Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

A Supervised Sparse Learning Framework to Solve EEG Inverse Problem for Discriminative Activations Pattern

Feng Liu
Department of Industrial, Manufacturing & Systems Engineering,
University of Texas at Arlington, Arlington, TX 76019, USA

Abstract

Electroencephalography (EEG) is one of the most important noninvasive neuroimaging tools that provides excellent temporal accuracy. As the EEG electrode sensors measure electrical potentials on the scalp instead of direct measuring activities of brain voxels deep inside the head, many approaches are proposed to infer the activated brain regions due to its significance in neuroscience research and clinical application. However, since mostly part of the brain activity is composed of the spontaneous neural activities or non-task related activations, task related activation patterns will be corrupted in strong background signal/noises. In our research, we proposed a sparse learning framework for solving EEG inverse problem which aims to explicitly extract the discriminative sources for different cognitive tasks by fusing the label information into the inverse model. The proposed framework is capable of estimation the discriminative brain sources under given different brain states where traditional inverse methods failed. We introduced two models, one is formulated as supervised sparse dictionary learning and the other one is the graph regularized discriminative source estimation model to promote the consistency within same class. Preliminary experimental results also validated that the proposed sparse learning framework is effective to discover the discriminative task-related brain activation sources, which shows the potential to advance the high resolution EEG source analysis for real-time non-invasive brain imaging research.

Background and Motivation

To infer brain cortex activations from the scalp recorded EEG signals belongs to the class of *inverse problem*. Precise localization of neuronal activity inside the brain can offer an insightful understanding of how brain is functioning given certain cognitive and motion tasks. According to previous research (Raichle 2006), more than 80% of the brain energy is devoted to the non-task related energy. To extract discriminative sources given different brain status is extremely difficult and not well explored. In this research, we aim to calculate extract discriminative sources to facilitate the understanding of brain mechanism under different cognitive tasks or different neurological disorders by incorporating a simple linear classifier which can be interpreted as discriminative filters for different brain patterns.

To the best of our knowledge, there is no literature addressing simultaneously estimation of brain sources and distinguishing different sources under different brain status. We propose a new supervised formulation of the inverse problem and with efficient algorithms to solve it. The new formulation is composed of two ingredients, source reconstruction and supervised source classification. The contributions of this paper is threefold, including: **(a)** First proposed a supervised inverse model with discriminative capability by leveraging the label information; **(b)** First describe the EEG inverse problem as an overcomplete dictionary learning problem and show the opportunities of using algorithms from compressive sensing and computer vision community; **(c)** Propose efficient algorithms to solve the optimization model good accuracy.

Problem formulation

The electromagnetic field measured by EEG can be described as the following optimization model:

$$\arg\min_{S} \|X - LS\|_F^2 + \lambda\Theta(S) \tag{1}$$

where $X \in \mathbb{R}^{N_c \times N_t}$ is the EEG data measured at a set of N_c electrodes for N_t time points, $L \in \mathbb{R}^{N_c \times N_d}$ is the lead field matrix which maps the source signal to sensors on the scalp, each column of L represents the activation pattern of a particular source to the EEG electrodes, $S \in \mathbb{R}^{N_d \times N_t}$ represents the corresponding driving potential in N_d sources locations for all the N_t time points.The penalty function $\Theta(S)$ is to discourage unnecessary complicated source configurations and enforces neurophysiologically plausible solution. In our research, we propose a source reconstruction formulation fused with classification information and the label information is coded in H, the goal function is given below:

$$\begin{aligned}\langle W, S\rangle &= \arg\min_{W,S} \|X - LS\|_F^2 + \beta\|H - WS\|_F^2 \\ &+ \lambda\|W\|_F^2 \; s.t.\; \forall i, \;\; \|s_i\|_0 \leqslant T\end{aligned} \tag{2}$$

where $H \in \mathbb{R}^{m \times N_t}$ is the label matrix where non-zero entry in each column denotes the corresponding class, there are m classes corresponding m different brain status. W is the linear classifier parameter to be trained. For Eqn.2, it can be rewritten as

$$\begin{aligned}\langle W, S\rangle &= \arg\min_{W,S} \left\| \begin{pmatrix} X \\ \sqrt{\beta}H \end{pmatrix} - \begin{pmatrix} L \\ \sqrt{\beta}W \end{pmatrix} S \right\|_F^2 \\ &+\lambda\|W\|_F^2 \qquad s.t.\; \forall i, \;\; \|s_i\|_0 \leqslant T\end{aligned} \tag{3}$$

We show that the problem can be regarded a dictionary learning problem and We employed a reviesed version of DK-SVD algorithm to solve Problem.3.
Due to the fact that true sources are usually corrupted with noises, we come up with a graph regularized model to promote in-class consistency and out-class discrimination.

$$\langle W, S\rangle = \arg\min_{S} \|X - LS\|_F^2 + \gamma\|S\|_{1,1}$$
$$+\frac{\beta}{2}\sum_{i,j=1}^{N} \|s_i - s_j\|_2^2 M_{ij} + \alpha\sum_{i=1}^{N}\|h_i - Ws_i\|_2^2, \tag{4}$$

where M_{ij} is defined to be 1 if s_i and s_j belongs to the same class, and -1 otherwise. G is the Laplacian matrix of M, which is defined as $G = D - M$ and $D_{ii} = \sum_j M_{ij}$.

$$\langle W, S\rangle = \arg\min_{W,S} \|X - LS\|_F^2 + \gamma\|S\|_{1,1} + \beta(Tr(S^T GS))$$
$$+ \alpha\sum_{i=1}^{N}\|h_i - Ws_i\|_2^2 \tag{5}$$

The above problem 5 can be solved with feature-sign search algorithm.

Preliminary Result

Numerical simulations were conducted given different SNR. We compared our proposed framework with two different baseline methods, namely Efficient Projections onto the ℓ1-Ball (EP-ℓ1B) (Liu and Ye 2009), MxNE (Gramfort, Kowalski, and Hämäläinen 2012).

We used three different accuracy criteria to measure the reconstructed source accuracy. The first one is perfect reconstruction accuracy (PRA), which compare the calculated source location and the exact ground true. The second measurement is to use Baillet-Garnero's reconstruction accuracy (BRA) criteria (Baillet and Garnero 1997). The third measurement is to use the criteria proposed in (Haufe and Ewald 2016) denoted as Haufe reconstruction accuracy (HRA), which is to measure whether the reconstructed source is located in the ROI. To make the solution be more informative, a sparse solution is always preferred for its interpretability. The averaged number non-zero entries (NZE) in the solution is also included to measure the sparsity. $\|X - LS\|_F$ is the reconstruction error (RE). Fig.1 illustrates the EEG potentials before and after the application of our method, distinctive source activation patterns can be clearly retracted.

Table 1: Performance comparison at SNR=1.2

Method	Time	PRA	BRA	HRA	NZE	RE
DKSVD	**2.04**	**0.77**	**0.81**	**0.93**	**4.00**	**0.91**
$\ell_1 SR$	10.9	0.33	0.37	0.52	308	118
MxNE	10.6	0.22	0.25	0.50	449	85.6

Conclusion and Future Work

We aim to reconstruct discriminative sources given different brain status. A label guided dictionary learning formulation was given for the first time with ℓ_0-norm and is solved using our revised version of DK-SVD algorithm. Through numerical simulations, we showed that in terms of accuracy and speed, our method is better than the ℓ_1 or ℓ_2 related ones. The reason is high coherence of lead field matrix and sparsity constraints is easy to fail, which we shows in details in our future coming research. The classification component trained a W matrix with each row corresponding certain type of brain status, which is physically meaningful, we termed as discriminative filter. Our proposed framework can achieve satisfactory result compared to traditional methods and can be extended to more specific priors such as spatially smoothness requirement or depth compensation requirement.

Table 2: Performance comparison at SNR=0.5

Method	Time	PRA	BRA	HRA	NZE	RE
DKSVD	**2.38**	**0.63**	**0.64**	**0.68**	**4.0**	**12.4**
$\ell_1 SR$	11.4	0.30	0.32	0.50	410	111
MxNE	12.0	0.25	0.28	0.50	507	89.6

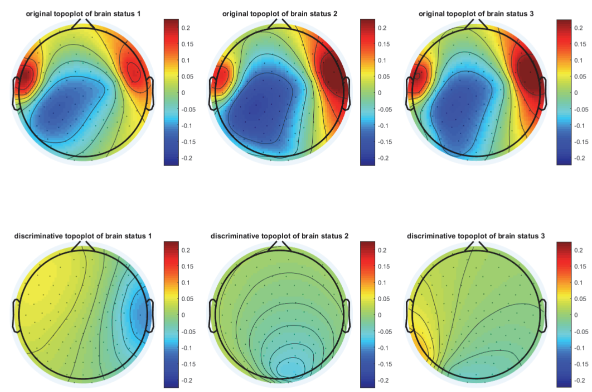

Figure 1: Discriminative filtered topoplots for 3 different brain status: the top 3 topoplots is cooresponding to 3 different brain tasks with high background noise or resting state potentials, it's very hard to distinguish them. Below is the topoplots after we applied our methodology to extract the discriminative expression for different brain tasks.

References

Baillet, S., and Garnero, L. 1997. A bayesian approach to introducing anatomo-functional priors in the eeg/meg inverse problem. *IEEE Transactions on Biomedical Engineering* 44(5):374–385.

Gramfort, A.; Kowalski, M.; and Hämäläinen, M. 2012. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Physics in Medicine and Biology* 57(7):1937.

Haufe, S., and Ewald, A. 2016. A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies. *Brain topography* 1–18.

Liu, J., and Ye, J. 2009. Efficient euclidean projections in linear time. In *Proceedings of the 26th Annual International Conference on Machine Learning*, 657–664.

Raichle, M. E. 2006. The brain's dark energy. *Science* 314(5803):1249–1250.