Published in final edited form as:

Mach Learn Med Imaging. 2017 ; 2017: 307–315.

Fusion of High-Order and Low-Order Effective Connectivity Networks for MCI Classification

Yang Li¹, Jingyu Liu¹, Ke Li², Pew-Thian Yap³, Minjeong Kim³, Chong-Yaw Wee⁴, and Dinggang Shen³

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

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

³Department of Radiology and BRIC, UNC Chapel Hill, Chapel Hill, NC, USA

⁴Department of Biomedical Engineering, National University of Singapore, Kent Ridge, Singapore

Abstract

Functional connectivity network derived from resting-state fMRI data has been found as effective biomarkers for identifying patients with mild cognitive impairment from healthy elderly. However, the ordinary functional connectivity network is essentially a low-order network with the assumption that the brain is static during the entire scanning period, ignoring the temporal variations among correlations derived from brain region pairs. To overcome this weakness, we proposed a new type of high-order network to more accurately describe the relationship of temporal variations among brain regions. Specifically, instead of the commonly used undirected pairwise Pearson's correlation coefficient, we first estimated the low-order effective connectivity network based on a novel sparse regression algorithm. By using the similar approach, we then constructed the high-order effective connectivity network from low-order connectivity to incorporate signal flow information among the brain regions. We finally combined the low-order and the high-order effective connectivity networks using two decision trees for MCI classification and experimental results obtained demonstrate the superiority of the proposed method over the conventional undirected low-order and high-order functional connectivity networks, as well as the low-order and high-order effective connectivity networks when they were used separately.

1 Introduction

Mild cognitive impairment (MCI) is considered to be a transitional state between normal senility and Alzheimer's disease (AD) [1, 2]. Around 10% to 15% of MCI patients deteriorate to AD every year, and more than half of them develop into AD within 5 years [1, 2]. Due to its serious negative impacts on the healthcare system, economy and society, it is thus crucial to accurately identify MCI at its early stage so that appropriate actions can be taken to minimize the burdens. However, MCI is difficult to diagnose because of its relative subtle symptoms of the cognitive impairment [3]. Functional connectivity, defined as the temporal correlations of brain regions [4, 5], is often described through the functional connectivity network using the knowledge of graph theory [6–9].

In conventional modeling approaches [3, 7, 9, 10], the vertices of functional connectivity networks correspond to the brain regions and the edges correspond to the correlation among brain regions. This kind of network is regarded as low-order network and the correlation of low-order network is regarded as low-order correlation [6]. It is noteworthy that the low-order correlation is normally calculated using the whole time series. This correlation is a fixed value with no temporal variations of the correlation are encoded [6]. However, the brain activities are indeed not static across the entire scanning period and the correlations among brain regions varied across time. Therefore, the conventional methods, which ignoring the temporal variations among correlations, may fail to diagnose MCI accurately. Recently, a high-order network modeling method [6] has been proposed to preserve the dynamic correlation information neglected in the conventional methods. Specifically, the vertices of high-order network correspond to the brain region pairs while the edges correspond to the relevance among temporarily correlated time series. However, the existing high-order network modeling approach [6] is based on pairwise Pearson's correlation coefficient, thus inconsistent with the sparse nature and small-world characteristic of most biologically networks [11]. Furthermore, the high-order correlation-based networks are undirected networks that cannot characterize the directed causal influence between the brain regions. To overcome these drawbacks, we enhance the network modeling approach with a novel sparse regression algorithm. Particularly, different from the traditional sparse regression model with a l_1 -norm penalization [12] which leads to different network structures at individual level, the proposed sparse model utilizes a $l_{2,1}$ -norm penalization to encourage an identical network structure among subjects. Moreover, an ultra-least squares (ULS) criterion is employed to extract more information from fMRI time series, such as the information of the classical dependent relation of the fMRI data and the dependent relation of the associated weak derivatives. While the low-order network ignores the temporal variations of correlation, the high-order network unable to characterize the holistic correlation calculated on the whole time series. Therefore, to incorporate both the low-order correlation and the temporally dynamic information encoded in the high-order correlation for better classification performance, we first constructed a decision tree (DCT) for each type of correlations and then fused their classification scores to provide the final classification decision. The fused DCT model considers not only the integrated correlation calculated on the whole time series but also the temporal variations of correlation. We compared our proposed framework (i.e., fusion of high- and low-order effective connectivity networks) with other related methods on a dataset and experimental results demonstrate the superiority of the proposed framework for MCI classification.

2 Materials

In this study, the fMRI data were collected from 28 MCI individuals and 33 healthy controls. All the subjects were scanned using a standard echo-planar imaging (EPI) sequence on a 3 T Siemens TRIO scanner with following parameters: TR = 3000 ms, TE = 30 ms, acquisition matrix = 74×74 , 45 slices, and voxel thickness = 3 mm. One-hundred and eighty resting-state fMRI volumes were acquired. Standard preprocessing pipeline of the fMRI images was performed using Statistical Parametric Mapping 8 (SPM8) software package which includes removal of first 10 fMRI volumes, slice timing correction, head-motion correction,

regression of nuisance signals (ventricle, white matter, global signal, and head-motion with Friston's 24-parameter model [9]), signal de-trending, and band-pass filtering (0.01–0.08 Hz). Next, the brain space was parcellated into 90 ROIs based on the automated anatomical labeling (AAL) atlas [6].

3 Methods

3.1 The ULS Criterion

Suppose there is a linear system with k inputs and one output, the ordinary least squares regression problem can be solved via the least squares criterion as follows:

$$J_{LS} = \left\| y - \sum_{i=1}^k \theta_i x_i \right\|_2^2, \quad (1)$$

where y , x_i denote the system output and input variables, θ_i is the parameter and e is the system noise, respectively. Supposing $[0, T]$ is the time span of observed data and $\hat{y}(t)$ is prediction function of $y(t)$, it is obvious that the least squares criterion only measures the discrepancy between $y(t)$ and $\hat{y}(t)$ on the whole interval $[0, T]$, ignoring how the discrepancy distributed at every individual time point. Therefore, the least squares criterion cannot accurately describe the similarity of function shape and discards the information of correlations among the data points, thus leading to the common overfitting problem for the identification of dynamic system [13].

In order to overcome this drawback, a new ULS criterion has been adopted to produce a more accurate evaluation standard about the model fitness, which is defined as the combination of the least squares criterion and the weak derivative:

$$J_{ULS} = \left\| y - \sum_{i=1}^k \theta_i x_i \right\|_2^2 + \sum_{l=1}^L \left\| D^l y - \sum_{i=1}^k \theta_i D^l x_i \right\|_2^2, \quad (2)$$

where D^l denotes the l -th order weak derivative. The new criterion considers not only the discrepancy between two functions at a whole level, but also the discrepancy between weak derivatives of the two functions.

3.2 Effective Network Structure Estimation via Ultra-Group Lasso

Suppose we have M ROIs and N subjects, the objective function of widely used Lasso with a l_1 -norm penalization is defined as follows:

$$f(\theta_m^n) = \left\| y_m^n - A_m^n \theta_m^n \right\|_2^2 + \lambda \left\| \theta_m^n \right\|_1, \quad (3)$$

where y_m^n denotes the M ROI time series of m -th ROI for n -th subject,

$A_m^n = [y_1^n, \dots, y_{m-1}^n, y_{m+1}^n, \dots, y_M^n]$ is data matrix which includes all ROIs time series except

the m -th ROI, $\theta_m^n = [\theta_1^n; \dots; \theta_{m-1}^n; \theta_{m+1}^n; \dots; \theta_M^n]$ denotes the weight vector and $\lambda > 0$ is the regularization parameter that controls the sparsity level of the regression model. Constructing sparse effective network structure can be considered as an optimization problem, i.e. minimizing the above objective function.

To minimize inter-subject variability, a Group-Lasso algorithm has been used to estimate an identical network structure across subjects by replacing l_1 -norm penalization with $l_{2,1}$ -norm penalization. The objective function of Group-Lasso is given as

$$f(\theta_m) = \sum_{n=1}^N \|y_m^n - A_m^n \theta_m^n\|_2^2 + \lambda \|\theta_m\|_{2,1}, \quad (4)$$

where $\theta_m = [\theta_m^1, \dots, \theta_m^{n-1}, \theta_m^{n+1}, \dots, \theta_m^N]$ denotes the weight matrix. By incorporating the ULS criterion with the Group-Lasso algorithm, we proposed an Ultra-Group Lasso (UG-Lasso) algorithm to estimate a more accurate effective network structure via the following objective function

$$f(\theta_m) = \sum_{n=1}^N \|y_m^n - A_m^n \theta_m^n\|_2^2 + \sum_{n=1}^N \sum_{l=1}^L \|D^l y_m^n - D^l A_m^n \theta_m^n\|_2^2 + \lambda \|\theta_m\|_{2,1}. \quad (5)$$

3.3 Effective Connectivity Strength Estimation via UOLS

It is important to note that the estimated coefficient based on Eq. (5) cannot be regarded as the effective connectivity strength, because they are biased as the result of group-constrained sparse penalization. Furthermore, even some of the coefficients are negative, leading to difficulty in interpreting and analyzing the effective connectivity network. Therefore, we proposed to utilize an Ultra-Orthogonal Least Squares (UOLS) algorithm to estimate the effective connectivity strength between the time series of two brain regions [13]. Suppose there are P ROIs have been found correlated to m -th ROI based on the UG-Lasso in Eq. (5), the effective connectivity between m -th ROI and the other P ROIs can be estimated by minimizing the objective function:

$$f(\alpha_m^n) = \|y_m^n - \Phi_m^n \alpha_m^n\|_2^2 + \sum_{l=1}^L \|D^l y_m^n - D^l \Phi_m^n \alpha_m^n\|_2^2, \quad (6)$$

where $\Phi_m^n = [y_1^n, y_2^n, \dots, y_P^n]$ is the data matrix including time series of P ROIs and $\alpha_m^n = [\alpha_1^n, \alpha_2^n, \dots, \alpha_P^n]$ is the weight vector, i.e. the strength of effective connectivity.

3.4 Construction of High-Order UG-Lasso-UOLS-Based Networks

The first step in high-order network construction is applying a sliding window to partition ROIs time series into multiple overlapping segments. Suppose the size of the sliding window is S and the step size between adjacent windows is r . For one ROI time series containing Z

temporal image volumes, y_m^n , the total number of time series segments generated using the sliding window can be computed as $K = [(Z-S)/T] + 1$. Let $y_m^n(k)$ be the k -th segment generated from y_m^n , for n -th subject, the k -th segments of all M ROIs can be represented in a matrix form as $Y^n(k) = [y_1^n(k), y_2^n(k), \dots, y_M^n(k)] \in R^{S \times M}$. Further, the set of k -th segments for all subjects and all ROIs can be expressed as $Y(k) \in R^{S \times M \times N}$. Thus, by applying sliding window to $fMRI$ series, we can obtain K sets of time series segments $Y(k)$.

For each set of segments $Y(k)$, we applied an UG-Lasso across subjects to derive the network structure and then applied an UOLS at individual level to estimate the connectivity strength of each derived connection, producing a total of $K \times N$ temporal low-order effective connectivity networks. Taking each ROI with $\{y_m^n(k)\}$ as the vertex and $\{C_{ij}^n(k)\}$ as the strength of connection for each pair of vertices, the temporal low-order effective connectivity network can be expressed as $G^n(k) = (\{y_m^n(k)\}, \{C_{ij}^n(k)\}) (k = 1, 2, \dots, K)$. A larger value of $\{C_{ij}^n(k)\}$ indicates a stronger connection between i -th ROI and j -th ROI in the k -th window [6].

For the n -th subject, the correlation time series of each ROI pair (i, j) $C_{ij}^n = [C_{ij}^n(1), C_{ij}^n(2), \dots, C_{ij}^n(K)] \in R^K$ can be obtained by concatenating all $C_{ij}^n(k) (k = 1, 2, \dots, K)$. Different from y_m^n representing the time series of a ROI, C_{ij}^n characterizes the variations of correlation of ROI pair (i, j) across time [6]. Considering the low-order network is directional, the total number of correlation time series $\{C_{ij}^n | 1 \leq i \leq M, 1 \leq j \leq M, i \neq j\}$ for the n -th subject is $M(M-1)$. For the set of all correlation time series $\{C_{ij}^n\}$, we employed an UG-Lasso to estimate the network structure and then employed an UOLS to estimate the connectivity strength, establishing a total of N high-order networks with identical network structure. By considering each ROI pair (i, j) with $\{C_{ij}^n\}$ as vertices and $\{E_{ij, pq}^n\}$ as the weights of edges, high-order network can be expressed as $G^n = (\{C_{ij}^n\}, \{E_{ij, pq}^n\})$, where $\{E_{ij, pq}^n\}$ represents the correlation for each pair of correlation time series $(\{C_{ij}^n\}, \{C_{pq}^n\})$. Therefore, high-order network is devoted to describe the relationship of temporal correlations among brain regions.

However, there is an obvious deficiency about high-order network, i.e., the scale of high-order network is too large [6]. As mentioned above, the number of vertices $\{C_{ij}^n\}$ is $M(M-1)$, thus the number of edges is proportional to M^4 . The large-scale of high-order network leads to a large amount of computation and poor generalization performance. To overcome this deficiency, we applied the Ward's hierarchical grouping [14] to group the correlation time series into clusters. Then, the mean correlation time series of each cluster was used as the vertices of high-order network. As a result, the scale of high-order network can be significantly reduced and the generalization capability of the high-order network can be significantly improved.

3.5 Feature Extraction and Selection

In this work, we extract four frequently used network features from low-order and high-order effective connectivity networks, respectively, including the weighted clustering coefficient, betweenness centrality, in-degree and out-degree [15]. For a node i of an effective connectivity network, the weighted clustering coefficient is defined as

$$WC_i = \frac{\sum_{j \neq i} \sum_{h \neq (i,j)} (c_{i,j}^{1/3} + c_{j,i}^{1/3})(c_{i,h}^{1/3} + c_{h,i}^{1/3})(c_{j,h}^{1/3} + c_{h,j}^{1/3})}{2[d_i(d_i - 1) - 2d_i^{\leftrightarrow}]}, \quad (7)$$

where $c_{i,j}$ is the effective connectivity strength from node i to j , d_i represents the number of adjacent points of node i , $d_i^{\leftrightarrow}$ denotes the number of bilateral edges between i and its adjacent nodes.

After feature extraction, we use a filter-based approach which based on the correlation coefficient between the features and the class labels of training samples to select the most discriminative features for classification. The features selected in the first step will form a feature set S_1 . Then, for the diversification of the feature selection methods, we apply the Relief algorithm [16] to compute the weights for the features which are not belong to S_1 . The features whose weights are larger than the predefined threshold will be selected to form a feature set S_2 . Finally, $S = S_1 \cup S_2$ will be the final feature pool for MCI classification.

3.6 Classification via Fusion DCT Model

After completing the feature selection, we performed 10-fold cross-validation for MCI classification to evaluate the performance of our proposed method. For each fold of cross-validation, two DCT submodels were constructed based on the features selected from both the low-order UG-Lasso-UOLS-based network and the high-order UG-Lasso-UOLS-based network, respectively. For a testing subject, each DCT submodel can provide a classification score for each subject, and we get the final classification scores for each subject via combining the classification scores from the two DCT submodels with equal weight. We repeated the cross-validation for 20 times and reported our results as the mean of these 20 repetitions.

4 Results and Discussions

In this work, we compared the proposed fusion UG-Lasso-UOLS-based method with other methods including low-order correlation-based method, high-order correlation-based method [6], fusion correlation-based method [6], low-order UG-Lasso-UOLS-based method (combination of [3, 13]), and high-order UG-Lasso-UOLS-based (combination of [3, 6, 13]) method on the same dataset. To characterize performance comparison between different methods, we used four typical performance measures including accuracy (ACC), sensitivity (SEN), specificity (SPC), and area under curve (AUC). As shown in Table 1, the proposed fusion UG-Lasso-UOLS-based method yields the best accuracy of 85.5%, the best sensitivity of 86.6% and the best AUC of 0.896, respectively. Although the best specificity

of 88.9% belongs to high-order UG-Lasso-UOLS-based method, it also indicates dynamic correlation information is important and should be incorporated for MCI classification. Furthermore, our proposed method performed significantly better than all competing methods in terms of ACC, SPC, and AUC based on two-sample t-test results on 20 repetitions. Table 1 also shows the p -values for the comparison of AUC between the proposed method and other methods.

The most discriminative brain regions and clusters are defined as the ones with the highest selected frequency in 20 repetitions of 10-fold cross-validation corresponding to the low-order and the high-order network, respectively [6]. The most discriminative brain regions selected from low-order network includes hippocampus left, orbitofrontal cortex (superior) right, inferior parietal lobule right, inferior frontal gyrus (opercular) right, anterior cingulate gyrus left, and these regions are frequently reported as highly associated with AD/MCI pathology [6, 10]. Direction information of effective connection among the top 20 most discriminative ROIs is provided in Fig. 1. As for high-order networks, the ROI pairs in the most discriminative clusters includes middle occipital gyrus left and middle occipital gyrus right, Heschl gyrus left and Superior temporal gyrus left, etc. It means that the temporal variations of correlation between these ROI pairs are discriminative for MCI classification. Additionally, Fig. 2 shows the difference of correlation among the top 20 most discriminative clusters between MCI and NC subjects.

5 Conclusions

In this paper, we proposed a novel fusion approach to infer high- and low-order effective connectivity networks for MCI classification. By using the UG-Lasso and UOLS, our approach can minimize inter-subject variability and extract the temporal information of connections among fMRI data, which can be further used to construct the effective connection for low-order and high-order networks. Additionally, by fusing DCT models constructed using low-order and high-order networks, our proposed approach considers integrated correlation that computed both the entire time series and the temporal variations of correlation, simultaneously. Promising results obtained demonstrate the superiority of our proposed method and the importance of dynamic temporal correlation information for MCI classification.

References

1. Petersen RC, et al. Current concepts in mild cognitive impairment. Arch Neurol. 2001; 58:1985–1992. [PubMed: 11735772]
2. Gauthier S, et al. Mild cognitive impairment. Lancet. 2006; 367:1262–1270. [PubMed: 16631882]
3. Wee, CY., et al. Constrained sparse functional connectivity networks for MCI classification. MICCAI; 2012. p. 212-219.
4. Aertsen AM, et al. Dynamics of neuronal firing correlation: modulation of “effective connectivity”. J Neurophysiol. 1989; 61:900–917. [PubMed: 2723733]
5. van den Heuvel MP, et al. Exploring the brain network: A review on resting-state fMRI functional connectivity. Eur Neuropsychopharm. 2010; 20:519–534.
6. Chen XB, et al. High-order resting-state functional connectivity network for MCI classification. Hum Brain Mapp. 2016; 37:3282–3296. [PubMed: 27144538]

7. Jie, B., Shen, D., Zhang, D. Brain Connectivity Hyper-Network for MCI Classification. In: Golland, P.Hata, N.Barillot, C.Hornegger, J., Howe, R., editors. LNCS; MICCAI 2014; Cham: Springer; 2014. p. 724-732.
8. Wee CY, et al. Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification. *Brain Imaging Behav.* 2016; 10:342–356. [PubMed: 26123390]
9. Wee CY, et al. Resting-state multi-spectrum functional connectivity networks for identification of MCI patients. *PLoS ONE.* 2012; 7:e37828. [PubMed: 22666397]
10. Wee CY. Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage.* 2012; 59:2045–2056. [PubMed: 22019883]
11. Supekar K, et al. Network Analysis of Intrinsic Functional Brain Connectivity in Alzheimer's Disease. *PLoS Comput Biol.* 2008; 4:e1000100. [PubMed: 18584043]
12. Lee H, et al. Sparse brain network recovery under compressed sensing. *IEEE Trans Med Imaging.* 2011; 30:1154–1165. [PubMed: 21478072]
13. Li Y, et al. Time-varying system identification using an ultra-orthogonal forward regression and multiwavelet basis functions with applications to EEG. *IEEE Trans Neural Netw Learn Syst.* 2017:1–13.
14. Ward JH Jr. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc.* 1963; 58:236–244.
15. Rubinov M, et al. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage.* 2010; 52:1059–1069. [PubMed: 19819337]
16. Kira K, et al. The feature selection problem: traditional methods and a new algorithm. *AAAI.* 1992:129–134.

Fig. 1.
Comparison of effective connections from low-order effective connectivity networks among the top 20 most discriminative ROIs with a threshold weight of 0.2.

Fig. 2.
Comparison of correlation difference from high-order effective connectivity networks among the top 20 most discriminative clusters.

Table 1

Performance comparison from different network modeling methods

Method	ACC(%)	SEN(%)	SPC(%)	AUC	<i>p</i> -values
Low-order correlation-based	62.4	51.4	71.7	0.600	2.7×10^{-17}
High-order correlation-based	69.3	63.9	73.9	0.711	3.8×10^{-13}
Fusion correlation-based	69.7	63.0	75.3	0.730	1.8×10^{-12}
Low-order UG-Lasso-UOLS	81.5	74.3	87.6	0.756	2.0×10^{-13}
High-order UG-Lasso-UOLS	81.6	72.9	88.9	0.755	5.6×10^{-11}
Fusion UG-Lasso-UOLS	85.5	86.6	84.6	0.896	–