HHS Public Access

Author manuscript

Mach Learn Med Imaging. Author manuscript; available in PMC 2018 June 13.

Published in final edited form as:

Mach Learn Med Imaging. 2017 ; 2017: 316–324.

Novel Effective Connectivity Network Inference for MCI Identification

Yang Li[1], **Hao Yang**[1], **Ke Li**[2], **Pew-Thian Yap**[3], **Minjeong Kim**[3], **Chong-Yaw Wee**[4], and **Dinggang Shen**[3]

[1]School of Automation Sciences and Electrical Engineering, Beihang University, Beijing, China

[2]School of Aeronautic Science and Engineering, Beihang University, Beijing, China

[3]Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, USA

[4]Department of Biomedical Engineering, National University of Singapore, Singapore, Singapore

Abstract

Inferring effective brain connectivity network is a challenging task owing to perplexing noise effects, the curse of dimensionality, and inter-subject variability. However, most existing network inference methods are based on correlation analysis and consider the datum points individually, revealing limited information of the neuron interactions and ignoring the relations amongst the derivatives of the data. Hence, we proposed a novel ultra group-constrained sparse linear regression model for effective connectivity inference. This model utilizes not only the discrepancy between observed signals and the model prediction, but also the discrepancy between the associated weak derivatives of the observed and the model signals for a more accurate effective connectivity inference. What's more, a group constraint is applied to minimize the inter-subject variability and the proposed modeling was validated on a mild cognitive impairment dataset with superior results achieved.

1 Introduction

Mild cognitive impairment is considered as the clinical stage between normal aging and dementia. MCI patients suffer from a cognitive decline that does not interfere notably with activities of daily living. Anatomical and physiological researches suggest that cognitive process is greatly associated with the interactions among distributed brain regions [1].

Constructing functional and effective brain connectivity from neuroimaging data holds great promise for understanding the functional interactions between brain activities. Recently, many connectivity modeling approaches based on functional magnetic resonance imaging (fMRI) have been proposed and employed disease for identification, e.g., Alzheimer's disease (AD) and MCI from normal controls (NCs).

Recent works demonstrated that sparse learning techniques provide excellent performances in a series of neuroimaging applications [2, 3]. The use of certain sparsity connectivity

Correspondence to: Hao Yang.

modeling can elucidate robust connections from a set of noisy connections and increase the discriminative power for disease diagnosis. Lee et al. [2] adopted a least absolute shrinkage and selection operation (Lasso) to construct the functional connectivity network. Wee et al. [3] introduced Group Lasso based connectivity network by adopting a $l_{2,1}$ regularizer to the original Lasso. Both methods achieved a relatively high accuracy in disease classification. However, these methods consider only the datum points (brain regions in our case) individually, ignoring the inter-datum connections which are represented by the derivatives of the signals. The absence of interconnections information may lead to overfitting problems in effective connectivity network modelling.

To address this issue, in this paper, we presented a novel sparse linear regression model to infer effective connectivity network and used it for accurate identification of MCI patients from NCs. Specifically, ultra-fMRI time series were first generated by concatenating the original fMRI signal and its corresponding weak derivatives. The structure of effective connectivity network was then determined using an ultra-group Lasso method. Based on this structure, an ultra-orthogonal forward regression (UOFR) algorithm was employed to estimate the strength of each effective connection. The proposed method was applied for MCI identification and superior classification performance was achieved.

2 Materials and Methodology

2.1 Data Acquisition and Preprocessing

The present study involved 61 participants (28 MCI patients and 32 controls) who were diagnosed based on a battery of general neurological examination, collateral and subject symptom, neuropsychological assessment evaluation, and functional capacity reports. Data acquisition was performed using a 3 T Siemens TRIO scanner. One-hundred and eighty resting-state fMRI volumes of each participant were collected with the following parameters: TR = 3000 ms, TE = 30 ms, acquisition matrix = 74 74, 45 slices, and voxel thickness = 3 mm.

The preprocessing pipeline including slice time correction, head motion correction, spatial smoothing, and template wrapping was performed using Statistical Parametric Mapping 8 (SPM8) software package. It should be noted that nuisance signals were band-pass filtered within frequency interval [0.01 ≤ f ≤ 0.08 Hz]. The mean fMRI time series of each region-of-interest according to AAL atlas was then computed for each subject by averaging the fMRI time series over all voxels in each ROI.

2.2 Network Structure Detection via Ultra-group Lasso

Suppose there are M ROIs and N subjects, the mean time series of m-th ROI for n-th subject can be represented as $y_m^n = [y_m^{(n)}(1); y_m^n(2); \ldots; y_m^n(T)]$ with T being the number of time points in the time series. For each ROI, its mean time series y_m^n can be modeled the linear combination of time courses of other ROIs as

$$y_m^n = A_m^n \theta_m^n + e_m^n \quad (1)$$

where $A_m^n = [y_1^n, \ldots, y_{m-1}^n, y_{m+1}^n, \ldots, y_M^n]$ denotes a data matrix that includes all mean time series except for the m-th ROI, θ_m^n and e_m^n denote the weight vector and noise.

For a dynamic system, the datum points are time dependent and are connected between each other through the derivatives of time continuous functions. These interconnections convey many important characteristics of a dynamic system. However, the standard least squares criterion considers the datum points individually, discarding the connections among them. The absence of the information conveyed by these interconnections may lead to overfitting problems of dynamic systems [4]. To address this issue, an ultra-least squares (ULS) criterion was introduced by incorporating the weak derivatives into the least squares criterion. The weak derivatives $D^l y$ is defined in Sobolev space as

$$D^l y(t) = \int_{[0,T]} y(t) \mathrm{D}^l \varphi(t) dt = (-1)^l \int_{[0,T]} \varphi(t) \mathrm{D}^l y(t) dt \quad (2)$$

for any test function $\varphi(t) \in C_0^\infty([0,T])$, which is smooth and possesses compact support on [0, T] [4]. In this study, the (m + 1)-th order B-spline functions were employed as the modulating functions. Then, the derivatives $D^l y$ can be redefined as

$$D^l y(\tau) = \int_{[\tau, T_0+\tau]} y(t) \mathrm{D}^l \varphi(t-\tau) dt,\ \tau \in [0, T-T_0] \quad (3)$$

Based on the description above, the ULS criterion is defined as

$$J_{ULS} = \left\| y_m^n - A_m^n \theta_m^n \right\|_2^2 + \sum_{l=1}^k \left\| D^l y_m^n - (D^l A_m^n) \theta_m^n \right\|_2^2 \quad (4)$$

where $D^l A_m^n = [D^l y_1^n, \ldots, D^l y_{m-1}^n, D^l y_{m+1}^n, \ldots, D^l y_M^n]$.

By concatenating the original signals and the weak derivatives together as $\tilde{y}_m^n = [y_m^n; D^1 y_m^n; D^2 y_m^n; \cdots; D^k y_m^n]$, $\tilde{A}_m^n = [A_m^n; D^1 A_m^n; D^2 A_m^n; \cdots; D^k A_m^n]$, the ULS criterion can be rewritten as

$$J_{ULS} = \left\| \tilde{y}_m^n - \tilde{A}_m^n \theta_m^n \right\|_2^2 \quad (5)$$

which has the same form as the standard least squares criterion.

By applying the ULS criterion to the Lasso algorithm, we introduce the ultra-Lasso as

$$J_{Ultra-Lasso} = \left\| \tilde{y}_m^n - \tilde{A}_m^n \theta_m^n \right\|_2^2 + \lambda \left| \theta_m^n \right| \quad (6)$$

where $\lambda > 0$ is the regularization parameter controlling the 'sparsity' of the model, $\tilde{y}_m^n$ is the ultra-version of the target signal, $\tilde{A}_m^n$ represents the set of ultra-regressors and θ_m^n is the regression parameter, respectively.

However, since the sparsity constraint in Lasso is applied at an individual level, the nonzero elements in θ_m^n differ across subjects. This inevitably causes inter-subject variability which may influence further group analysis. To minimize the inter-subject variability and gain the same model structure for multiple subjects, a group constraint [3] was imposed into ultra-Lasso as

$$J_{UGL} = \sum_{j=1}^{n} \left\| \tilde{y}_m^j - \tilde{A}_m^j \theta_m^n \right\|_2^2 + \lambda \left\| \theta_m \right\|_{2,1} \quad (7)$$

where $\theta_m = [\theta_m^1, \theta_m^2, \cdots, \theta_m^N]$ and $||\theta_m||_{2,1}$ is the summation of l_2-norms of row vectors in θ_m. Specifically, the weights corresponding to certain parameters across different subjects are grouped together. This promotes a common connection topology, while in the meantime allows for variation of coefficient values between subjects. By employing the ultra-group Lasso, a subset of region-of-interests (ROIs) with nonzero weights is selected to be considered connecting with the target ROI.

2.3 Effective Connectivity Construction via UFOR

The coefficients θ_m estimated via the ultra-group Lasso can simply be regarded as the effective connectivity (connection weights) between ROI to construct an effective connectivity network. However, these estimated coefficients are unscaled and biased, and it may lead to difficult interpretation and analysis of the effective network. Thus, an UOFR algorithm [4] was employed to estimate the connectivity strength based on the structure detected by the ultra-group Lasso. Given the ultra-target signal $\tilde{y}_m^n$ and candidate regressor dictionary $D_m^n = \{\tilde{x}_i \mid \tilde{x}_i \in \tilde{A}_m^n, \theta_m^n(i) \neq 0\}$, where $\tilde{x}_i = [y_i^n; D^1 y_i^n; D^2 y_i^n; \cdots; D^k y_i^n]$ represents

ultra-time series of ROIs selected via the ultra-group Lasso, the values of error reduction ratio (ERR) can be computed as

$$ERR_i = \frac{\langle \tilde{x}_i, \tilde{y}_m^n \rangle^2}{\langle \tilde{x}_i, \tilde{x}_i \rangle \langle \tilde{y}_m^n, \tilde{y}_m^n \rangle} \quad (8)$$

Regressor (defined as $\tilde{x}_{max}$) with the greatest ERR was first removed from the dictionary D_m^n, and was regarded as the connection weight between the corresponding ROI and the target ROI. The remaining regressors in D_m^n were then orthogonalized with $\tilde{x}_{max}$ using a Gram-Schmidt algorithm. This process was repeated until the regressor dictionary D_m^n becomes empty. All the maximum ERR values were arranged into an effective connectivity matrix of size $M \times M$ for M-dimensional ROIs, where the matrix contains every possible effective connectivity of ROIs pairs [5].

2.4 Feature Selection and Classification

To characterize the brain networks with a small number of neurobiologically meaningful and easily computable measures, topological properties, such as out-degree and in-degree, weighted-clustering, betweenness centrality, were extracted as features from the connectivity matrix following [6]. In order to ensure that all features were within the same scale and to minimize bias in feature selection, the feature vectors are scaled to the range [0, 1] individually across subjects.

After features extraction, a feature selection method based on the importance scores from a standard random forest was adopted. The importance scores have been shown to select a highly reduced subset of discriminative features and was detailed in [7] and we briefly outline it here. At each node τ within the binary trees T of the random forest, the Gini impurity measures how well a potential split is separating the samples of the two classes in this particular node [7]. It is defined and calculated as

$$G(\tau) = 1 - p_1^2 - p_0^2 \quad (9)$$

where $p_k = \frac{n_k}{n}$ denotes the fraction of the n_k samples from class k = {0, 1} out of the total n samples at node τ. By splitting and sending the samples to two sub-nodes τ_l and τ_r (with respective samples fractions $p_l = \frac{n_l}{n}$ and $p_r = \frac{n_r}{n}$) with a threshold t_θ on variable θ, the Gini coefficient decreases $\Delta G(\tau)$ is calculated as

$$\Delta G(\tau) = G(\tau) - p_l G(\tau_l) - p_r G(\tau_r) \quad (10)$$

By searching over all variables θ and all possible thresholds t_θ, the maximum $\Delta G(\tau)$ is determined. Individually for all variables θ, the decrease in Gini impurity is recorded and accumulated for all nodes τ and all trees T in the forest:

$$I_G(\theta) = \sum_T \sum_\tau \Delta G_\theta(\tau, T) \quad (11)$$

The Gini importance I_G indicates the frequency a particular feature is selected for a split and the discriminative power of this feature for the classification problem. Based on this criterion, features ranked and selected prior to the training of a classifier. The Gini importance criterion has shown robustness against noise and effectiveness in selecting useful features [7].

Finally, a linear SVM was trained for MCI classification using the features selected based on the Gini importance. A nested leave-one-out cross-validation (LOOCV) scheme was adopted in this study to evaluate the classification performance.

3 Experiment Results

3.1 Classification Performance

We compared our proposed method with several other related methods for connectivity network based MCI classification in Table 1. Experiment results demonstrate that, by the use of ultra least criterion and the group Lasso, the proposed method models the relationship among brain regions more accurately and achieves much improved performance in identifying MCI subjects from NC. It indicates excellent diagnostic power of proposed classification framework and also validates the effectiveness of the modeling method.

By comparing results of methods with the ultra-least criterion and those without, we find that methods with ultra-least criterion obtained 3.28% accuracy increase. This implies that the ultra-least criterion efficiently increases the noise resistibility and robustness of the method by incorporating the weak derivatives into the least squares criterion and thus helps improve the classification performance afterwards. It should also be noted that the Group Lasso method achieves an accuracy of 77.05% which is 3.28% higher than ordinary Lasso. By grouping together the weights corresponding to certain features across different subjects, the group constraint reduces inter-subject variability and thus enables relatively easier differentiation between MCI subjects and healthy controls [3].

3.2 Brain Regions Involved in Classification

As a LOO strategy is employed to evaluate the proposed method, features selected at each loop might be quite different. To evaluate the importance of brain regions, the frequency that features being selected are counted and features with highest selected times are considered

to be significant in the classification of MCI. As each ROI corresponds to several features, the selected times of features corresponding to the same ROI is added up. The top ten ROIs and their locations are listed in Table 2.

It worth noticing that these top 10 regions locate in the frontal lobe (e.g. Superior frontal gyrus, orbital part) and the temporal lobe (e.g. Superior temporal gyrus). And regions such as hippocampus are also found to be associated with MCI/AD diagnosis. These results are exemplified in previous literatures [8–10].

As topological properties are extracted as features, brain regions are evaluated by selected frequency in classification at a nodal level. To further evaluate the significant differences of each connection, a standard two-sample t-test was employed on full dataset. Connections with a p-value smaller than 0.01 are listed in Table 3.

Figure 1 graphically shows the differences of these connections between MCI and NC (the thickness of edges indicates the strength of connections). It is interesting to note that the effective connectivity for most of the optimal connections are much smaller in MCI patients than that of NCs. Moreover, several ROIs (Hippocampus, Superior frontal gyrus, orbital part; Angular gyrus) selected in classification showed significant differences between two groups based on the two-sample t-test, further justify their contribution to MCI pathology.

4 Discussion and Conclusion

In summary, we have proposed a novel sparse effective connectivity network estimation method by imposing a group constraint and an ULS criterion into ordinary Lasso. We use an ultra-group Lasso to detect the network structure and re-estimate the connectivity strength via an UOFR algorithm. The network structure detection process prunes the ROI candidates, thus reducing the parameters of the model to prevent overfitting problems in effective network modelling. Promising experiment results demonstrated the efficacy of the proposed approach for MCI classification.

References

1. Sporns O. Towards network substrates of brain disorders. Brain. 2014; 137:2117–2118. DOI: 10.1093/brain/awu148 [PubMed: 25057132]
2. Lee HLD, Kang H, Kim BN, Chung MK. Sparse brain network recovery under compressed sensing. IEEE Trans Med Imaging. 2011; 30(5):1154–1165. [PubMed: 21478072]
3. Wee CY, Yap PT, Zhang D, Wang L, Shen D. Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. Brain Struct Funct. 2014; 219(2):641–656. DOI: 10.1007/s00429-013-0524-8 [PubMed: 23468090]
4. Li Y, Cui WG, Guo YZ, Huang T, Yang XF, Wei HL. Time-varying system identification using an ultra-orthogonal forward regression and multiwavelet basis functions with applications to EEG. IEEE Trans Neural Netw Learn Syst. 2017; (99):1–13. DOI: 10.1109/TNNLS.2017.2709910
5. Li Y, Wee CY, Jie B, Peng ZW, Shen DG. Sparse multivariate autoregressive modeling for mild cognitive impairment classification. Neuroinformatics. 2014; 12(3):455–469. DOI: 10.1007/s12021-014-9221-x [PubMed: 24595922]
6. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. Neuroimage. 2010; 52(3):1059–1069. DOI: 10.1016/j.neuroimage.2009.10.003 [PubMed: 19819337]

7. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, Hamprecht FA. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics. 2009; 10

8. Nir TM, Jahanshad N, Villalon-Reina JE, Toga AW, Jack CR, Weiner MW, Thompson PM. Alzheimer's Disease Neuroimaging Initiative (ADNI): Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging. Neuroimage Clin. 2013; 3:180–195. DOI: 10.1016/j.nicl.2013.07.006 [PubMed: 24179862]

9. Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I. Alzheimer's Disease Neuroimaging Initiative: Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. Front Neurosci. 2015; 9:307.doi: 10.3389/Fnins.2015.00307 [PubMed: 26388719]

10. Jie B, Zhang DQ, Gao W, Wang Q, Wee CY, Shen DG. Integration of network topological and connectivity properties for neuroimaging classification. IEEE Trans Bio Med Eng. 2014; 61(2): 576–589. DOI: 10.1109/Tbme.2013.2284195

(a)NC (b) MCI

Fig. 1.

Comparison of connectivity strengths based on the most discriminative connections

Table 1

Comparison of classification performance of different connectivity-networks

Method	ACC(%)	SEN(%)	SPE(%)	BAC(%)	AUC
Pearson correlation	70.49	75.00	66.67	70.83	0.69
Lasso	73.77	67.86	78.79	73.32	0.75
Ultra Lasso	77.05	75.00	78.79	76.89	0.73
Group Lasso	77.05	78.57	75.76	77.16	0.72
Proposed	80.33	75.00	84.85	79.92	0.81

Mach Learn Med Imaging. Author manuscript; available in PMC 2018 June 13.

Table 2

The most selected brain regions in classification

Index	Full name	Location
37	Hippocampus	Limbic lobe
71	Caudate nucleus	Subcortical gray nuclei
6	Superior frontal gyrus, orbital part	Frontal lobe
28	Gyrus rectus	Frontal lobe
31	Anterior cingulate and paracingulate gyri	Limbic lobe
42	Amygdala	Subcortical gray nuclei
52	Middle occipital gyrus	Occipital lobe
54	Inferior occipital gyrus	Occipital lobe
65	Angular gyrus	Parietal lobe
81	Superior temporal gyrus	Temporal lobe

Table 3

The most discriminative connections. (STG = Superior temporal gyrus; PCG = Posterior cingulate gyrus; TPOmid = Temporal pole, middle temporal gyrus; CUN = Cuneus; HIP = Hippocampus; PHG = Parahippocampal gyrus; ORBsup = Superior frontal gyrus, orbital part; ANG = Angular gyrus; PCL = Paracentral lobule; ROL = Rolandic operculum; ORBmid = Middle frontal gyrus, orbital part; TPO-sup = Temporal pole, superior temporal gyrus; PUT = Lenticular nucleus, putamen; AMYG = Amygdala; ORBinf = Inferior frontal gyrus, orbital part; PreCG = Precentral gyrus; PCL = Paracentral lobule; MFG = Middle frontal gyrus; L = Left; R = Right)

Selected ROIs	Direction of connectivity	Neighbors of selected ROIs	p-values
PHG.R	←	STG.R	0.0054
CUN.R	→	TPOmid.L	0.0061
ANG.R	←	CUN.L	0.0063
PCL.R	←	PUT.L	0.0064
ROL.L	→	CUN.L	0.0068
ORBmid.R	→	ORBinf.R	0.0070
TPOsup.R	←	AMYG.L	0.0070
PCG.R	→	HIP.L	0.0077
ORBsup.R	→	PreCG.R	0.0080
PCL.L	→	MFG.L	0.0094
IPL.R	→	SMG.R	0.0089
AMYG.L	←	Vermis9	0.0099