

$$\mathbf{w}^T\mathbf{x} + b = 0, \tag{1}$$

where $\mathbf{w}$ is the normal to the hyperplane and b is a bias. The optimization problem can be formulated as the minimization of

$$f(\mathbf{w}, \xi) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i, \tag{2}$$

subject to the constraints,

$$\begin{aligned} y_i(\mathbf{x}_i\mathbf{w} + b) &\geq 1 - \xi_i \\ \xi_i \geq 0 \quad \forall i, \end{aligned} \tag{3}$$

where ξ_i are positive slack variables and the cost parameter C can be chosen by the users. A larger C is associated with assigning a higher penalty to errors. By solving a quadratic programming optimization problem, the SVM solutions are achieved. The following is the decision function,

$$f(\mathbf{x}) = sgn[\sum_{i=1}^{m} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b]. \quad (4)$$

where α_i are the Lagrangian multiplier for each data sample, $K(\mathbf{x}, \mathbf{x}_i)$ is the kernel function and m is the number of SVs. The SVs, which are the data points lying at the border of the margin, have non-zero optimal solutions for their coefficients in the final discriminant, whereas the coefficients for other data points converge to zero. Usually the SVs are only a small fraction of the original training samples. The kernel parameters, such as kernel width, σ^2, in the Gaussian kernel can be chosen by the users.

We apply a Gaussian kernel SVM on single-trial ERP detection. The inputs of the classifier are the EEG measurements and the outputs of the classifier are the likelihood values, which are used to label the EEG epochs according to whether or not they contain an ERP pattern. We adopt 10-fold cross validation on the training session to adjust two regularization parameters of the SVM: the kernel width of Gaussian kernels, σ^2, and the cost parameter, C, for each subject. We set a discrete set of the kernel size σ^2 range from 0.01 to 500 and a discrete set of the cost parameter C range from 1 to 10^6. We vary σ^2 and C over the grid formed by the selected values above. The SVM classifier is trained using the σ^2 and C giving the best validation performance.

2.3. Incremental Learning for ERP Detection

The essential property of the SVM algorithm is that only the SVs contribute to the decision boundary and the remaining training examples may be regarded as redundant. Based on this property, Syed et al. proposed an incremental learning with SVM to deal with large datasets [1]. They segmented a huge dataset into small partitions to available memory, and incrementally trained the SVM with the small partitions. Their results demonstrated that the SVs selected by the SVM was a minimal set. Any further removal of data samples significantly deteriorated the performance because the loss of SVs led to loss of vital information about the class distribution.

We apply incremental learning for cross session ERP detection. The basic idea of the incremental learning ERP detection is to train an SVM on a subset of training EEG data. The SVs found from training on each subset are preserved and combined with training samples from another data set. For the cross session EEG data in Section 2.1, there are 10 datasets, S_1 to S_{10} . Instead of training on all previous data as S_1, $S_1 \cup S_2$, ..., $S_1 \cup ... \cup S_{10}$, we preserve the SVs from the previous training sets and discard the redundant data. Let V_i represent the SVs in session i. We train using S_1, $V_1 \cup S_2$, $V_{1,2} \cup S_3$,..., $V_{1,2}$, ...,$_9 \cup S_{10}$.

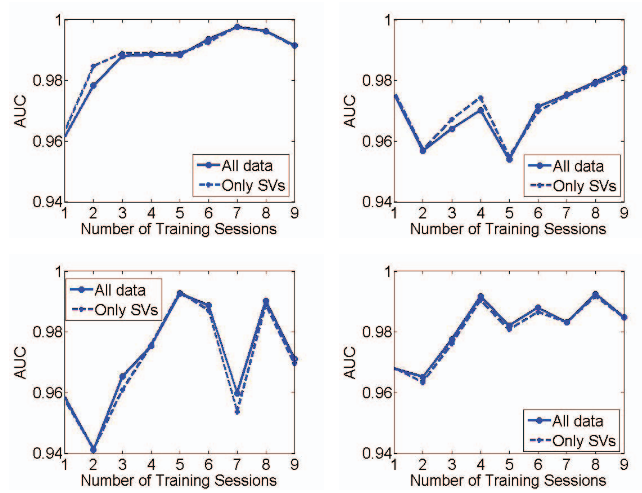

Fig. 2. The SVM test performance in terms of area under ROC (AUC) as a function of the number of training sessions for four subjects. The solid curve is the batch learning (uses all previous data for training) performance and the dash curve is the incremental learning (only the SVs are propagated) performance. The naive learning performance is the first dot for one training session.

3. RESULTS

Because the cost of missing a target in this application is extremely high, we adopt area under the receiver operating characteristic (ROC) curve (AUC) and minimum false alarm rate at zero miss (MFAR) to estimate the quantitative efficacy. The evaluation of incremental learning on single-trial ERP detection is conducted on the cross session dataset.

The ERP detection performance of the incremental learning method on the cross session data is compared with those of the batch learning mode and the naive learning mode. Figure 2 shows the cross-session ERP detection performance in terms of the AUCs for four subjects. The incremental learning SVM where only the SVs are propagated achieves similar AUCs as the batch approach for all subjects. The AUC exhibits a generally increasing trend with the inclusion of additional training data from subsequent sessions for four subjects. The incremental learning sessions have higher AUC than the naive learning using only one training session in most cases. Figure 3 shows the cross-session ERP detection performance in terms of the MFARs for four subjects. The incremental learning SVM achieves similar MFARs as the batch approach for all subjects. The MFAR exhibits a generally decreasing trend with the inclusion of additional training data from subsequent sessions for four subjects. The incremental learning sessions have lower MFAR than the naive learning in most cases. The computational load in terms of the number of training data of the incremental learning is compared with that of the batch mode. Figure 4 shows the number of train-

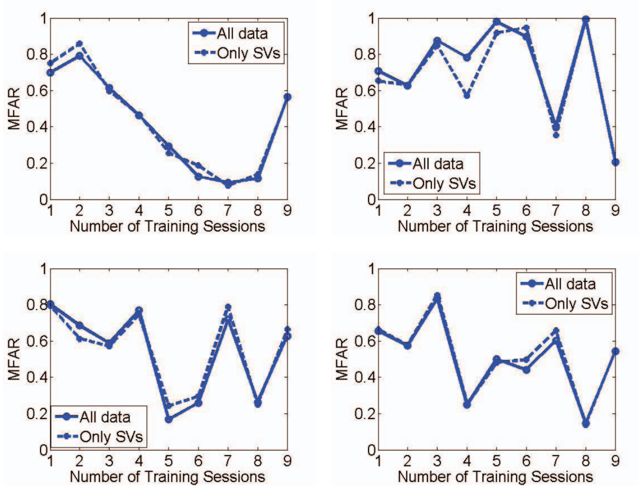

Fig. 3. The SVM test performance in terms of minimum false alarm rate at zero miss (MFAR) as a function of the number of training sessions for four subjects. The solid curve is the batch learning (uses all previous data for training) performance and the dash curve is the incremental learning (only the SVs are propagated) performance. The naive learning performance is the first dot for one training session.

ing samples for subject 1 for both incremental learning and batch learning. The other three subjects have similar results. The incremental learning using only SVs, which is a small fraction of all data, substantially decreases the computational load after the aggregated data grow over time.

4. DISCUSSION

This research uses an incremental learning method for single-trial ERP detection on the task of target image detection. The incremental learning method using only SVs performs better than the naive method and achieves a performance similar to the batch method for cross session ERP detection with substantially less computational load. Results show the feasibility of the incremental learning on the ERP-based target detection system. With more training samples, the cross session methods outperform the naive method. The results demonstrate the inter-session variances do not deteriorate the performance. The incremental learning performs as well as the batch mode due to the fact that only the SVs contribute to the decision boundary. Since the incremental learning compacts the previous training data to the SVs and then incorporates only the SVs with the new dataset, it is more computationally efficient than the batch learning method. In the future we will investigate an exact online method, based on the incremental SVM method by Cauwenberghs and Poggio [2], to construct the solution recursively, one point at a time on the single-trial ERP detection.

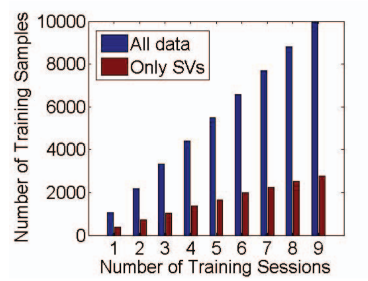

Fig. 4. The number of training samples for different number of training sessions for subject 1 using the batch learning (uses all previous data for training) and the incremental learning (only the SVs are propagated)

5. REFERENCES

[1] N.A. Syed, H. Liu, and K.K. Sung, "Incremental learning with support vector machines," in *Proc. of the Int. Joint Conf. on Artificial Intelligence*, Stockholm, Sweden, 1999.

[2] G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," in *Proc. of the Conf. on Advances in neural information processing systems*, Vancouver, Canada, 2001, pp. 409–415.

[3] J.S. Johnson and B.A.Olshausen, "Timecouse of neural signatures of object recognition," *Journal of Vision*, vol. 3, pp. 499–512, sep 2003.

[4] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, pp. 520–522, 1996.

[5] P.Sajda, A.D. Gerson, M.G. Philiastides, and L.C. Parra, *Towards Brain-Computer Interfacing*, chapter Single-trial analysis of EEG during rapid visual discrimination: Enabling cortically-coupled computer vision, MIT Press, 2007.

[6] Y. Huang, D. Erdogmus, S. Mathan, and M. Pavel, "Mixed effects models for eeg evoked response detection," in *Proc. of IEEE Workshop on Machine Learning for Signal Processing*, Cancun, Mexico, 2008.

[7] Y. Huang, D. Erdogmus, S. Mathan, and M. Pavel, "Large-scale image database triage via eeg evoked responses," in *Proc. of IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Las Vegas, NV, 2008, pp. 429–432.

[8] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[9] C. J.C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.