AN FDR-CONTROLLED, EXPLORATORY GROUP MODELING FOR ASSESSING BRAIN CONNECTIVITY

Aiping Liu[⋆] *Junning Li*[†] *Z. Jane Wang*[⋆] *Martin J. McKeown*[⋆†]

[⋆] Department of Electrical and Computer Engineering, University of British Columbia, Canada, {aipingl, zjanew}@ece.ubc.ca
[†] AUG Signals Ltd, Toronto, Canada, junningl@gmail.com
[⋆†] Department of Medicine (Neurology) and Pacific Parkinson's Research Centre, University of British Columbia, Canada, mmckeown@interchange.ubc.ca

ABSTRACT

Recent development in the inference of brain connectivity from neuroimaging data such as functional magnetic resonance images (fMRI) provides better understanding of brain activities and functions. The group analysis of fMRI data usually focuses on functional connectivity, while exploratory graphical modeling of effective connectivity is generally designed for the single-subject case. In this paper, a new group-level network learning method, the group PC_{fdr} algorithm, is proposed by combining group inference with exploratory graphical modeling for estimating brain connectivity, with inter-subject variances taken into consideration. The simulation results demonstrate that this group method can accurately recover underlying group-level structures and steadily control the false discovery rate at the user specified level. It is also applied to a real fMRI dataset of normal subjects and Parkinson's Disease patients to infer the brain connectivity at the group level.

***Index Terms*—** fMRI, group analysis, effective connectivity, exploratory graphical modeling

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a non-invasive neuroimaging technology that measures relative alterations in deoxygenated hemoglobin as a result of ongoing brain activity. Due to growing recognition of the importance of connectivity between brain regions in normal and disease brain functioning, much effort has been dedicated towards suitable brain connectivity network modeling and analysis. By "connectivity", here we do not simply refer to different brain regions that are covarying, but rather that one brain region has *direct* influence over another.

Various models have been proposed for modeling brain connectivity, such as Structural Equation Modeling [1], Dynamic Causal Models [2], and Dynamic Bayesian Networks [3]. As biomedical research typically studies a group of subjects rather than focusing on an individual subject, group analysis plays an important role in statistical analysis of fMRI data. However, compared with various choices of group-level methods available for the analysis of amplitude and fluctuation of Blood-Oxygen-Level-Dependent (BOLD) signals [4, 5], the problem of group-level effective-connectivity analysis is relatively less well studied, probably due to the fact that it requires accommodating not only the variances and correlations across subjects, but also accounting for the potentially complicated structures of brain connectivity networks.

Currently, most group-level methods for brain connectivity analysis fail to satisfactorily address the efficiency and accuracy concerns. One common practice is to first learn connectivity networks individually for each subject, and then perform group analysis on the subject-level networks. While this approach allows enough freedom for inter-subject variability, the group-level and the subject-level models are not well integrated. Another common practice is to make inference from pooled group data, but this approach ignores inter-subject variability, and is suitable only when the group is relatively homogeneous. We have previously shown that this may be a suitable strategy for fMRI studies involved in control subjects but is unsuitable when dealing with disease populations who tend to be more heterogeneous [6]. Bayesian model selection [7] proposed by Stephan, Penny, etc. in 2009 elegantly handles inter-subject variability and error control, however its current design cannot efficiently handle a large number of networks, making it more suitable for confirmatory, rather than an exploratory research.

In this paper, we propose a group-level exploratory method for brain effective connectivity inference by utilizing an efficient search strategy whilst simultaneously controlling the false discovery rate (FDR). This method is a combination of the PC_{fdr} algorithm and a mixed-effect model, a widely used method for handling inter-subject variability.

Thanks to the CIHR grant (CPN-80080 MJM), the NSERC grant (STPGP 365208-08) and the NSERC fellowship (6715-2009-395708).

2. METHODS

2.1. FDR-controlled exploratory group modeling

Before discussing the algorithms, some useful notations must be first introduced. Acronym "DAG" stands for a directed acyclic graph, that is a graph with directed edges and without directed cycles. V denotes the vertex set for a graph; lowercase a and b denote graph vertices; uppercase A and B denote vertex sets. X_a and X_A denote the variables for vertex a and vertex set A, respectively; Adj(a,G) denote vertices adjacent to a in graph G.

Graphical models usually denote their variables as vertices and variable-wise relationships as edges. In analysis effective connectivity, we can use vertices to represent brain regions, and edges to represent interactions between brain regions.

The PC_{fdr} algorithm developed by Li and Wang [8] (Step 6(a) of Table 1) is capable of asymptotically controlling the False Discovery Rate (FDR) under pre-specified levels [8]. (Interested readers may refer to [8] for an extensive discussion.) The FDR is one of the important error control criteria for multiple testing. Compared with the traditional type I and type II error rates, FDR is more reasonable in some applications such as bioinformatics and neuroimaging since it is directly related with the uncertainty of the discovered positive results.

Here we propose extending the PC_{fdr} algorithm from the single subject-level to the group-level by considering a mixed-effect model (Step 6(b) of Table 1). Suppose we have m subjects in a group. Then for subject i, the conditional independence between the activities of two brain regions a and b given other regions C can be measured by the partial correlation coefficient between $X_a(i)$ and $X_b(i)$ given $X_C(i)$, denoted as $r_{ab|C}(i)$. Here the index i indicates that these variables are for subject i. By definition, the partial correlation coefficient $r_{ab|C}(i)$ is the correlation coefficient between the residuals of projecting $X_a(i)$ and $X_b(i)$ onto $X_C(i)$, and can be estimated by the sample correlation coefficient as

$$\hat{r}_{ab|C}(i) = \frac{Cov[Y_{a|C}(i), Y_{b|C}(i)]}{\sqrt{Var[Y_{a|C}(i)]Var[Y_{b|C}(i)]}}, \tag{1}$$

where

$$\beta_{a|C}(i) = \arg\min_{\beta} |X_a(i) - X_C(i)\beta|^2, \tag{2}$$

$$\beta_{b|C}(i) = \arg\min_{\beta} |X_b(i) - X_C(i)\beta|^2, \tag{3}$$

$$Y_{a|C}(i) = X_a(i) - X_C(i)\beta_{a|C}(i), \tag{4}$$

$$Y_{b|C}(i) = X_b(i) - X_C(i)\beta_{b|C}(i). \tag{5}$$

To focus on the group-subject relationship, in the following discussion we omit the subscript "$ab|C$", and simply use index "i" to emphasize that a variable is associated with subject i. To study the group-level conditional-independence relationships, a group-level model should be introduced for r_i.

Table 1: The PC_{fdr} algorithm in [8] and the proposed group-level method.

1: Form the complete undirected graph G, and set search depth d = 0.
2: **repeat**
3: **repeat**
4: Select an ordered pair of variable vertices a and b such that they are adjacent in G and Adj$(a, G)\backslash b$ has at least d vertices in it.
5: **for** vertex subset $C \subseteq$ Adj$(a, G)\backslash b$, and $|C| = d$ **do**
6: (a) test hypothesis $X_a \perp X_b | X_C$ and calculate the p-value for edge $a - b$ at the **subject level**, **or**
(b) test hypothesis $X_a \perp X_b | X_C$ and calculate the p-value for edge $a - b$ at the **group level**.
7: Input edge p-value to FDR control procedure and judge if any edge should be removed, and update G.
8: **if** the edge between a and b is removed **then**
9: Break loop at step 5.
10: **end if**
11: **end for**
12: **until** all existing edges have been tested.
13: $d = d + 1$.
14: **until** Cannot find a triple (a, b, C) for step 5.

In Step 6, either (a), or (b) is executed. (a) is the original subject-level PC_{fdr} algorithm in [8], for controlling the FDR at the subject level; (b) is the proposed method, for controlling the FDR at the group level.

Since partial correlation coefficients are bounded and their sample distributions are not Gaussian, we apply Fisher's z-transformation to convert (estimated) partial correlation coefficients r to a Gaussian-like distributed z-statistic z, which is defined as

$$z = Z(r) = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right), \tag{6}$$

where r is a (estimated) partial correlation coefficient, and z is its z-statistic.

The group model we employ is

$$z_i = z_g + e_i, \tag{7}$$

where e_i follows a Gaussian distribution $N(0, \sigma_g^2)$ with zero mean and σ_g^2 variance. Consequently, the group-level testing of conditional independence is to verify whether the null hypothesis $z_g = 0$ is true or not.

Because z_i is unknown and can only be estimated, the inference of z_g should be conducted with $\hat{z}_i = Z(\hat{r}_i)$. If $X_a(i)$, $X_b(i)$ and $X_C(i)$ jointly follow a multi-variate Gaussian distribution, then $\hat{z}_i$ asymptotically follows a Gaussian distribution $N(z_i, \sigma_i^2)$ with $\sigma_i^2 = 1/(N_i - p - 3)$, where N_i is the

sample size of subject i's data and p represents the number of variables in $X_C(i)$. Therefore, based on Eq. (7), we have

$$\hat{z}_i = z_g + e_i + \epsilon_i, \tag{8}$$

where ϵ_i follows $N(0, \sigma_i^2)$, and e_i follows $N(0, \sigma_g^2)$.

This is a mixed-effect model where ϵ_i denotes the within-subject randomness and e_i denotes the inter-subject variability. At the group level, $\hat{z}_i$ follows a Gaussian distribution $N(z_g, \sigma_i^2 + \sigma_g^2)$. Note that unlike regular mixed-effect models, the within-subject variance ${\sigma_i}^2$ in this model is known, because N_i and p are known given the data $X(i)$ and C. In general, $\sigma_i^2 = 1/(N_i - p - 3)$ is not necessarily equal to σ_j^2 for $i \neq j$, and the inference of z_g should be conducted in the manner of mixed-models, such as estimating σ_g^2 with the Restricted Maximum Likelihood (ReML) approach. However, if the sample size of each subject's data is the same, then σ_i^2 equals σ_j^2. In this balanced case, which is typically true in fMRI applications and as well the case in this paper, we can simply apply a t-test to $\hat{z}_i$'s to test the null hypothesis $z_g = 0$. We employed it in our simulations and the analysis on the fMRI data presented in this paper. Replacing Step 6(a) of the single-subject PC_{fdr} algorithm (i.e. the within-subject hypothesis test) with the test of $z_g = 0$, we can extend the single-subject version of the algorithm to its group-level version. Such a testing approach significantly simplifies the estimation process and our simulation results presented later demonstrates that this method could still control the FDR at a user specified error rate level.

2.2. Simulations

The simulations in this section served to verify whether the proposed group PC_{fdr} algorithm for modeling brain connectivity can control the FDR at the group level; and to compare its performance with that of single-subject PC_{fdr} algorithm by pooling data together[8].

The simulations were conducted as follows.

(1) Randomly generate a directed acyclic graph (DAG) as the group-level network, and associate each connection with a coefficient. The connection coefficients were assigned as random samples from the uniform distribution $U(\beta_1, \beta_2)$, where β_1 and β_2 characterized the coefficient strength.

(2) Subject-level networks were derived from the group-level network by randomly adding and deleting connections. More specifically, for each of the existing connections, the connection was deleted with probability 0.05, and for each of the absent connections, a connection was added with probability 0.01. The corresponding connection coefficients were randomly sampled from the uniform distribution $U(\beta_1, \beta_2)$.

(3) Given a subject-level network, the subject-specific data was generated from a Gaussian Bayesian network, with the additional Gaussian noise following the standard Gaussian distribution of $N(0, 1)$.

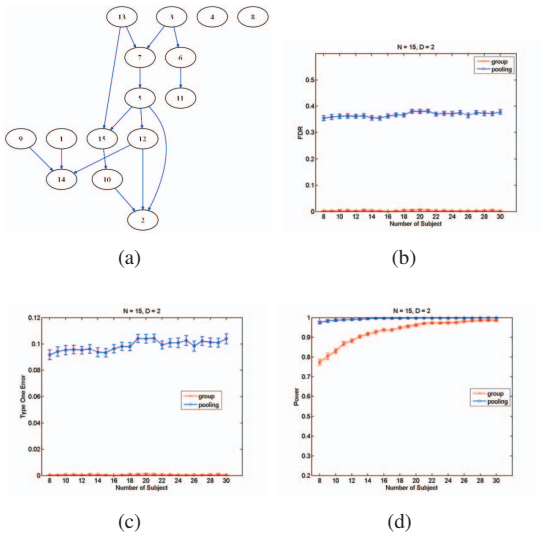


Fig. 1: Simulation: Assessing the effects of increasing the number of subjects on learned networks. (a) The group-level network, with 15 nodes and an average of two connections per node. (b) The FDR curves (with standard deviation marked) of the group-level PC_{fdr} algorithm and the single-subject PC_{fdr} algorithm. (c) The Type I error rate curves. (d) The power curves. The x-axis represents the number of subjects within the group.

In the simulation, we tested the performances of the algorithms as a function of the number of subjects within the group. The group-level network was the DAG in Figure 1(a), and the number of subjects increased from eight to thirty. At the network learning stage, we set the target FDR to be 5%. Figure 1(b) demonstrate the FDR results as a function of the number of subjects. The proposed group PC_{fdr} algorithm was able to keep the FDR below the specified level. The detection powers estimated in Figure 1(d) gradually increased as the number of subjects increased.

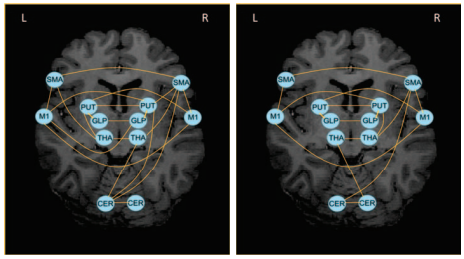
The simulation results here show that the proposed method is able to accurately discover the underlying group network and steadily control the false discovery rate at a user-specified level. The group-level PC_{fdr} method is shown to be much more reliable than simply pooling together the data from all subjects.

3. APPLICATION TO FMRI DATA

We apply group PC_{fdr} algorithm to a real fMRI dataset collected from 10 PD patients and 9 age-matched controls to learn brain connectivity. The 10 PD subjects had typical Parkinson's symptoms with no other neurological or psychiatric conditions. All experiments were approved by the University of British Columbia Ethics Committee. The subjects were asked to perform a bulb squeezing task. All of

them were instructed to squeeze the bulb with their right hand to control the width of an inflatable ring which was shown as a horizontal black bar in the screen.

We chose 12 brain regions of interest (ROIs), including L/R cerebellar hemispheres (l/rCER), L/R globus pallidus (l/rGLP), L/R Putamen (l/rPut), L/R supplementary motor area (l/rSMA), L/R thalamus (l/rTHA) and L/R primary motor cortex (l/rM1). The connectivity networks were computed by the proposed group PC_{fdr} algorithm (with the target FDR being 5%) for the normal and PD groups.



(a) (b)

Fig. 2: (a) Learned effective connections for the normal group. (b) Learned effective connections for the PD group. Here "L" and "R" refer to the left and right sides respectively.

The connectivity networks learned by the proposed group PC_{fdr} algorithm (with the target FDR of 5%) for the normal and PD groups are compared in Figure 2. Comparing the group networks from the PD subjects to that of the control group, the following connections were lost: the connectivity between the supplementary motor area (lSMA) and the thalamus (lTHA) contralateral to the hand used in the task and the contralateral primary motor cortex (lM1) to the ipsilateral globus pallidus (rGLP) (part of the expected striatothalamocortical loop) and the connection from the contralateral cerebellar hemisphere (lCER) to the ipsilateral thalamus (rTHA), (part of the cerebellothalamocortical loop). However, some connections were only detected in the PD group, including connections between the contralateral cerebellar hemisphere (rCER) and the ipsilateral supplementary motor area (rSMA), and between the ipsilateral cerebellar hemisphere (rCER) and ipsilateral thalamus (rTHA).

4. DISCUSSION AND CONCLUSION

In this paper, the group PC_{fdr} algorithm is proposed for inferring brain connectivity, with inter-subject variances taken into consideration. The simulation results show that this new method is able to accurately discover the underlying group structures and steadily control the FDR at the user specified level. The networks learned with this method are far more reliable than the result of pooling subject data together. The proposed group PC_{fdr} method is likely more suitable for sparse network inference, as empirically we observed that increasing the density of the group network may lead to a decrease in detection power. Nevertheless, it is an attempt of error-rate-controlled group-level modeling of brain effective connectivity.

Utilizing the proposed method, we found connections that were unique to PD subjects not seen in the normal subjects. Consistent with prior studies, these new connections detected by the proposed method likely represent compensatory recruitment [9], although it can be often difficult to distinguish direct and indirect disease-related changes from compensatory ones.

5. REFERENCES

[1] AR McIntosh, CL Grady, LG Ungerleider, JV Haxby, SI Rapoport, and B Horwitz, "Network analysis of cortical visual pathways mapped with pet," *The Journal of Neuroscience*, vol. 14, no. 2, pp. 655–666, 1994.

[2] K.J. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273 – 1302, 2003.

[3] Junning Li, Z.J. Wang, J.J. Eng, and M.J. McKeown, "Bayesian network modeling for discovering dependent synergies among muscles in reaching movements," *Biomedical Engineering, IEEE Transactions on*, vol. 55, no. 1, pp. 298 –310, jan. 2008.

[4] K. J. Worsley, C. H. Liao, J. Aston, V. Petre, G. H. Duncan, F. Morales, and A. C. Evans, "A general statistical analysis for fmri data," *NeuroImage*, vol. 15, no. 1, pp. 1 – 15, 2002.

[5] K.J. Friston, K.E. Stephan, T.E. Lund, A. Morcom, and S. Kiebel, "Mixed-effects and fmri studies," *NeuroImage*, vol. 24, no. 1, pp. 244 – 252, 2005.

[6] Junning Li, Z. Jane Wang, Samantha J. Palmer, and Martin J. McKeown, "Dynamic bayesian network modeling of fmri: A comparison of group-analysis methods," *NeuroImage*, vol. 41, no. 2, pp. 398 – 407, 2008.

[7] Klaas Enno Stephan, Will D. Penny, Jean Daunizeau, Rosalyn J. Moran, and Karl J. Friston, "Bayesian model selection for group studies," *NeuroImage*, vol. 46, no. 4, pp. 1004 – 1017, 2009.

[8] Junning Li and Z. Jane Wang, "Controlling the false discovery rate of the association/causality structure learned with the pc algorithm," *J. Mach. Learn. Res.*, vol. 10, pp. 475–514, June 2009.

[9] S.J. Palmer, J. Li, Z.J. Wang, and M.J. McKeown, "Joint amplitude and connectivity compensatory mechanisms in parkinson's disease," *Neuroscience*, vol. 166, no. 4, pp. 1110 – 1118, 2010.