Graph Regularized EEG Source Mapping with in-Class Consistency and Out-Class Discrimination

Author names removed due to blind review policy

No Institute Given

Abstract. Electroencephalography (EEG) is one of the most promising noninvasive neuroimaging tools that provides high 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 is proposed to infer the activated brain sources due to its significance in neuroscience research and clinical applications. However, since mostly part of the brain activity is composed of the spontaneous neural activities or non-task related activations, true task relevant activation sources will be corrupted in strong background signal/noises. For decades, the EEG inverse problem was solved in an unsupervised way without any utilization of the label information that comes from different brain states (e.g. happiness, sadness, and surprise). We argue that by leveraging label information, the task related discriminative sources can be much better retrieved. A novel model for solving EEG inverse problem called Laplacian Graph Regularized Discriminative Source Reconstruction (LGRDSR) which aims to explicitly extract the discriminative sources by implicitly coding the label information into the graph regularization term. The proposed model is capable of estimation the discriminative brain sources under given different brain states and reduce the variation within classes simultaneously. The proposed model is general and can be extended with different neurophysiological assumptions or prior. Simulated result show the effectiveness of our proposed framework.

Keywords: EEG, Inverse Problem, Discriminative Source, Graph Regularization

1 Introduction

Due to its low cost, easy portability, high temporal resolution and no exposure to radioligands, electroencephalography (EEG) has become one of the most popular brain imaging tools. Compared to other techniques such as positron emission topography (PET) and functional magnetic resonance imaging (fMRI), EEG is a direct measurement of real time electrical neural activities, and EEG is more suitable to answer exactly *when* different brain modules are activated and hence in what processing steps each module is involved [9]. PET and fMRI cannot be used to assess rapidly varying neuronal activity due to the slow response of metabolism [6]. Successful applications of EEG can be found in several clinical environments, such as real time monitoring of patients' sleep apnea[14], detection and prediction of epilepsy seizures [2].

To infer the brain source from the recorded EEG is termed as *inverse problem*. In order to solve the inverse problem, different priors or assumptions can be made. The most traditionally used priors in EEG source reconstruction are based on the ℓ_2 norm, leading to what is known as the minimum norm (MN) inverse solver [5]. This MN inverse solver leads to a minimum norm estimates (MNE) of the sources. However, ℓ_2-based solvers suffer from several limitations, e.g. the solution inside the brain will be diffuse for the overestimation of activated sources. Other assumption priors are also presented with different methodologies, such as, Multiple Signal classifier (MUSIC) and Recursively applied and projected MUSIC (RAP MUSIC) [10][11] which adopted spatio-temporal independent topographies (IT) model with recursive subspace projection; low resolution brain electromagnetic tomography (LORETA) [12] and stardardized LORETA [13] which enforces spatial smoothness of the source located on neighboring voxels, focal under determined system solution (FOCUSS) [3] which combines the advantages of distributed dipole modeling method and linear estimation method by allowing current sources to take arbitrary shape with high resolution;

As summarized above, based on different assumptions or prior knowledge, different algorithms solving the inverse problem were proposed. Usually, when we designed a sequence of experiments to record the EEG and asked the subjects to perform different psychological tasks within certain time windows, the label information on the recorded EEG data can be easily obtained. Traditional algorithms solve the EEG inverse problem independently without leveraging the label information, that will make it hard to compare the reconstructed sources under different brain states due to its low SNR(Signal-to-noise ratio) of the brain. EEG signal is highly non-stationary and contains much noises, to explicitly extract factual sources and promote the in-class consistency and out-class discrimination, we implemented the graph regularized version of discriminative source reconstruction, and tested on simulated data and real EEG data.

The contributions of this paper is listed in the following: (1) We proposed to use label information to solve the EEG inverse problem in a supervised way. (2) A graph regularized EEG inverse model is presented that can promote in-class consistency and out-class discrimination. (3) A Voting Orthogonal Matching pursuit algorithm is given to decompose the common sources motivated by the "cross-and-bouquet" model [16].

2 The Inverse Problem

Under the quasi-static approximation of Maxwell's equations, the measured EEG signal X can be described as the following linear function of current sources S:

$$X = LS + E \tag{1}$$

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 brain 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 neural potential in N_d sources locations for all the N_t time points. $E \in \mathbb{R}^{N_c \times N_t}$ is additive noise. L is a matrix with number of columns far much greater than the number of rows, making the inverse problem ill-posed. an estimate of S can be found by minimizing the following cost function, which is composed of a quadratic error and a regularization term:

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

The regularization term $\Theta(S)$ is to discourage complicated source configurations temporally or spatially and enforces neurophysiologically plausible solutions, and $\|\cdot\|_F$ is the Frobenius Norm. The regularization term take the form of ℓ_2, ℓ_1 or mixed norm, spatially smooth formulation as in LORETA estimation or spatially sparse formulation with least absolute shrinkage and selection operator estimate. For example, to restrict the total number of activated sources less than a scalar T, the following ℓ_0-Norm formulation can be used:

$$\arg\min_{S} \|X - LS\|_F^2 \;\; s.t. \; \|s_i\|_0 \leqslant T, \tag{3}$$

As is well know that ℓ_0-norm is the best intuitive formulation to restrict number of activated sources, almost of neuroscience researchers, if not all, for solving EEG inverse problem use approximated norm such as ℓ_1 to avoid the problem being NP-hard. For the ith time point, the ℓ_1 regularized formulation is given below:

$$s_i = s^*(x_i, L) = \arg\min_{x} \|x_i - Ls_i\|_2^2 + \gamma \|s_i\|_1 \tag{4}$$

The ill-posed problem of Eq.1 originates from the fact thathe L matrix as a dictionary, and each column in L is an atom of the dictionary. Given the EEG recordings at a time point, which is denoted as ith column x_i of X matrix, we want to represent the signal with minimum error by trying to find the best linear representation from activation patterns (atoms) in the over-complete dictionary L. The solution s_i is the sparse coding for the x_i in the dictionary L, the non-zero entries in s_i corresponding to a column in the dictionary matrix L represent the activated regions inside the brain.

3 Proposed Framework

3.1 Discriminative Source Reconstruction with Graph Regularization

Due to the fact that EEG signal is non-stationary and typically the SNR is very low, it's important to get consistent inverse solution given under same brain status and discriminative solutions under different brain status. Inspired by the successful applications of graph regularization in computer vision community [4][1], the proposed model in form of discriminative dictionary learning is described, which is termed as Laplacian Graph Regularized Discriminative Source Reconstruction (LGRDSR), and comprises the source reconstruction fidelity term and label guided in-class consistency and out-class discrimination term:

$$\langle S \rangle = \arg\min_{S} \|X - LS\|_F^2 + \gamma \|S\|_{1,1} + \frac{\beta}{2} \sum_{i,j=1}^{N} \|s_i - s_j\|_2^2 M_{ij}, \tag{5}$$

where the first term is the fidelity term, the second term is the cost of sparse coding, $\|\cdot\|_{1,1}$ is the ℓ_1 norm notation for a matrix, equal to the sum of the absolute value of all elements in a matrix. The third term is the graph regularization term that requires all the sparse coder within the same category remains similar pattern while making the sparse representation for different class distinct from each other. The definition of M matrix can be written as:

$$M_{ij} = \begin{cases} +1, \text{ if } (s_i,\text{s}_j) \text{ belong to the same class} \\ -1, \text{ if } (s_i,\text{s}_j) \text{ belong to different classes} \end{cases}$$

The goal of this formulation is to find discriminative sources while maintaining the robustness of in-class reconstructed sources.
Remarks on *design of M matrix*
When (s_i, s_j) belong to the same class, design the value of M_{ij} to be positive will add penalty when the intrinsic geometry (s_i, s_j) is different, thus promoting in-class consistency of the source. When (s_i, s_j) belong to different classes, design the value of M_{ij} to be negative will explicitly promote out-class discrimination of the source. In practice, if we care more about in-class consistency, we can set $M_{ij} = 0$ when (s_i, s_j) belongs to different classes. The magnitude of M_{ij} can also be adjusted to tailor the relative weight between in-class consistency and out-class discrimination.

Define D as a diagonal matrix whose entries are column or row sums of the symmetric matrix M, $D_{ii} = \sum_j M_{ij}$, define $G = D - M$, G is called graph Laplacian [1], The third term of Eq.5 can be further derived in the following way:

$$\sum_{i,j=1}^{N} \|s_i - s_j\|_2^2 M_{ij} = \sum_{i,j=1}^{N} ({s_i}^T s_i + {s_j}^T s_j - 2{s_i}^T s_j) M_{ij} = 2tr(S^T G S) \tag{6}$$

As a result, Eq.5 is further derived as:

$$\langle S \rangle = \arg\min_{S} \|X - LS\|_F^2 + \gamma \|S\|_{1,1} + \beta(Tr(S^T G S)) \tag{7}$$

Eq.7 can be solving using feature-sign search algorithm [1][8], which converted the nondifferentiable ℓ_1 norm problem to a ℓ_1-regularized problem which better convergence rates.

3.2 Common Sources Decomposition with Voting Othogonal Matching Pursue(VOMP)

Under the assumption of strong common spontaneous source activation pattern, the contribution of discriminative sources to the EEG recorded data is relatively small, making the solution space for different classes highly correlated, which is hard to solve. In order to address that, we use the idea of a useful step is to decompose of X, which is very helpful in the reconstruction of discriminative sources. Similar procedure The Voting Othogonal Matching Pursue (OMP) is proposed. The aim is

Algorithm 1 Decomposition of Non-discriminative Sources with Voting OMP (VOMP)

INPUT: Lead field matrix L, preprocessed EEG signal matrix X, label matrix H, Prior knowledge of maximum number of common activated sources T_{max}, Minimum Voting Acceptance percentage p
OUTPUT: S_c, result of removed common sources X_r
Initialization: $T \leftarrow 1$, $\Omega = \emptyset$, $R = X$, $R_{new} = X$, $S' = 0$

```
while Stopping criteria is not met do
    for i ∈ 1, ..., N_t do
        s_i ← OMP(L, x_i, 1)                ▷ Orthogonal Matching Pursuit with only 1 iteration
        q_i ← nonzero index of s_i          ▷ q is used to denote nonzero element index
    end for
    q_best ← most frequent q_i              ▷ q_best is the voted nonzero index
    if T = T_max or frequency of f(q_best) < p then
        break;
    else
        Ω ← Ω ∪ q_best ; L' = (L_{:,i} | i ∈ Ω) ; S' ← pinv(L')X
        S' ← mean(S')                       ▷ make the common source consistent across all classes
        R_new ← X − L'S'
    end if
    for k ∈ 1, ..., C do                    ▷ Loop for all the classes
        R^k_new = {R_new(i) | i ∈ class k} ; R^k = {R(i) | i ∈ class k}
    end for
    if ||R^k_new|| < ||R^k|| for k ∈ 1, ..., C then
        continue;                           ▷ Continue if and only if the residual for all classes are decreasing
    else break;
    end if
    T ← T + 1; R ← R_new
end while
X_r = R_new; S_c = S'
return S_c, X_r
```

to recover the common sources across all classes. As we assumed, the strong spontaneous source activation pattern is very strong, thus making the convex hull spanned by all the source configuration to a tiny portion of the space [16]. We propose the denoising model and an very efficient algorithm to solve it. We used orthogonal matching pursuit algorithm to solve it.

$$\langle S_c \rangle = \arg\min_{W_c} \|X - LS_c\|_F^2 \quad s.t. \ \|s_i\|_0 \leqslant T_{max}, \quad i = 1, 2, ...N_d \tag{8}$$

Algorithm 2 Proposed framework of solving Problem 7

INPUT: Lead field matrix L, preprocessed EEG signal matrix X, label matrix H
OUTPUT: Discriminative source S_d
Initialization: $T \leftarrow 1$, $\Omega = \emptyset$, $R = X$, $R_{new} = X$, $S' = 0$

while Stoping criteria not met **do**
(1) Using VOMP algorithm for common source decomposition;
(2) Solve the following sparse coding problem for $< s(i) >= \arg\min_{s(i)} L(s_i) + \gamma \|s_i\|_1$ using the feature-sign search algorithm [8] ;
(3) Adjust the voting threshold p;
end while

4 Numerical Result

We used a recently developed realistic head model called ICBM-NY or "New York Head" [7] which is based on highly detailed standardized finite element model (FEM) of the non-linear averaged anatomical template-ICBM152. In the simulation experiments, we designed common sources

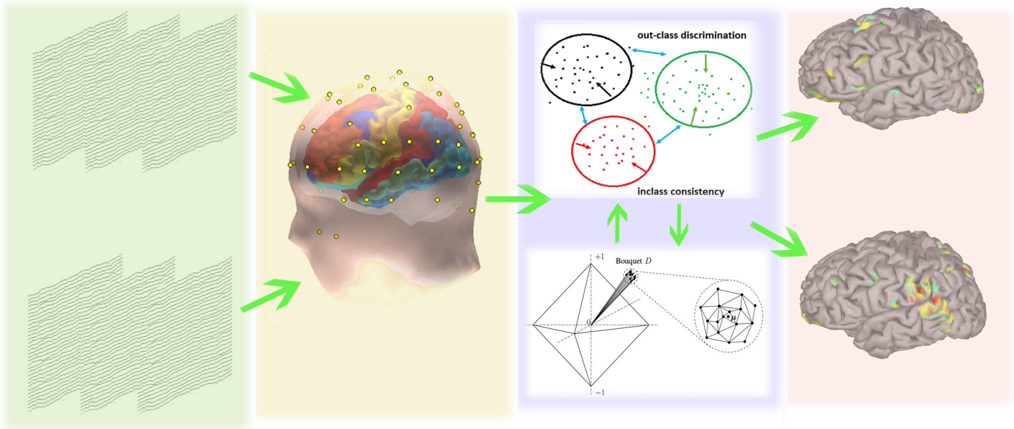

Fig. 1: Procedures of our framework: After gathering labeled EEG recorded data, the brain model is constructed using finite element method (BEM) based on MRI images, then we first use the VOMP algorithm to decompose the primary common source starting with a high minimum voting percentage, and then solve it using feature-sign search algorithm, do these steps alternative until converged, last step is to map discriminative sources to the cortex.

which has much higher magnitude and three discriminative sources under three assumed brain conditions with smaller magnitude compared to common sources. The dimension of lead field matrix is 108×2004, representing 108 channels and 2004 voxels. Like the authors in Ref.[15] estimating source when source signal is around the peak, we designed 2 neighbor voxels to be the spontaneous common sources with a magnitude of 0.8 with standard deviation to be 0.1 and another 2 neighbor source as the secondary source with magnitude to be 0.2 with standard deviation of 0.05. We designed three classes, all the three class share the same common sources and have their own discriminative source. We sampled 200 time points for each class, and did the experiment 5 time to get the average accuracy of the reconstructed source. For the LGRDSR parameter, we set β to be 0.05 and γ to be 0.1; The noise matrix is designed to effect the EEG recording together with the true source signal. For each time point, 3 random voxels are corrupted randomly, with controlled magnitude randomly within a region.The SNR is calculated as $SNR = 20 \log \frac{\|S\|}{\|N\|}$. The reconstruction performance of the proposed method as well as the benchmark methods is given in Table 1.

The 8 benchmark method includes ElasticNet, Homotopy, DALM, PDIPA, FISTA, sLORETA, WMN. The former 6 can be referred to Ref.[17]. In table 1, time is recorded with unit of second (s), PSA represent primary source (spontaneous common source) accuracy, which is to measure the capability of each algorithm to reconstruct the common sources. All the values except the Time column in the table represents the distance in (mm) of ground true source to the reconstructed source calculated using shortest path algorithm. F means failed when the reconstructed location is on a different hemisphere. EC1 represents error for class 1, which is to calculate the distance of the reconstructed discriminative source with the ground true. EC2 and EC3 are similarly defined. To illstrate the effect of the proposed framework, the ground truth of the activated pattern is given in Fig.2, with the reconstructed source by WMN, sLORETA and our method given in Fig.3–5. We can see from Table 1 and the Fig.3–5 that when the SNR is large, all the algorithms performs well in reconstructing primary source, as for the discriminative sources for different classes, our method can achieve perfect reconstruction performance. All other algorithms' except sLORETA and WMN performance are also acceptable when SNR is large. When we increase the noise, all the algorithms can still achieve high accuracy in finding the primary source. For the discriminative source, our algorithm perform much better. We also validated that, to solve a pure ℓ_1 EEG inverse problem, Homotopy algorithm performs better than other algorithms in the EEG inverse problem, which is in accordance with Ref.[17].

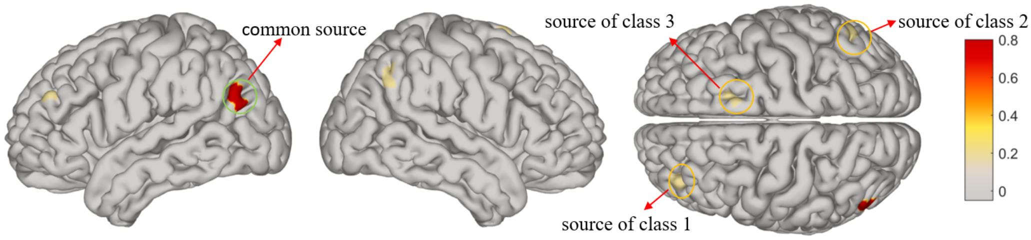

Fig. 2: Ground truth for all 3 classes

Table 1: Reconstruction Accuracy Summary

	SNR = 16.5					SNR = 24.5				
Methods	Time	PSA	EC1	EC2	EC3	Time	PSE	EC1	EC2	EC3
ElasticNet	0.011	6.96	F	F	F	0.01	6.96	F	F	F
Homotopy	0.56	0.004	124.0	86.1	129.1	0.19	0	3.02	3.29	3.43
DALM	0.30	0.36	97.9	69.8	158.5	0.29	0	3.29	4.22	3.90
PDIPA	1.78	0.37	117.1	74.9	135.0	1.07	0	4.53	4.21	4.45
L1LS	4.09	0.42	123.3	72.8	78.8	136.6	0	3.13	3.01	3.87
FISTA	2.63	10.18	138.5	75.9	192.0	2.63	0	8.64	9.15	8.30
sLORETA	0.055	1.71	F	F	F	0.54	0	F	F	F
WMN	0.0004	8.43	F	F	F	0.0004	8.71	F	F	F
LGRDSR	0.28	0	4.33	3.97	4.01	0.26	0	0	0	0

5 Conclusion

In this paper, we proposed to use label information to retrieve discriminative sources corresponding to each different brain status. Although solving sparse representation with graph regularization in computer vision and compressive sensing community is nothing new, its application in EEG inverse problem that implicitly using label information has never been proposed. Our model is presented with a Laplacian graph regularized term that can boost the in-class similarity and discourage the out-class similarity, thus making the source solution from the same class more robust to the noise. We bring up the idea of cross-and-bouquet in the inverse problem and present an efficient algorithm to address the highly coherence of the reconstructed signals given high background spontaneous source signals. We illustrated that current algorithms can't find the discriminative sources and the superior of our algorithm given certain amount of noises.

References

1. Cai, D., He, X., Han, J., Huang, T.S.: Graph regularized nonnegative matrix factorization for data representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(8), 1548–1560 (2011)
2. Chen, G.: Automatic EEG seizure detection using dual-tree complex wavelet-fourier features. Expert Systems with Applications 41(5), 2391–2394 (2014)
3. Gorodnitsky, I.F., George, J.S., Rao, B.D.: Neuromagnetic source imaging with focuss: a recursive weighted minimum norm algorithm. Electroencephalography and clinical Neurophysiology 95(4), 231–251 (1995)
4. Guo, H., Jiang, Z., Davis, L.S.: Discriminative dictionary learning with pairwise constraints. In: Asian Conference on Computer Vision. pp. 328–342. Springer (2012)
5. Hämäläinen, M.S., Ilmoniemi, R.J.: Interpreting magnetic fields of the brain: minimum norm estimates. Medical & biological engineering & computing 32(1), 35–42 (1994)
6. Haufe, S., Nikulin, V.V., Ziehe, A., Müller, K.R., Nolte, G.: Combining sparsity and rotational invariance in EEG/MEG source reconstruction. NeuroImage 42(2), 726–738 (2008)

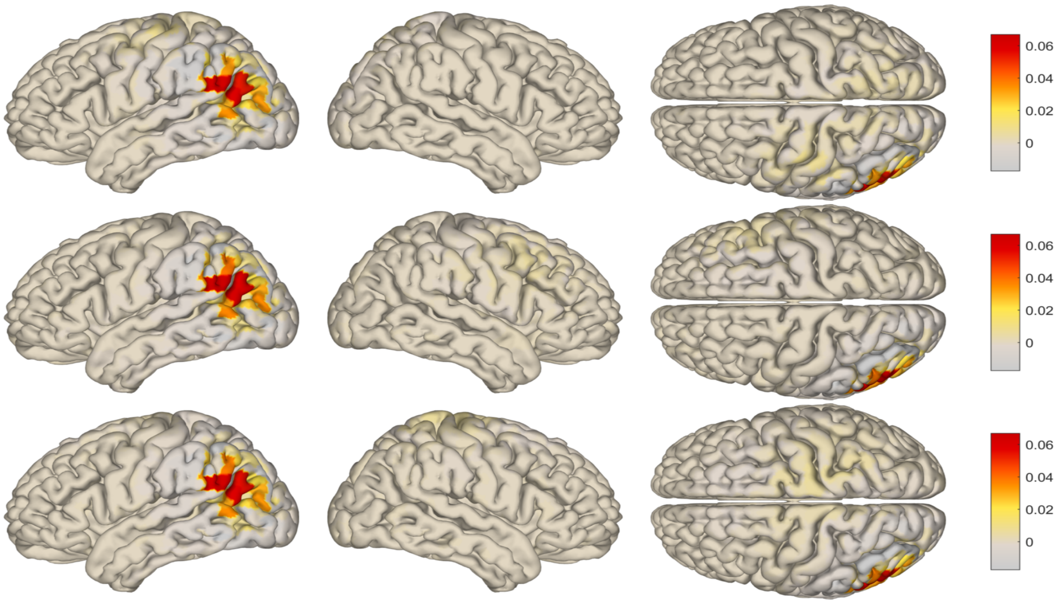

Fig. 3: WMN solution: The above row is the WMN solution for class 1; Class 2 and class 3 is illustrated in the middle and bottom row. The solution WMN gives is not sparse, with too many spurious sources of small magnitude.

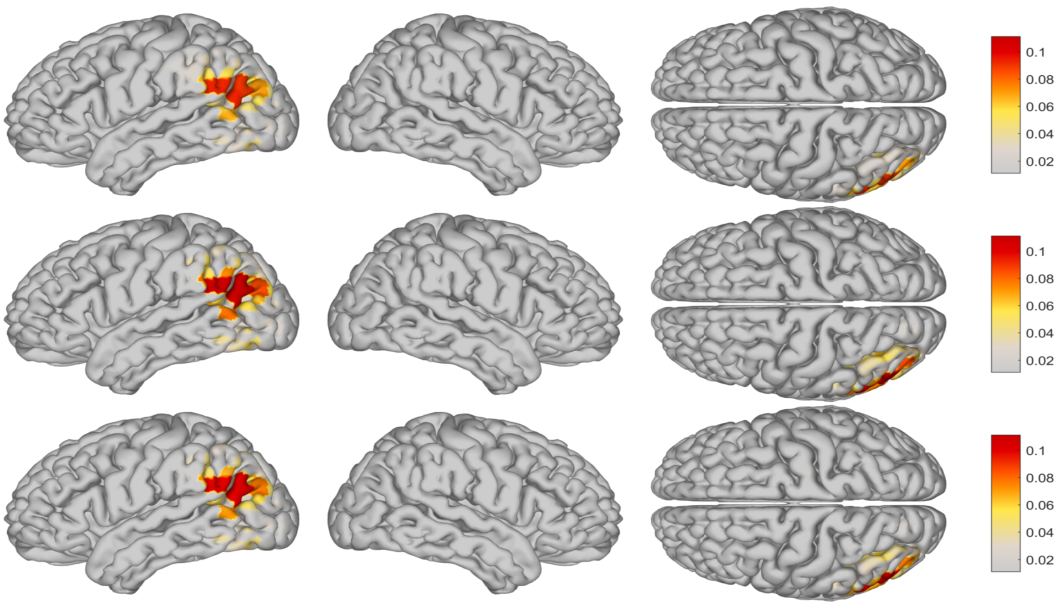

Fig. 4: sLORETA inverse solution: The above row is the sLORETA solution for class 1; Class 2 and class 3 is illustrated in the middle and bottom row. sLORETA can successfully reconstruct the primary source, however the secondary source is not successfully reconstructed. Compared to the solution of WMN, sLORETA can suppress the numerous spurious sources with small magnitude.

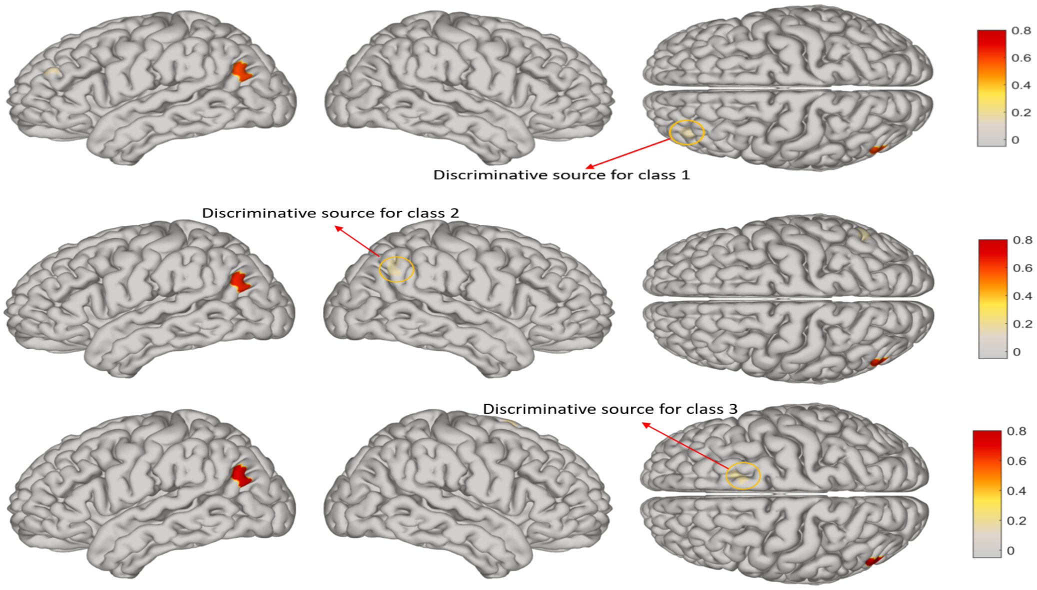

Fig. 5: LGRDSR reconstructed source: The reconstruction solutions for 3 classes are given in each row. As can be seen from the illustration, the discriminative source can be successfully reconstructed compared to other methodologies.

7. Huang, Y., Parra, L.C., Haufe, S.: The new york head -a precise standardized volume conductor model for EEG source localization and tes targeting. NeuroImage 140, 150 – 162 (2016), transcranial electric stimulation (tES) and Neuroimaging
8. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: Advances in neural information processing systems. pp. 801–808 (2006)
9. Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., de Peralta, R.G.: EEG source imaging. Clinical neurophysiology 115(10), 2195–2222 (2004)
10. Mosher, J.C., Leahy, R.M.: Recursive music: a framework for EEG and meg source localization. IEEE Transactions on Biomedical Engineering 45(11), 1342–1354 (1998)
11. Mosher, J.C., Leahy, R.M.: Source localization using recursively applied and projected (rap) music. IEEE Transactions on signal processing 47(2), 332–340 (1999)
12. Pascual-Marqui, R.D., Lehmann, D., Koenig, T., Kochi, K., Merlo, M.C., Hell, D., Koukkou, M.: Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia. Psychiatry Research: Neuroimaging 90(3), 169–179 (1999)
13. Pascual-Marqui, R.D., et al.: Standardized low-resolution brain electromagnetic tomography (sloreta): technical details. Methods Find Exp Clin Pharmacol 24(Suppl D), 5–12 (2002)
14. Perrier, J., Clochon, P., Bertran, F., Couque, C., Bulla, J., Denise, P., Bocca, M.L.: Specific EEG sleep pattern in the prefrontal cortex in primary insomnia. PloS one 10(1), e0116864 (2015)
15. Sohrabpour, A., Lu, Y., Worrell, G., He, B.: Imaging brain source extent from eeg/meg by means of an iteratively reweighted edge sparsity minimization (ires) strategy. NeuroImage 142, 27–42 (2016)
16. Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T.S., Yan, S.: Sparse representation for computer vision and pattern recognition. Proceedings of the IEEE 98(6), 1031–1044 (2010)
17. Yang, A.Y., Sastry, S.S., Ganesh, A., Ma, Y.: Fast 1-minimization algorithms and an application in robust face recognition: A review. In: Image Processing (ICIP), 2010 17th IEEE International Conference on. pp. 1849–1852. IEEE (2010)