COMPARISON OF SUPERVISED AND UNSUPERVISED LINEAR METHODS FOR RECOVERING TASK-RELEVANT ACTIVITY IN EEG

An Luo, Adam Gerson, Paul Sajda

Dept. of Biomedical Engineering, Columbia University, New York, NY 10027

ABSTRACT

In this paper we compare three linear methods, Independent Component Analysis (ICA) [1], Common Spatial Patterns (CSP) [2], and Linear Discrimination (LD) [3] for recovering task relevant neural activity from high spatial density electroencephalography (EEG). Each linear method uses a different objective function to recover underlying source components by exploiting statistical structure across a large number of sensors. We test these methods using a dual-task event-related paradigm. While engaged in a primary task, subjects must detect infrequent changes in the visual display, which would be expected to evoke several well-known event-related potentials (ERPs), including the N2 and P3 [4]. We find that though each method utilizes a different objective function, they in fact yield similar components. We note that one advantage of the LD approach is that the recovered component is easily interpretable, namely it represents the component within a given time window which is most discriminating for the task, given a spatial integration of the sensors. Both ICA and CSP return multiple components, of which the most discriminating component may not be the first. Thus, for these methods, visual inspection or additional processing is required to determine the significance of these components for the task.

1. INTRODUCTION

EEG provides a high-temporal resolution imaging modality for relating brain activity to cognitive function. More recently, high-spatial density EEG (≥ 64 sensors) has been used to improve the spatial resolution of the modality, as well as provide additional observations for solving the inverse-problem associated with dipole localization. The large number of sensors also enables various array processing methods to be employed.

As the spatial-temporal data acquired from EEG are high dimensional, both non-linear and linear methods are often used for analysis and dimensionality reduction. Compared with non-linear methods, linear methods are less susceptible to over-fitting, have a significantly lower computational cost and are consistent with the linearity of volume conduction.

Currently, linear methods for analyzing multi-channel EEG can be categorized as either supervised or unsupervised. Traditionally implemented as an unsupervised method, ICA [1] decomposes signals into several independent components that have their own time series and are viewed and analyzed separately. Unsupervised methods do not take into account "ground-truth" labels associated with experimental events, such as stimulus type. Supervised methods, on the other hand, exploit training labels given knowledge of the task and/or subject responses. CSP [2] is an example of supervised source recovery widely used in EEG. The method weights electrodes according to the power captured for two classes of data. It finds orientations in the sensor space in which the power is simultaneously maximized for one class and minimized for the other. An alternative to maximizing power is to maximize the discrimination between the two classes. Parra et al. [3] proposed an LD method that spatially integrates sensor values in well-defined temporal windows to recover sources that maximally discriminate two classes given labeled EEG.

In this paper we compare these supervised and unsupervised linear methods, applying the algorithms to EEG data acquired during a two-task visual detection experiment. We show that the three methods recover similar components, though in some cases one method (LD) is preferable given interpretability and stability of the results.

2. EXPERIMENT AND DATA ACQUISITION

2.1 Task description

Our experiment consists of two tasks. Subjects detect task-relevant visual changes in a secondary task while engaged in a complex primary task. Both tasks are visual in nature (they occur within a simple video game) and are event driven. The visual-motor nature of the task with rather infrequent secondary task events predicts N2 and P3b ERP activity [4] locked to changes detected for the secondary task.

2.2 Subjects and data acquisition

Two right-handed subjects (one male and one female) completed the study. A 64-channel EEG system was used to recording. The sampling rate was 1000Hz. Eye-blink and eye-movement activities are recorded separately so that these artifacts could be removed from the EEG. Events were recorded and data were locked to visual "change" events (in this case color changes in the stimulus). Subject responses were also recorded (button press or not). Trials are labeled based on whether or not the subject detected a change. Trials are labeled as either those in which change was detected by the subject or trials in which change was not detected (but occurred). Thus each trial is associated with a binary label.

3. DATA ANALYSIS

3.1 Preprocessing

After EEG is acquired, preprocessing is applied to the raw data. Eye-blink and horizontal and vertical eye motion artifacts are removed by using principle component analysis [5]. Low frequency drifts, 60Hz noise, and 120Hz harmonic noise was removed by filtering.

3.2. Linear methods

All three methods linearly transform the observed signals as,

$$Y = WX \, , \tag{1}$$

where X are the observations (the original EEG signal matrix), W is the transform matrix (or vector) that is calculated using the different linear approaches, and Y is the resulting source matrix (or vector) representing the recovered sources. Note that (1) represents the recovery of sources Y given an underlying linear mixture of sources observed in X.

3.2.1 ICA

ICA is widely used for solving the blind source separation problem. Spatial filters derived using ICA blindly separate the input EEG data into a sum of temporally independent and spatially fixed components arising from distinct or overlapping brain or extra-brain sources [1].

In ICA, X in Equation (1) corresponds to an $N \times T \times R$ matrix, where N is the number of recording electrodes, T is the number of time points and R is the number of trials. The trials are appended together and then unwrapped after computing the "unmixing matrix". The unmixing matrix W **,** estimated by ICA, linearly decouples the multi-channel EEG data, with the columns of W^{-1} giving the relative projection strengths of the respective components at each of the electrodes. The rows of the output matrix Y are time courses of the ICA components.

3.2.2 CSP

CSP is based on the decomposition of the EEG into spatial patterns using simultaneous diagonalization of two covariance matrices [2]. These patterns maximize the difference of the power between the two populations.

Essentially, optimal spatial filters are determined through joint diagonalization of two covariance matrices derived from each task related class (in this case detected change vs. undetected change). The normalized covariance matrix of each single trial NxT matrix X, where N is the number of channels and T is the number of samples, is determined as $C = XX'/\mathrm{trace}(XX')$. The average of covariance matrices from class 1 (C_1) and class 2 (C_2) trials are then summed to produce a composite covariance matrix $C_c = C_1 + C_2$. The eigenvectors and eigenvalues of this spatial covariance matrix yield a whitening transformation $P = (\lambda_c)^{-1/2}U_c'$ where $C_c = U_c\lambda_cU_c'$. Transforming the average covariance matrices corresponding to the two classes, $S_1 = PC_1P'$ and $S_2 = PC_2P'$, assures that S_1 and S_2 share common eigenvectors such that $S_1 = B\lambda_1B'$ and $S_2 = B\lambda_2B'$ where $\lambda_1 + \lambda_2 = I$. The first and last eigenvectors of B then represent optimal projections associated with class 1 and class 2 respectively. The projection matrix is then defined as $W = (B'P)'$ and an EEG trial is transformed as $Y = WX$.

As with ICA, raw EEG data is represented as the $N \times T$ matrix X . Y corresponds to the decomposition (mapping) of a trial. The variances of Y are used for of two populations of single-trial EEG recordings**.** The columns of W^{-1} are the common spatial patterns and can be seen as time-invariant EEG source distribution vectors.

3.2.3 LD

LD is also a supervised method and can be used to compute the optimal spatial integration of a large array of sensors for discrimination between two classes [3].

In this case in Equation (1), X is a 2-D $N \times (M \times t)$ matrix representing M trials ($M = I + J$, I trials for class 1 and J trials for class 2) of EEG data at t time points, and the number of recording electrodes is N . W is the spatial weighting coefficient vector which is calculated by logistic regression such that the $M \times t$ dimensional vector Y represents a hyperplane maximally separating the two classes, which in this case, is the detection of change versus not detecting change. . Timing information is exploited by discriminating and averaging within a short time window relative to a given external event (in this case the change in the stimulus).

4. RESULTS

4.1 ICA

After they are acquired and preprocessed, EEG data are decomposed using ICA tools provided by EEGLAB [6]. The first 10 components that are locked to the onset of visual change in the secondary task are presented in Figure 1. The time course of the 8th component is shown in Figure 2. In Figure 2, the onset of the change is locked at 0ms. The ordinate corresponds to different trials.

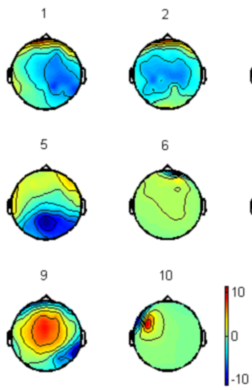

Figure 1. Scalp projections for first 10 components recovered using ICA for subject 1 in response to color changes. Ordering is based on the fraction of variance captured by each component.

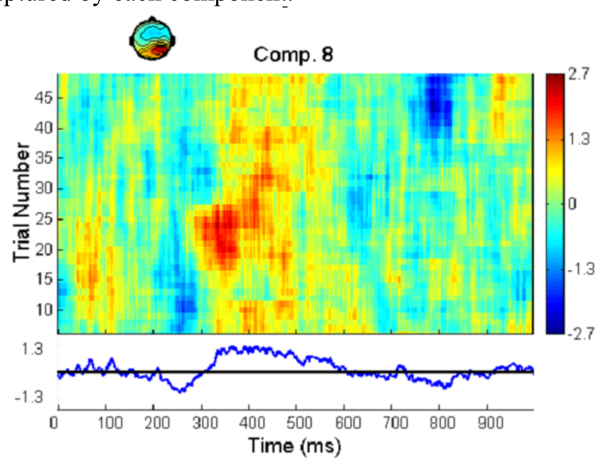

Figure 2. Time course of component 8 recovered using ICA. The plot on the bottom shows an average across trials. Note negativity beginning around 200msec and positivity around 300msec.

4.2 CSP

Figures 3 and 4 show results using CSP. We see that despite the sign and magnitude difference (due to a difference in the weighting coefficients in the transform matrix), component 8 recovered via ICA and pattern 1 recovered via CSP are quite similar in terms of both scalp projection and time course.

In addition, the two components found with ICA (component 8) and CSP (component 1) correlate well the expected N2 and P3b activity, which appear in the trial averaged EEG during stimulus classification and target processing [7].

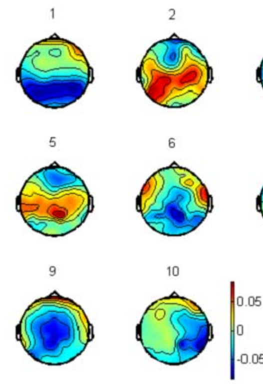

Figure 3. Scalp projections for first 10 components recovered using CSP for subject 1 in response to color changes. Ordering is based on the fraction of variance captured by each component.

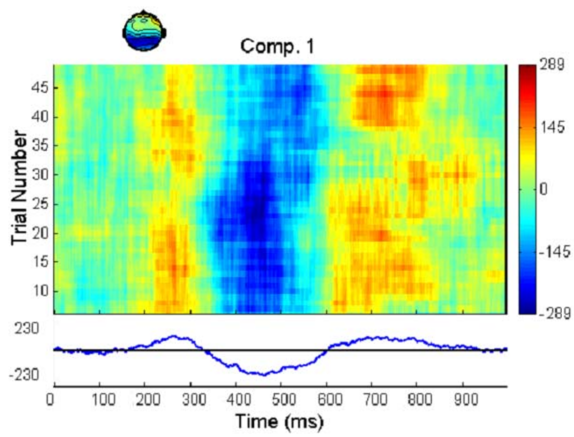

Figure 4. Time course for pattern 1 recovered using CSP. Note positivity beginning around 200msec and negativity around 300msec (the sign difference in both the scalp projection and component time course cannot be unambiguously recovered. Thus this is equivalent to component 8 recovered using ICA).

4.3 LD

Since LD computes a single linearly discriminating orientation in sensor space, a single component is recovered which may be interpreted as most discriminating component between two classes. In our analysis, the training window starts at 400ms after the onset of the change and spans 100ms. The recovered component is displayed in Figure 5 with the inset showing its scalp projection. From the LD result, we see that this component is very similar to the 8th component recovered via ICA and the 1st pattern in CSP.

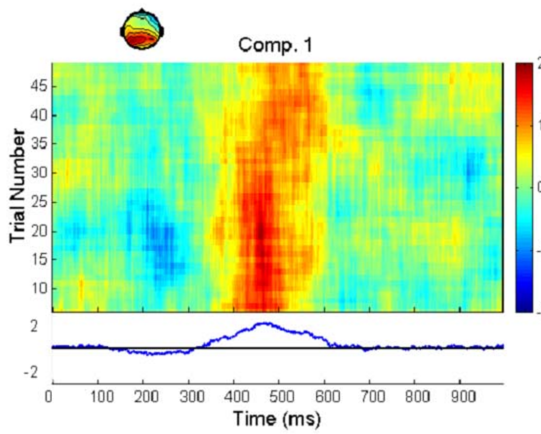

Figure 5. Time course and scalp projection of discriminating source recovered via training in a 100msec window starting at 400sec. Note negativity beginning around 200msec and positivity around 300msec.

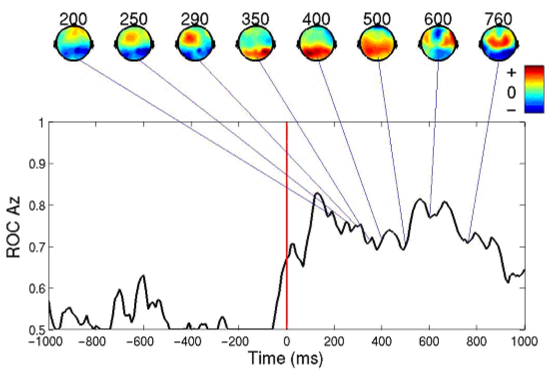

Figure 6. Scalp projections and level of discrimination (plot as Az) for subject 1 computed using LD trained for different temporal windows.

LD results using different temporal windows for discrimination are shown in Figure 6. Results are illustrated using leave-one-out Receiver Operator Characteristic (ROC) analysis [8] to evaluate the performance of the discriminator. Abscissa shows the time index relative to when the change occurs (locked at 0ms), and ordinate shows the area under the ROC curve (Az). We see that the LD approach recovers different discriminating activity, depending on the location of the training window. For example there is a clear change of discriminating activity from posterior to central regions as the window progresses from 300-800msec. This is likely related to the strong motor response that occurs roughly around 650 msec (average response time is 688ms for this subject).

5. DISCUSSION AND CONCLUSIONS

ICA, CSP and LD are all methods used for EEG analysis that apply a linear transformation to the original sensor data. ICA computes the transform matrix so that the data are separated into a set of temporally independent and spatially fixed components; CSP decomposes signals into spatial patterns that maximize the power differences between the two populations; LD computes optimal spatial integration of a large array of sensors during a time window so that the two populations of data are maximally discriminated.

From the results presented we see that similar components are recovered by ICA and CSP. However the ordering of these components is different. However, LD yields a single component that is also similar to components recovered with the other approaches. For both ICA and CSP the optimal discriminating component over a specific temporal window is not ranked first, therefore visual inspection or additional pattern classification is required to identify it. Also important to note is that CSP and LD are stable in that they consistently obtain the same result from the same data. Sensitive to initial random seeds of starting weights, data presentation order, and learning rates, ICA may yield components with different scalp maps, time courses, and ranking across training runs.

REFERENCES

[1] T. Jung, S. Makeig, etc., “Analysis and Visualization of Single-Trial Event-Related Potentials,” *Human Brain Mapping*, 14: pp. 166-85, 2001.

[2] H. Ramoser, J. Muller-Gerking, etc., “Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement,” *IEEE Trans Rehabil Eng.,* 8(4): pp. 447-56, Dec. 2000.

[3] L. Parra, C. Alvino, et.al., “Linear Spatial Integration for Single-Trial Detection in Encephalography,” *NeuroImage*, 17: pp. 223–30, 2002.

[4] D. Friedman, Y.M. Cycowics, H. Gaeta, “The Novelty P3: an Event-Related Brain Potential (ERP) Sign of the Brain’s Evaluation of Novelty”, *Neuroscience and Biobehavioral Reviews*, 25: pp. 355-73, 2001.

[5] L. Parra, C. Spence, A. Gerson and P. Sajda, “Response Error Correction - A Demonstration of Improved Human-Machine Performance Using Real-Time EEG Monitoring,” *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, 11(2): pp. 173-7, 2003.

[6] A. Delorme, S. Makeig, “EEGLAB: an Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis,” *Journal of Neuroscience Methods*, in press.

[7] M.H. Giard, F. Peronnet, “Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study,” *J. of cognitive neuroscience*, 11: pp. 473-90, 1999.

[8] J.A. Swets “ROC Analysis Applied to the Evaluation of Medical Imaging Techniques,” *Investigative Radiology* 14:109-121, 1979.