Determining Gender from Local Network Synchronization in the Frontal Cortex

P. Sun[1], Z. Yang[1], A. Korzekwa[1], M. T. Sutherland[1], and A. C. Tang[1]

University of New Mexico, Albuquerque NM 87106, USA
{pengsun, zhyang, akorzekw, msuther, akaysha}@unm.edu

Abstract. We investigate the possibility that one can differentiate male and female brains by a few parameters derived from the dynamics of brain activity from a local network of the frontal cortex over a small time window. With a least-square support vector machine (LS-SVM) with the SOBI-derived frontal neuronal sources power in the theta, alpha, beta, and gamma bands as inputs, we achieved a remarkable accuracy over 86% on gender classification using 5 min of resting EEG data.

1 Introduction

Numerous EEG studies have shown that scalp recorded EEG signals from the human brain differ between the males and females(e.g. [1, 2]). Given the existence of such differences, we consider the problem of determining the gender of the brain based solely on a short wave of brain signals recorded using high density EEG. In contrast to most EEG studies where sex differences were reported, we recorded EEG signals during passive visual processing, mental imagery formation, and during eyes-open and eyes-closed resting. This problem is interesting because it may shed light on how the dynamics of ongoing brain activity, independent from any goal or task, differ between the male and female brains. The differences in ongoing network dynamics may reveal gender differences at a more fundamental level of neural processing than could be revealed under specific task conditions.

We approach this problem by a combination of SOBI [3]-based feature selection method and a SVM classifier [4]. We first extracted neuronal signals from a specific region of the brain, the frontal cortex, using second-order blind identification (SOBI) as described in [5]. In the this first stage of feature selection, one specific SOBI component that corresponds to frontal source activity was selected. In the second state, for this time series of frontal source, we measured powers in four distinct frequency bands from successive 10 sec time windows and used these power values as feature dimensions of inputs to the SVM. In contrast to the automatic feature selection approach used by Schoder and colleagues [6], we constructed the features based on the unique role that the frontal lobe play in human brain function.

2 Methods and Results

Support Vector Machines. Support vector machine (SVM) is a powerful methodology for solving problems in nonlinear classification, function estimation and density estimation, which has also led to many other recent developments in kernel based methods in general [4]. For many practical problems, SVM showed superior performance over other classification algorithms, including the multilayer neural networks (for review see ref [7]). Originally, it has been introduced within the context of statistical learning theory and structural risk minimization. It guarantees the generalization ability by finding the maximum margin separating hyper-plane. We consider data

$$\{(x_1, c_1), (x_2, c_2), \ldots, (x_n, c_n)\}, \quad x_i \in \mathbb{R}^{\mathrm{p}}, \quad c_i \in (-1, 1) \ ,$$

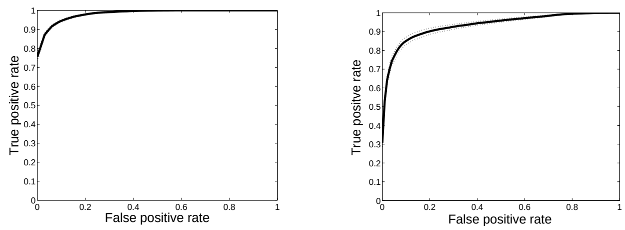
where the c_i is either 1 or -1, a constant denoting the class to which the point x_i belongs. Each x_i is a p-dimensional vector. In this paper, we applied the Least-Square Support Vector Machines (LS-SVM) with the nonlinear Radial Basis Function (RBF) as its kernel, which maps the original p-dimensional feature space to infinite dimensional space.

Data. Up to 5 min of EEG data were collected during each of the following conditions: (1) eyes-closed resting; (2) eyes-open resting; (3) video-viewing (a silently played nature video); and (4) forming mental images of scenes from the video. EEG signals were recorded using a 128-channel SynAmps EEG system (NeuroScan, El Paso, TX) with tin electrodes mounted in a custom-made cap (ElectroCap International, Eaton, OH) continuously sampled at 1000 Hz and bandpass filtered between 0.1 and 200 Hz. Then we used SOBI to extract the frontal neuronal signals from high density EEG (see refs [5]). The time series of the frontal neuronal signal from each of 16 subjects (8 males and 8 females) is truncated into short windows (10 sec window, no overlapping); for each window, the power level for each of the following four frequency bands is calculated: θ 3–7 Hz, α 8–12 Hz, β 13–20 Hz, and γ 21–35 Hz. These four power measures provide a 4–dimensional vector to feed into the SVM. Note that the inputs to the classifier is *signals from a single EEG sensor* but signals from a single inferred underlying neuronal source.

Training and Testing. We adopted the During the training phase of the SVM, 10% of the data from each of the subjects was randomly selected to make up the training set and the remaining 90% as the testing data set. This process is repeated either 50 (Exp.1) or 100 times (Exp.2) in different experiments and average and standard error of means of the accuracy over these repeated sampling will be reported in the table. A two-fold cross-validation method was used to search for the optimal regularization parameters in LS-SVM.

Classification Experiment One. With only 5 minutes of signal from the frontal source at a resting state, an accuracy of 85.8% $\pm$ 0.3% can be achieved. The area-under-ROC (AUC) is 0.99 (Fig. 1a, dashed lines indicate error bars), which is remarkable in comparison to a classification accuracy of 0.82 for an easier problem of determining whether the left or right hand was pressing a button [8]. Note that this result is obtained by using non-overlapping data which typically reduces the performance.

Classification Experiment Two. With four times of the resting data lengths and additional information from more active processing states, i.e., data from all four experimental conditions, the accuracy was only $84.5\% \pm 0.8\%$ and the AUC 0.94 (Fig. 1b). Both measures were lower than those obtained with the 5 min resting data. This decrease may have to do with an increased commonality in frontal source activity between the males and females during active mental processing. An evidence is for eyes-closed resting, while for active resting.



(a) Eyes closed Resting (b) Active Resting

Fig. 1. ROC curves: probability of correctly classifying a male sample (true positive rate) versus the probability of misclassifying a female sample (false positive rate).

Classification Experiment Three. To determine whether neuronal signals in a specific frequency band are particularly useful for gender classification, we compared classification accuracy across all sub-spaces of the 4-dimensional feature space. This experiment was performed using only data from eye-closed resting condition. Table 1 shows when single band power was used as input, classification accuracy was in the range of 61.7%–71.2%, with the highest accuracy resulting from Gamma band power. Table 2 shows that when combinations of 2 bands were used, in most cases, classification accuracy was increased to 62.6%–82.7% in comparison to its single-band accuracy. Finally, using a 3 band combination (Table 3), classification accuracy further increased to 78.9%–84.1%. It appears that higher accuracies were associated with the use of γ band and the highest accuracy of 85.8% was achieved with the use of all four bands.

3 Summary

We showed that using 10% of 5 minutes of SOBI extracted frontal source signal as training set and using only the powers of theta, alpha, beta, and gamma band as input dimensions to the SVM classifier, short 10 sec of waves of electrical signals from the frontal cortex can be classified according the gender of the brain with

Table 1. Classification accuracy as function of single frequency band (mean ± sem)

Band	θ	α	β	γ
Accuracy (%)	61.7 ± 0.2	61.9 ± 0.4	65.2 ± 0.2	71.2 ± 0.2

Table 2. Classification accuracy as function of 2 band combination.

Band	θ, α	θ, β	θ, γ	α, β	α, γ	β, γ
Accuracy (%)	62.6 ± 0.3	75.6 ± 0.2	82.7 ± 0.2	80.1 ± 0.3	82.1 ± 0.3	71.9 ± 0.2

Table 3. Classification accuracy as function of 3 band combination.

Band	θ, α, β	θ, α, γ	θ, β, γ	α, β, γ
Accuracy (%)	78.9 ± 0.2	83.6 ± 0.2	82.8 ± 0.3	84.1 ± 0.3

an accuracy as high as 86%. We found that using more EEG data collected during active cognitive processing did not lead to improvement in classification accuracy and that the gamma band power appeared to be the most important feature in classifying male and female brain signals. Note that these results are achieved using only activity from one neuronal source within the frontal cortex. It remains to be explored whether activity from other brain regions can further improve classification of gender.

Acknowledgment. Support from the Sandia National Laboratory to ACT.

References

1. Barry, R. J., Clarke, A. R., McCarthy, R., Selikowitz, M. Johnstone, S. J., Rushby, J. A.:Age and gender effects in EEG coherence: I. Developmental trends in normal children. Clinical neurophysiology **115** (2004) 2252-2258
2. volf, N. V., Razumnikova, O. M.:Gender difference in hemispheric spatiiotemporal EEG patterns upon reproduction of verbal information. Human Physiology **30** (2004) 274-280
3. Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., Moulines, E.: A blind source separation technique using second-order statistics. IEEE Trans. Signal Process. **5** (1997) 434-444
4. Vapnik, V. N.: Statistical Learning Theory. New York: Wiley, (1998)
5. Tang, A. C., sutherland, M. T., McKinney, C. J.:Validation of SOBI components from high-density EEG. NeuroImage, **25** (2005) 539-553
6. Schroder, M., Bogdan, M., Rosenstiel, W., Hinterberger, T., Birbaumer, N.: Automated EEG feature selection for brain computer interfaces. Proc. 1st International IEEE EMBS conference on neural engineering. Capri Island, Italy (2003)
7. Müller, K. R., Krauledat, M., Dornhege, G., Curio, G., Blankertz, B.:machine learning techiniques for brain-computer interfaces. Biomedical Engineering **49** (2004) 11-22
8. Parra,L., Alvino C., Tang, A., Pearlmutter, B., Yeung, N., Osman, A., Sajda, P.:Linear Spatial Integration for Single Trial Detection in Encephalography **17** (2002) 223-230