ISSN 2616-5775 Vol. 4, Issue 1: 87-93, DOI: 10.25236/AJCIS.2021.040114

Classification of MCI Brain Network Based on Orthogonal Minimum Spanning Tree

Fei Han, Miao Song*

College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China
**Corresponding Author: miaosong@shmtu.edu.cn*

Abstract: *Brain network based on resting-state functional magnetic resonance (rs-fMRI) is the most popular method for brain disease diagnosis, which is expected to provide accurate and effective biomarkers. The original fully connected resting-state networks (RSNs) are too dense and must be filtered to get the real network model. In this study, orthogonal minimum spanning trees (OMSTs) was used to filter the connection matrix of 49 age-matched healthy controls (HC) and 50 patients with mild cognitive impairment (MCI). At the same time, we also used global cost efficiency (GCE) algorithm to filter brain network for comparison with OMSTs. We calculated the topological metrices of brain network. Fisher score was used to select features, and the optimal feature subset was used to construct SVM classifier. The classification accuracy of OMSTs was 87%, while that of GCE algorithm was 81%. The experimental results show that the classification accuracy is greatly improved by using OMSTs, which is an effective brain network filtering method.*

Keywords: *brain network, graph theory, orthogonal minimum spanning tree, support vector machine*

1. Introduction

Functional magnetic resonance (fMRI) is a non-invasive neuroimaging technique, which is used to explore the neural activity of the brain [1]. Brain network based on rs-fMRI regards the brain as a graph to explore the relationship between different brain regions from the perspective of functional integration [2]. In the process of brain network construction, the filtering of connection matrix is an essential step. Some studies have found that brain regions are sparsely connected anatomically, while the original fully connected resting state brain network is too dense, so it is necessary to filter the brain network. In recent years, many filtering methods have been proposed for brain network filtering.

Proportional thresholding is the common brain network filtering technique. It sorts the connection weights, and then retains a certain proportion of strong connections. This method is easy to operate, but it is not easy to choose the proportion. If the value is too small, it will form isolated nodes in the network, and if the value is too large, it can not achieve the filtering effect. In order to determine a suitable threshold, some researchers have used the area under the curve of brain network attributes under continuous threshold interval as a feature to construct a classifier. The GCE formula is defined as the difference between the global efficiency of the brain network and the cost of maintaining the network. Khazaee and his colleagues combined the GCE formula with the proportional threshold method [3]. The GCE formula is used as the basis of proportion selection in proportional threshold method, and the proportion that maximizes GCE is searched in continuous threshold range. The minimum spanning tree assumes that the information between brain regions flows through the shortest path, and extracts the core structure of the network. The advantage of this method is that it can keep strong connections and weak connections at the same time. Researches show that the combination of strong and weak connectivity can improve the reliability of biomarkers. But the brain network constructed by minimum spanning trees (MSTs) is too sparse and loses its original topological properties. In order to overcome these problems, OMSTs were introduced in previous studies [4]. The algorithm uses spanning tree theory to ensure the connectivity of all nodes in the network, and includes weak nodes and strong nodes in the network at the same time. In addition, it optimizes the GCE of a single network. GCE represents the ability to combine the specialized information of distributed brain regions with its cost (global efficiency cost) , providing a method to quantify the topology of economic small world.

In this study, we use OMSTs as brain network filtering method. The topological attributes are extracted from brain network, and combined with machine learning algorithm. An accurate and automatic classification method of mild cognitive impairment is developed.

ISSN 2616-5775 Vol. 4, Issue 1: 87-93, DOI: 10.25236/AJCIS.2021.040114

2. Method

2.1 Subjects

In this study, the functional and structural MRI images of 40 patients with MCI (average age 78.5 years, 20 female) and 30 age-matched HC (average age 75.2 years, 19 female) were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.ucla.edu).

1) The patients with MCI had MMSE scores between 24 and 30, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living and absence of dementia.

2) The HC subjects were non-depressed, non-MCI, non-demented and had an MMSE score of 24–30 and a CDR of 0. Demographic information of subjects is summarized in Table 1.

Table 1: Demographics of healthy control subjects and patients with AD

	Healthy Control	Patients with MCI
Number	39	40
Male/female	20/19	20/20
Age	75.2 ±4.1	78.5 ±3.3
MMSE score	26.8 ±2.5	21.2 ±3.4
CDR score	0.45 ±0.13	0.92 ±0.25

2.2 Data acquisition and preprocessing

The functional and structural MRI images were collected using 3-T Philips scanners.Acquisitions were performed according to the ADNI acquisition protocol. A total of 140 functional volumes (TR/TE 3000/30 mses, flip angle=80,3.331 mm slice thickness,48 slices) were obtained.

DPRAF toolboox and SPM8 pachage were used to preprocess the rs-fMRI data. The preprocessing steps were as follows [5]. Firstly, the first 10 volumes of each subject were removed, and the remaining 130 volumes were subject to sliced timing and motion correction. The subjects with horizontal head movement greater than 3 mm and rotation greater than 3 degrees were removed. Then, the fMRI volumes after motion correction were spatially standardized, and transformed into the MNI standard space of 3 mm voxels by 12 dimensional optimal affine transformation. T he volumes were smoothed with 4 mm full width half height to improve the signal-to-noise ratio of the image. Finally, low frequency filtering was used to reduce low frequency drift and high frequency biological noise.

2.3 Brain network analysis

Automated anatomical labeling (AAL) atlas were used to divide the brain into 116 distinct regions [6], with 90 regions located in the cerebral cortex and 26 regions in the cerebellum. In this study, only 90 regions of cerebral cortex in AAL were used as nodes of brain network. The time series of voxels within each region was averaged and the resulted signal was used as the representative signal of the node. Then, the Pearson correlations between nodes were calculated to denote the edge between nodes.The resulted functional connectivity network has 4005 (89*90/2) weighted edges and thus is a dense network.

2.4 Topological filtering of brain network by using OMST

OMSTs combined the strategies of MSTs and GCE, which is a data-driven brain network filtering method [7]. The steps are as follows.

1) The MSTs were extracted by iteratively applying Kruskal's algorithm on the inverted weighted Functional Connectivity Graphs. N-1 edges were extracted to connect all the N nodes.

2) In order to maintain orthogonality, the N-1 edges selected in the 1st MST were substituted with zeros in the original network. Then a 2nd MST was estimated that connects all of the N nodes with minimal total distance, satisfying the constraint that it is orthogonal with the 1st MST. Next, the N-1 connections of the 2nd MST were substituted with zeros and a 3rd MST was estimated that connected the nodes with the minimal total weight, subject to the constraint that it is orthogonal to the previous two constructed MST's (1st and 2nd). In general, an mth-MST is orthogonal to all the previous (m-1)th MSTs, having exactly m(N-1) edges.

ISSN 2616-5775 Vol. 4, Issue 1: 87-93, DOI: 10.25236/AJCIS.2021.040114

3) Connections were aggregated across OMSTs in order to optimize the function of global efficiency (GE) minus Cost over Cost.

4) For each added connection to the aggregated network, the objective function of Global Cost Efficiency (GCE) = GE-Cost was estimated, where Cost denotes the ratio of the total weight of the selected edges, over multiple iterations of OMST, divided by the total strength of the original fully-weighted graph. The values of this formula range within the limits of an economical small-world network for healthy control participants. The network which is considered as functionally optimal is the one associated with the maximum value of the following quality formula:

$$\mathrm{J}_{\mathrm{GCE}}^{\mathrm{OMSTs}} = \mathrm{GE} - \mathrm{Cos\,t}$$

2.5 Computation of graph measures

After filtering the original brain network, different graph measures (2 global measures and 5 local measures) were computed. Global measures assess ability of the brain in rapidly combining specialized information from distributed regions.Local measures characterize the ability of the brain for specialized processing to occur within densely interconnected groups of regions. The two global measures are network efficiency and small-worldness.The five local measures selected in this paper are clustering coefficient, shortest path length, betweenness centrality, degree centrality and node efficiency. Global measures have only one value for each graph whereas local measures have n values (n is the number of nodes in the graph) for each graph [8]. The size of final feature vector for each subject was 452 (90×5=450 local and 2 global features).

Clustering coefficient is used to quantify the possibility of neighbors becoming neighbors [9]. Assuming that node i has k_i neighbor nodes, there are at most $k_{i*}(k_i-1)/2$ edges between neighbor nodes, and the actual number of edges is E_i. Then the clustering coefficient of node i is:

$$C_i = \frac{E_i}{k_i(k_i - 1)/2} \tag{1}$$

The characteristic path length is defined as the number of shortest paths that a node needs to go through when its information reaches another node. The average shortest path of node i is equal to the average of the sum of the shortest path lengths L_i of node i and other nodes.

$$L_i = \frac{1}{N-1} \sum_{i,j \in G, i \neq j} l_{ij} \tag{2}$$

Betweenness centrality is defined as the importance of nodes as bridges. The number of the shortest paths of all node pairs in the network passing through node i.

$$B(i) = \sum_{i,s,t \in G, i \neq s \neq t} P_{st}(i) \tag{3}$$

Degree centrality is defined as the degree of connection between a node and other nodes. The degree centrality of node i is equal to the sum of the number of edges connected to it.

$$D_i = \sum_{i,j \in V, i \neq j} x_{ij} \tag{4}$$

Node efficiency is defined as the parallel transmission efficiency of nodes in the network. Node efficiency is defined as the reciprocal of the average shortest path.

$$E_i = \frac{1}{L_i} \tag{5}$$

Network efficiency is defined as the information transmission capacity of the network. Network efficiency is defined as the average efficiency of network nodes.

$$E = \frac{1}{N} \sum_{i \in G} E_i \tag{6}$$

ISSN 2616-5775 Vol. 4, Issue 1: 87-93, DOI: 10.25236/AJCIS.2021.040114

If the ratio of clustering coefficient of real network and random network is greater than 1, and the ratio of shortest path length is about 1, then σ is greater than 1, which means the network has small world property.

$$\sigma = \frac{\gamma}{\lambda} \tag{7}$$

$$\gamma = \frac{c_{real}}{c_{random}} \tag{8}$$

$$\lambda = \frac{L_{real}}{L_{random}} \tag{9}$$

2.6 Feature selection

An effective feature selection algorithm is the key part of machine learning method. In the feature selection stage, the optimal feature subset is selected from the original feature set. Feature selection can reduce the storage requirements and training test times, and improve the accuracy of classification. Feature selection algorithms can be roughly divided into two categories: filter and wrapper. This method is independent of the selected classifiers and selects feature subsets according to the general features of the data. However, the wrapper method needs a pre-determined classifier and evaluates the features according to its performance in classification. The most commonly used filtering method is Fisher algorithm. These algorithms are briefly described here.

Fisher score is a univariate feature selection algorithm [10]. The variance of features of the same class sample should be as small as possible, and the difference between features of different classes should be as large as possible. This is helpful to improve the classification accuracy of subsequent prediction results. Suppose m_i is the average of the i-th feature in all samples, m_{1i} is the average of the i-th feature in one sample, and m_{2i} is the average of the i-th feature in another sample. The Fisher score of each feature in the two types of problems is defined as:

$$FS(\mathrm{i}) = \frac{n_1(\mathrm{m}_{1i} - m_i)^2 - n_2(\mathrm{m}_{2i} - m_i)^2}{(n_1\sigma_{1\mathrm{i}}^2 - \mathrm{n}_2\sigma_{2\mathrm{i}}^2)} \tag{10}$$

In the formula, n_1 is the number of samples of the first type, n_2 is the number of samples of the second type, and σ_{1i}^2 is expressed as the i-th feature in the first type of sample. The variance in σ_{2i}^2 is expressed as the variance of the i-th feature of the second type of sample.

2.7 Classification and performance metrics

In this study, we use supervised machine learning method to construct classifier.The supervised machine learning algorithm is trained with a set of input data, namely the training data set, and produces the desired output. We use support vector machines (SVM) as supervised machine learning algorithms [10]. Support vector machine (SVM) is a supervised machine learning method originally developed for binary classification problems.

Data collection costs are high in human neuroimaging studies, so the number of available subjects is small. A small number of training samples and test samples will lead to the generalization ability of learning. Therefore, the use of cross-validation strategy is unavoidable. In this case, one of the most common methods is to omit cross-validation (LOOCV) [11]. LOOCV is repeated N times, leaving a sample in each iteration to test the classifier, and the remaining subjects are fed into the classifier for training. Repeat the procedure until all subjects use it once as a test sample. Finally, the final classification accuracy is obtained by averaging the results for each repetition.

In order to assess the performance of the classification results, the measures identified are as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Specifivity = \frac{TN}{TN + FP} \tag{13}$$

Where TP, TN, FP , and FN represent true positive, true negative, false positive, and false negative, respectively.

2.8 Experimental results and comparison

In this study, we compare OMSTs data-driven thresholding scheme and GCE algorithm. Both of them use the cost-efficiency formula as the objective criterion. The basic drawback of GCE algorithm is that it accumulates edges based on their strength and applying an iteratively absolute threshold searching approach from 0 to 1 without taking into consideration any topological constraint. Figure 1 is an example of applying the two methods to rs-fMRI data of patients with mild cognitive impairment. For GCE algorithm, when cost is 0.25 and cost-efficiency is 0.31, the optimal value is obtained. For OMST, when cost is 0.22 and cost-efficiency is 0.57, the optimal value is obtained.We can clearly see that OMST optimizes better the topological criterion compared to GCE algorithm. Figure 2 is the brain network filtered by OMSTs.

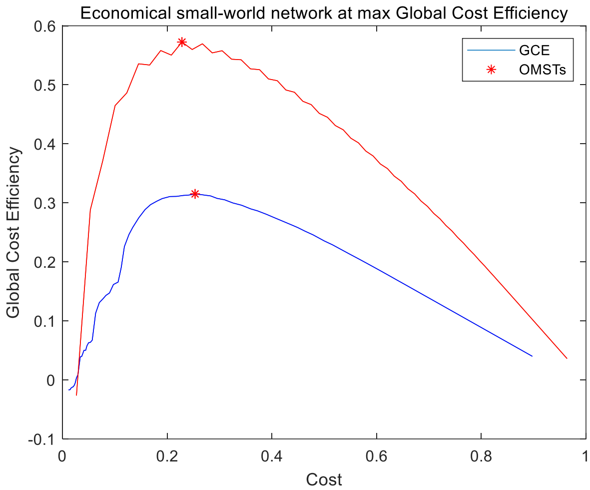

Figure 1: comparision of GCE algorithm and OMSTs data-driven thresholding scheme

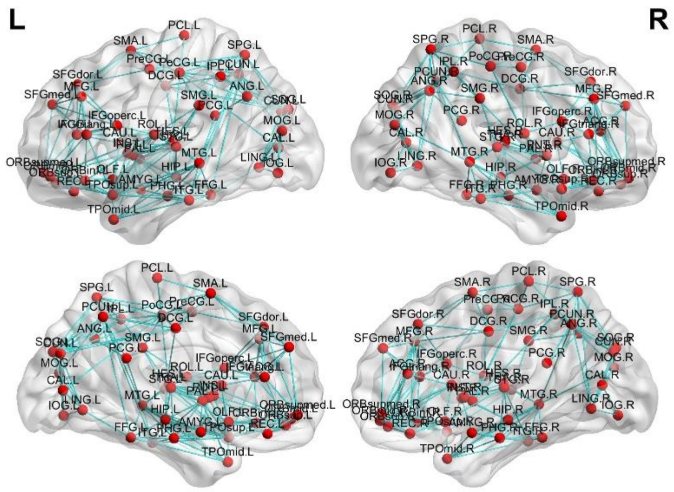

Figure 2: Brain network filtered by OMSTs

After filtering brain network with OMSTs, we calculate seven topological metrics of brain network. Because the feature dimension is too large, we use Fisher score for feature selection, and finally select 24 features. Figure 3 shows the number of features selected in each type of topological attribute.

ISSN 2616-5775 Vol. 4, Issue 1: 87-93, DOI: 10.25236/AJCIS.2021.040114

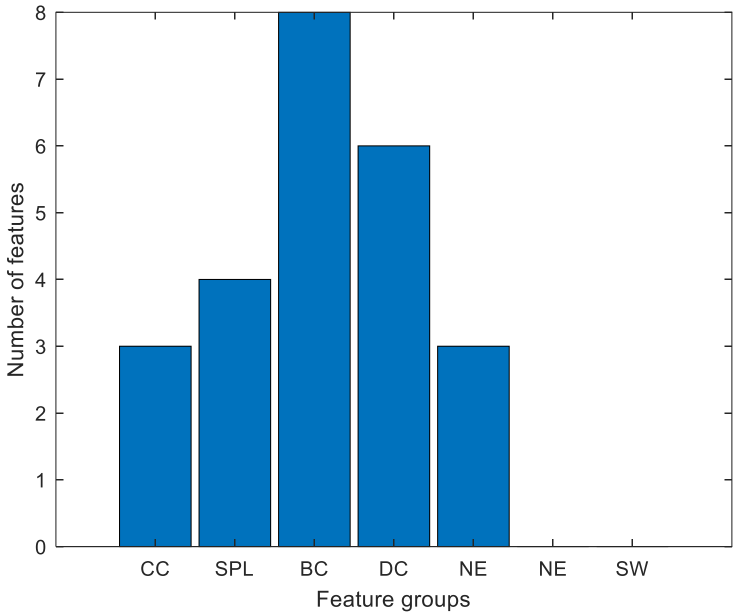

Figure 3: Number of the selected features from each group by the Fisher algorithm.

In this paper, OMSTs and GCE are used to filter the brain network and construct classifiers on the same dataset. The result is shown in the Table 2.

The classification accuracy based on OMSTs was 87%, which was significantly higher than 81% based on GCE. The experimental results show that the OMSTs-based filtering method can effectively improve the accuracy of classification.

Table 2: Classification performance corresponding to different features

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
OMSTs	0.87	0.82	0.86
GCE	0.81	0.79	0.80

4. Conclusion

In this study, we use OMSTs as brain network filtering strategy, and applied the graph theoretical approach and the pattern recognition method to classify two groups: HC and patients with MCI. The experimental results show that OMSTs is an effective topology filtering strategy, which can improve the classification accuracy of Alzheimer's disease diagnosis.

References

[1] KHAZAEE A, EBRAHIMZADEH A, BABAJANI-FEREMI A. Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease [J]. Brain Imaging Behav, 2016, 10(3): 799-817.
[2] SOARES J M, MAGALHAES R, MOREIRA P S, et al. A Hitchhiker's Guide to Functional Magnetic Resonance Imaging [J]. Front Neurosci, 2016, 10: 515.
[3] KHAZAEE A, EBRAHIMZADEH A, BABAJANI-FEREMI A. Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory [J]. Clin Neurophysiol, 2015, 126(11): 2132-2141.
[4] DIMITRIADIS S I, SALIS C, TARNANAS I, et al. Topological Filtering of Dynamic Functional Brain Networks Unfolds Informative Chronnectomics: A Novel Data-Driven Thresholding Scheme Based on Orthogonal Minimal Spanning Trees (OMSTs). [J]. Front Neuroinform, 2017, 11: 28.
[5] LV H, WANG Z, TONG E, et al. Resting-State Functional MRI: Everything That Nonexperts Have Always Wanted to Know. [J]. AJNR Am J Neuroradiol, 2018, 39(8): 1390-1399.
[6] ZHANG L, NI H, YU Z, et al. Investigation on the Alteration of Brain Functional Network and Its Role in the Identification of Mild Cognitive Impairment [J]. Front Neurosci, 2020, 14: 558434.
[7] CHAN Y L, UNG W C, LIM L G, et al. Automated Thresholding Method for fNIRS-Based Functional Connectivity Analysis: Validation With a Case Study on Alzheimer's Disease [J]. IEEE

ISSN 2616-5775 Vol. 4, Issue 1: 87-93, DOI: 10.25236/AJCIS.2021.040114

Trans Neural Syst Rehabil Eng, 2020, 28(8): 1691-1701.
[8] ZHANG T, ZHAO Z, ZHANG C, et al. Classification of Early and Late Mild Cognitive Impairment Using Functional Brain Network of Resting-State fMRI. [J]. Front Psychiatry, 2019, 10: 572.
[9] BULLMORE E, SPORNS O. Complex brain networks: graph theoretical analysis of structural and functional systems[J]. Nat Rev Neurosci, 2009, 10(3): 186-198.
[10] XU X, LI W, MEI J, et al. Feature Selection and Combination of Information in the Functional Brain Connectome for Discrimination of Mild Cognitive Impairment and Analyses of Altered Brain Patterns. [J]. Front Aging Neurosci, 2020, 12: 28.
[11] MAX A. LITTLE G V, SOHRAB SAEB, LUCA LONINI, ARUN JAYARAMAN, DAVID C. MOHR AND KONRAD P. KORDING. Using and understanding cross-validation strategies. [J]. Gigascience, 2017.