Discriminative Brain Network Using Sparse Regression

Hyekyoung Lee [1], Dong Soo Lee [1], Hyejin Kang [1], Boong-Nyun Kim [2], and Moo K. Chung [3,4]

[1] Department of Nuclear Medicine
[2] Department of Neuropsychiatry
Seoul National University College of Medicine, Seoul, Korea
[3] Department of Biostatistics and Medical Informatics
[4] Waisman Laboratory for Brain Imaging and Behavior
University of Wisconsin, Madison, WI 53706, USA
leehk@postech.ac.kr

Abstract. The correlation and the partial correlation are widely used for measuring connectivity of an undirected brain network. It is known that the brain network has a small-world and a scale-free topology, but its structure drastically changes depending on the criterion of how to threshold correlations. The exact threshold criterion has not been known yet, except for the statistical significance which is usually determined heuristically.

In this paper, we propose a novel framework for automatically determining the threshold based on the clustered structure of the network. By building sparse linear regression framework on correlations, we can utilize the inherent sparseness of the brain network and thus making the threshold easy to determine. We will show that our proposed method finds the biologically meaningful connectivity by making the best representation of the data characteristics and it discriminates brain networks between groups very well.

1 Introduction

The majority of connectivity analyses has been on thresholding correlation in detecting focal regions of correlated voxels [1–4]. The main limitation of connectivity analyses based on correlation is that they fail to explicitly factor out the confounding effect of other regions. To remedy this limitation, partial correlation has been naturally introduced in factoring out the dependencies of other regions [5–7] or eliminating the effect of experimental designs [8].

The correlation and the partial correlation are important measures for representing the brain connectivity, however, it depends on how we choose the threshold. Depending on the threshold, the characteristics of the brain network change from random to a more structured small-world network [9]. There are three possible scenarios. (1) If the threshold is low and the most of the nodes are connected, the averaged shortest path length is long and the clustering coefficient becomes high. Thus, the brain network with low threshold behaves like a random graph. (2) However, if we increase the threshold, at a certain threshold, all edges with long shortest path will be disconnected, but still the clustering coefficients will remain high. Therefore, the brain network with the specific threshold will become a small-world network. (3) If we increase the threshold even futher, the clustering coefficient become lower since the edges with the short shortest path length will also disappear. Therefore, the brain network again return to become more like a random graph. The specific threshold will maximizes the small-worldness of the brain network and it approximately corresponds to the threshold that maximizes the number of clusters.

In this paper, we propose a novel framework for automatically determining the threshold based on the number of clusters in the graph. Since the brain networks are known to be sparse and highly clustered, it is reasonable to incorporate the sparsity of network structures in estimating correlation and partial correlations [10, 6]. Since the partial correlation cannot find the exact solution under the small n large p setting, it is necessary to add the sparseness constraint such as the penalized inverse covariance estimation, [11–14]. We employ the adaptive least absolute shrinkage and selection operator (LASSO) which is the sparse linear regression with the weighted l_1 penalty [15, 16]. It allows us to find the sparse correlation and partial correlation as well as it helps to select the consistent variables under certain conditions [17]. We also impose the discriminability on the correlations using the separating hyperplane of SVM between two groups as the weight in the adaptive LASSO.

For numerical experiments, we have used both the correlation and the partial correlation in two different ways. We have applied the correlation to clustering which provides the global structure of the brain network while the partial correlation is used in classification in estimating the discriminant power of local connections. The proposed methods are applied to the 97 regions of interest (ROIs) extracted from FDG-PET data for 26 autism and 11 pediatric control. Numerical experiments shows that our method finds the proper threshold which reflects the meaningful data structure and discriminates between groups very well.

2 Background

Suppose that we are given data $\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_p\}$, measured at the p selected ROIs on the FDG-PET images. We assume that $\boldsymbol{f}_i$ is normally distributed with mean 0 and variance 1. The correlation coefficient between feature vectors $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ is $\rho_{ij} = \boldsymbol{f}_i^\top \boldsymbol{f}_j$. When the inverse of the covariance matrix is given by $\boldsymbol{\Pi} = [\pi_{ij}] \in \mathbb{R}^{p\times p}$, the partial correlation between $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ while accounting for the confounding effects of all other $(p-2)$ regions is $\theta_{ij} = -\pi_{ij}/\sqrt{\pi_{ii}\pi_{ij}}$.

3 Main Ideas

Our main contribution is to propose a new method for automatically selecting the threshold based on a clustering structure of the graph within the sparse

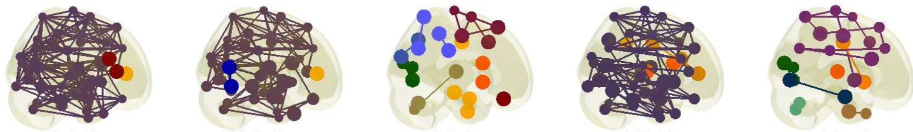

Fig. 1. Depending on the threshold, the structures of the brain networks of autism (a)-(c) and PedCon (d)-(e) are changed. '(a) and (d)' and '(b) and (e)' have the same significant level. '(c) and (e)' have the same number of edges. The numbers are the threshold, the significant level and the number of edges.

regression framework. We incorporate the correlation and partial correlation estimation into the adaptive LASSO regression framework. It allows us a consistent, sparse and discriminative solution under the small n and the large p condition. Within this new sparse regression framework, as main applications, we will consider network clustering and classification problems.

Clustering. A similarity-based graph partitioning helps to understand the global structure of graphs. Correlation is a popular similarity measure which have been used for data clustering and graph partitioning [18, 19]. It is known that the brain network in autism has the local overconnectivity and the global underconnectivity. The clustering of brain network can distinguish the autism and the normal controls in a biologically meaningful way. The sparse linear regression estimation and the proposed thresholding methods make a brain network clustering/partitioning much easier.

Classification. We propose to use partial correlation in group classification by using the residual in the sparse linear regression. Our method can classify the data as well as it finds the local connections which contribute to improving the classification accuracy.

3.1 Thresholding with Maximum Number of Clusters

Fig. 1 shows the brain networks when the thresholds are fixed with the significant levels .01 and .001, and the number of edges is fixed at 35. Although we use the popular significant levels, it is difficult to find the difference between the brain networks (a) and (d) or between (b) and (e). The graphs of autism and PedCon are more different shapes when the number of edges is fixed as shown in (c) and (e); however, we should heuristically select the proper number of edges by observing the graph structure with neurological prior knowledge about their differences.

As we increase the threshold, the number of clusters increase until a certain threshold. Then it decreases making it more or less concave (Fig.2 upper right). This interesting phenomenon is directly related to the small-worldness of graph, which is the ratio between the clustering coefficient and the characteristic path length (Fig. 2 lower right) [20]. This remarkable similarity is no coincidence. If the threshold is sufficiently small, the most nodes are connected to each other, thus making the number of clusters closes to one. When the threshold increases, some connections disappear, and the number of clusters starts to increase. At a certain

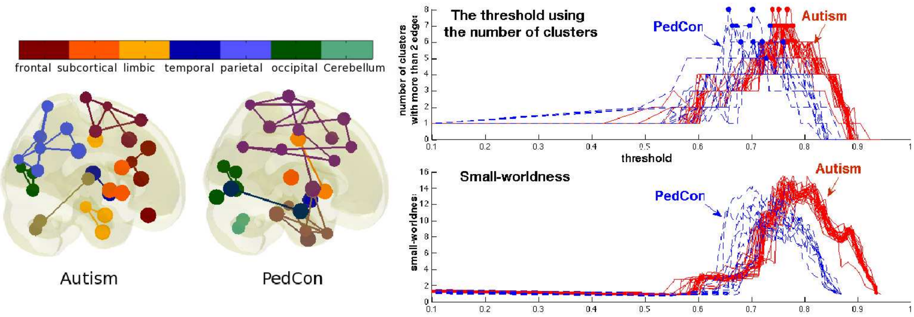

Fig. 2. Brain networks with the maximum number of clusters of autism (left) and PedCon (right). The upper right panel shows that the number of clusters with more than 2 edges varying the threshold and the lower right panel shows that the small-worldness varying the threshold. The red solid line is for 26 different subsets of autism data by jackknifing and the blue dashed line is for 11 subsets of PedCon. The point on each line in the right upper panel means the selected threshold. The edges in the brain networks of autism and PedCon appears with at least with probability 0.7 during jackknifing. The color of nodes indicates different lobes.

threshold, the number of clusters begins to decrease because weakly connected clusters, which contains small number of edges, are removed and only strongly connected clusters may survive. Maximizing the number of clusters means that disconnecting all edges which make the shortest path length long and maximizing the clustering coefficients in each cluster. Therefore, *the threshold which maximizes the number of clusters also likely to maximize the small-worldness* as directly demonstrated in Fig. 2. We propose to use this as the threshold for constructing network groups.

3.2 Sparse Regression

Correlation Coefficients. The correlation coefficients ρ_{ij} can be obtained by the linear regression as follows :

$$\boldsymbol{f}_i = \rho_{ij}\boldsymbol{f}_j + \boldsymbol{\epsilon}_i, \text{ for } i = 1, ..., p. \tag{1}$$

Partial Correlation. It is known that the partial correlation is estimated by the linear regression estimation as follows [21] :

$$\boldsymbol{f}_i = \sum_{j \neq i} \beta_{ij}\boldsymbol{f}_j + \boldsymbol{\epsilon}_i, \text{ for } i = 1, ..., p, \tag{2}$$

where $\boldsymbol{\epsilon}_i$ is uncorrelated with all variables except $\boldsymbol{f}_i$ and β_{ij} is considered as the measure of relationship between the i-th feature and the j-th feature given all other features. Then, the partial correlation θ_{ij} is given by $\theta_{ij} = -\pi_{ij}/\sqrt{\pi_{ii}\pi_{jj}} = \beta_{ij}\sqrt{(\pi_{ii}/\pi_{jj})}$, where $\text{var}(\boldsymbol{\epsilon}_i) = (1/\pi_{ii})$ and $\text{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \pi_{ij}/(\pi_{ii}\pi_{jj})$.

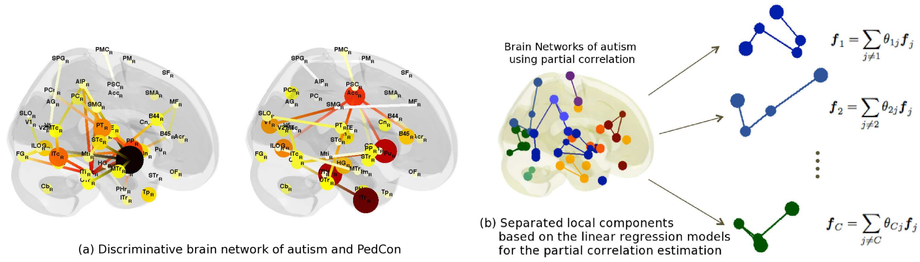

Fig. 3. (a) Discriminative networks of the autism (left) and the PedCon (right) and (b) the partial correlation graph which is divided into small trees.

LASSO. The linear regression (2) does not have the exact solution due to the small n large p problem. We impose the lasso shrinkage into the linear regression given by (2) as follows :

$$\hat{\beta}_{ij} = \arg\min \sum_{i=1}^{p} \parallel \boldsymbol{f}_i - \sum_{j\neq i} \beta_{ij} \boldsymbol{f}_j \parallel^2 + \lambda \sum_{i,j} |\beta_{ij}|, \tag{3}$$

where λ is the regularization parameter. It estimates the biased, but suboptimal solutions [22, 16]. The lasso penalization plays a role as a variable selector as well as it guarantees the consistent solution under certain conditions [17].

Adaptive LASSO. To add discriminability, we introduce the adaptive LASSO [15]. It use the weighted l_1 penalty with parameter γ such as

$$\hat{\beta}_{ij} = \arg\min \sum_{i=1}^{p} \parallel \boldsymbol{f}_i - \sum_{j\neq i} \beta_{ij} \boldsymbol{f}_j \parallel^2 + \lambda \sum_{i,j} |w_i|^{-\gamma} |\beta_{ij}|. \tag{4}$$

The larger the value of w_i is, the less penalized β_{ij} is. We have chosen w_i such that it satisfies $y_t(\boldsymbol{w}^\top \boldsymbol{x}_t - b) \geq 1$, the separating hyperplane in linear SVM, where $\boldsymbol{x}_t \in \mathbb{R}^p$ is a sample data measured in the p ROIs, $y_t \in \{-1, 1\}$ is the corresponding label and $\boldsymbol{w} = [w_1, \ldots w_p]^\top$ is the weights related with the ROIs.

3.3 Classification by Ensemble of Local Connections

The sparse partial correlation matrix and our thresholding method construct a union of the tree-like structured graphs [13]. Fig. 3b shows that the graph can be separated into small trees consisting of parent $\boldsymbol{f}_i$ and their connected children $\boldsymbol{f}_j$ ($j \neq i$ & $\beta_{ij} \neq 0$) with their strength of connection β_{ij} in Eq. (4). If a new data is given, each small tree become a classifier, which estimates a residual and determines whether it fits to the small tree or not. Then whether the data belongs to the graph, which is the collection of the small trees, is determined by an ensemble of the classifiers. In this way, we can classify a new data into a proper group label as well as estimate the discriminative power of each classifier, i.e., each connected brain regions.

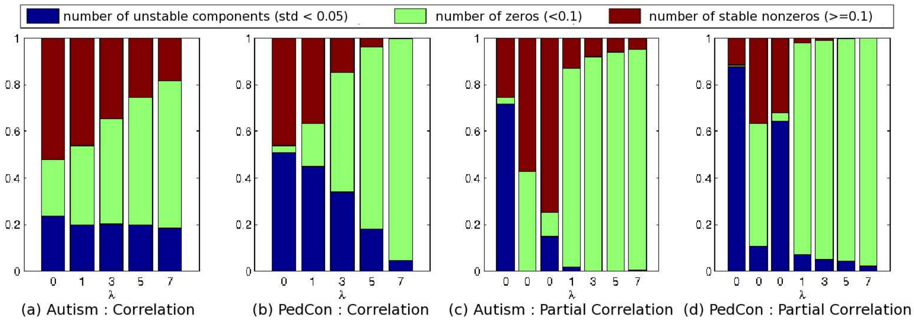

Fig. 4. The ratio of the unreliable elements, the nonzeros and zeros (a) autism and (b) the pediatric control in correlation matrix and (c) autism and (d) the pediatric control in partial correlation matrix . The λ value varied from 0 (standard methods) to 7 in the horizontal axis. The red, green and blue in each image represents the ratio of the number of the unstable elements, zeros and the stable nonzeros.

4 Experimental Results

4.1 Data Acquisition and Preprocessing

The data consists of two groups : 26 autism and 11 pediatric control (PedCon). PET images were preprocessed using Statistical Parametric Mapping (SPM) package. After spatial normalization to the standard template space, mean FDG uptake within 97 ROIs were extracted. The values of FDG uptake were globally normalized to the individual's total gray matter mean count.

4.2 Consistency Check using Jackknife

To verify the consistency of the correlation and partial correlation estimates, we apply the jackknife method which recalculates them using the subset of the sample data leaving out one sample at a time. Thus, we obtain the correlation and the partial correlation matrices as many as the number of trials of jackknife. We checked the consistency by determining whether the standard deviation is less than 0.05 or not, varying λ $(= 0, 1, 3, 5, 7)$. $\lambda = 0$ indicates the standard method without the sparseness constraint.

Our dataset is under high-dimension-small-sample-size setting where the exact partial correlation cannot be obtained properly, therefore, we borrowed 3 kinds of Schäfer's methods: the pseudoinverse of the covariance matrix, the inverse of the ensemble of the sample covariance matrix and the pseudoinverse of the ensemble of the sample covariance matrix, which are obtained by bootstrap aggregation (bagging) [23]. In Fig. 4, we show that the sparseness constraint provides the consistent correlation and partial correlation, reducing the red region which is the ratio of unreliable elements more than 0.05 standard deviation, as well as makes many zeros, increasing the green region. The partial correlation is more sensitive to λ than the correlation due to the small-n large-p problems. We choose $\lambda = 1$ for the simulation.

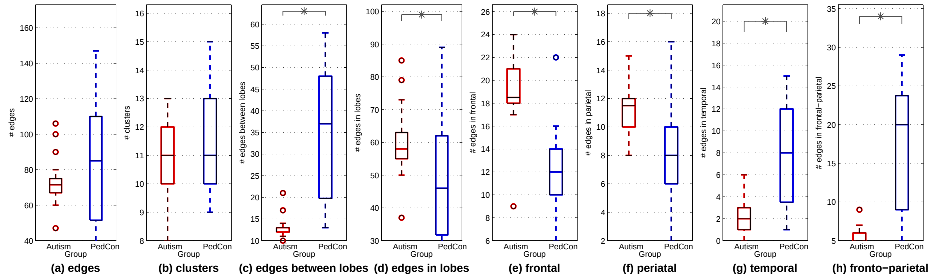

Fig. 5. Comparing (a) the number of edges, (b) the maximum number of clusters, (c) the number of connections between lobes, (d) the number of connections in lobe, the number of connections (e) in frontal lobe, (f) in parietal lobe, (g) in temporal lobe and (h) between frontal and parietal lobe. The red box for autism and the blue box for PedCon. The asterisk (*) represents the significant difference obtained by the Wilcoxon rank sum test.

4.3 Clustering based on Correlation

In Fig. 2, the representative clustered brain networks of autism and the pediatric control are shown. This result supports that the autistic brain network has local overconnectivity and global underconnectivity compared to the normal control. To verify the hypothesis, we performed the Wilcoxon rank sum test on the number of edges, the number of clusters, the number of edges between lobes, the number of edges in lobes, the number of edges in frontal lobe, parietal lobes and temporal lobes and the number of edges between fronto-parietal lobes (Fig. 5).

4.4 Classification based on Partial Correlation

For classification, we partitioned the dataset into training and test data sets, which consist of the randomly chosen two samples, from each group. After training the partial correlations which are coefficients of the p linear equations in Eq. (4), we estimate the p residuals from the regression and classify the data using p residuals. The classification accuracy is 84.44% when $\lambda = 1$ and $\gamma = 0.1$. We also calculate the discriminative power of each connected region (Fig. 3a). Thus, our method can find the local differences between different groups.

5 Conclusion

We proposed a novel thresholding framework in sparse regression on correlations. The threshold is chosen to maximize the number of clusters, which is related with the small-worldness. The sparse linear regression model was introduced for the consistent, sparse and discriminative correlation and the partial correlation estimation. The numerical experiments confirmed that (1) the partial correlation estimated in the linear regression framework can be used for classification, which can find the local differences between groups and (2) the correlation is a good measure for graph partitioning which represents the global difference between groups.

References

1. Cao, J., Worsley, K.: The geometry of correlation fields with an application to functional connectivity of the brain. Annals of Applied Probability **9** (1999) 1021–1057
2. Koch, M., Norris, D., Hund-Georgiadis, M.: An investigation of functional and anatomical connectivity using magnetic resonance imaging. NeuroImage **16** (2002) 241–250
3. Eguiluz, V., Chialvo, D., Cecchi, G., Baliki, M., Apkarian, A.: Scale-free brain functional networks. Physical Review Letters **94** (2005) 018102
4. Worsley, K., Chen, J., Lerch, J., Evans, A.: Comparing functional connectivity via thresholding correlations and singular value decomposition. Philosophical Transactions of the Royal Society B: Biological Sciences **360** (2005) 913
5. Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J., Benalia, H.: Partial correlation for functional brain interactivity investigation in functional mri. NeuroImage **32** (2006)
6. He, Y., Chen, Z., Evans, A.: Small-world anatomical networks in the human brain revealed by cortical thickness from mri. Cerebral Cortex **170** (2007) 2407–2419
7. Salvador, R., Suckling, J., Coleman, M., Pickard, J., Menon, D., Bullmore, E.: Neurophysiological architecture of functional magnetic resonance images of human brain. Cerebral Cortex **15** (2005) 1332–1342
8. McIntosh, A., Bookstein, F., Haxby, J., Grady, C.: Spatial pattern analysis of functional brain images using partial least squares. NeuroImage **3** (1996) 143–157
9. Kleinberg, J.: Navigation in a small world. Nature **406** (2000)
10. Achard, S., Bullmore, E.: Efficiency and cost of economical brain functional networks. PLoS Computational Biology **3** (2007)
11. Sun, L., Patel, R., Liu, J., Chen, K., Wu, T., Li, J., Reiman, E., Ye, Y.: Mining brain region connectivity for alzheimer's disease study via sparse inverse covariance estimation. In: the 15th ACM SIGKDD. (2009)
12. Cecchi, G., Rish, I., Thyreau, B., Thirion, B., Plaze, M., Paillere-Martinot, M., Martelli, C., Martinot, J., Poline, J.: Discriminative network models of schizophrenia. In: NIPS 2009. (2009)
13. Huang, S., Li, J., Sun, L., Liu, J., Wu, T., Chen, K., Fleisher, A., Reiman, E., Ye, J.: Learning brain connectivity of alzheimer's disease from neuroimaging data. In: NIPS 2009. (2009)
14. Huang, S., Li, J., Sun, L., Ye, J., Fleisher, A., Wu, T., Chen, K., Reiman, E.: Learning brain connectivity of alzheimers disease by sparse inverse covariance estimation. NeuroImage **50** (2010) 935–949
15. Zou, H.: The adaptive lasso and its oracle properties. Journal of the American Statistical Association **101** (2006) 1418–1429
16. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B **58** (1996) 267–288
17. Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the lasso. The Annals of Statistics **34** (2006) 1436–1462
18. Böhm, C., Kailing, K., Kröger, P., Zimek, A.: Computing clusters of correlation connected objects. In: SIGMOD 2004. (2003)
19. Bansal, N., Blum, A., Chawla, S.: Correlation clustering. Machine Learning **56** (2004) 89–113
20. Watts, D., Strogatz, S.: Collective dynamics of "small-world" networks. Nature **393** (1998) 440–442

21. Peng, J., Wang, P., Zhou, N., Zhu, J.: Partial correlation estimation by joint sparse regression models. Journal of the American Statistical Association **104** (2009) 735–746
22. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association **96** (2001) 1348–1360
23. Schäfer, J., Strimmer, K.: An empirical bayes approach to inferring large-scale gene association networks. Bioinformatics **21** (2005) 754–764