CONTROLLING THE FALSE DISCOVERY RATE IN MODELING BRAIN FUNCTIONAL CONNECTIVITY

Junning Li, Z. Jane Wang

Department of Electrical and Computer Engineering
University of British Columbia
Canada

Martin J. McKeown *

Pacific Parkinson's Research Centre
Department of Medicine (Neurology)
Brain Research Centre
University of British Columbia
Canada

ABSTRACT

Graphical models of brain functional connectivity have matured from confirming *a priori* hypotheses to an exploratory tool for discovering unknown connectivity. However, exploratory methods must control the error rate of "discovered" connectivity networks. Here we explore an error-rate-control method for graphical models which controls the false-discovery-rate (FDR) of the conditional-dependence relationships that a graphical model encodes. The application of this method to a group analysis of fMRI study on Parkinson's disease shows that it effectively controls the errors introduced by randomness, and yields meaningful and consistent results. The proposed approach appears promising for functional-connectivity modeling and deserves further investigation.

Index Terms— graphical model, false discovery rate, brain connectivity, functional magnetic resonance imaging (fMRI)

1. INTRODUCTION

Graphical models, such as structural equation models (SEM) [1], dynamic causal models (DCM) [2] and (dynamic) Bayesian networks (BN) [3], have attracted increasing attention in the field of modeling brain connectivity. These models generally represent the connectivity between brain regions as networks, and approximate the random dynamic interactions with certain regular stochastical processes. SEMs were first introduced [1], in an application to Positron Emission Tomography (PET), as a tool to validate a predefined connectivity network. After being used as a confirmatory tool for a long time, SEMs were extended to be an exploratory tool to discover unknown brain connectivity, in applications to functional magnetic resonance imaging (fMRI) [4, 5]. The usage of DCMs had a similar trend as that of SEMs, shifting from a confirmatory tool [2] to an exploratory tool in [6]. Another graphical model, the Bayesian network, was introduced in this field first as an exploratory tool [7, 8]. Recent literature has shown that applications of graphical models to brain connectivity have been more and more oriented for the purpose of unveiling unknown connectivity.

However, current learning algorithms of graphical models have not been adequately adapted to concerns regarding the error rate of claimed "discovered" networks. A desirable model is not only one which fits data well, but also accounts for the error rate of graphical features of the connectivity network, for example, the existence of a certain connection between two brain regions. Structure-learning methods for graphical models can be divided into three broad categories: (1) score-based searching methods which look for a suitable structure according to a certain criterion of goodness-of-fit, such as the Bayes factor [6], the Bayesian information criterion (BIC) [5], or the parsimony goodness-of-fit index (PGFI) [4]; (2) the Bayesian approach, which estimates the posterior probabilities of a set of candidate network structures and infers the error rate from the posterior probabilities [8]; and (3) methods based on conditional-independence (CI) tests which encode a set of tested conditional-independence relationships as graphs according to certain rules such as Markov properties [7]. Unfortunately, score-based methods do not allow explicit inferences on specific graphical features. The Bayesian approach is theoretically promising, but exact inferences demand intensive and impracticable computation, especially for large networks. CI-based methods control the error rate of each CI test, but the issue of simultaneously testing multiple CI hypotheses has not been addressed appropriately. Therefore, in this paper, we improve CI-based methods by controlling the error rate of their multiple testing according to the false discovery rate [9].

Graphical models are theoretically founded on graphical encoding of CI relationships among random variables [10]. As CI relationships are the backbone of graphical models, it is natural to construct a graphical model from a set of CI relationships. Several algorithms, named as "discovery algorithms" in [11], are available for this purpose, such as Spirtes, Glymour, and Scheines' SGS algorithm. Though these algo-

*This work was supported by NSERC/CIHR CHRP-(323602 - 06)

rithms construct different types of graphs, their first steps are the same: constructing an undirected graph from a set of CI relationships. The undirected graph will then be annotated into different types of graphs with different procedures, but the skeleton of the final output graph is kept the same as that of the undirected graph.

In this paper, we propose using a false-discovery-rate procedure to control the error rate: we first select a set of conditional-dependence (CD) relationships, with the false positive rate of the selected CD relationships controlled under a certain threshold such as 5%, and then input the selected CD relationships into these CI-based graph-construction algorithms. Because graphical models are graphical encodings of CI relationships, and CD and CI are complementary, controlling the error rate of the CD relationships a graphical mode encodes will also control the error rate of its structure. As inter-subject variability is a big concern in studies involving multiple subjects, we also propose a statistical procedure for group analysis, extracting a connectivity network that consistently appears in the connectivity networks of the subjects within the same group. An application of the proposed method to a real fMRI study on Parkinson's disease will be discussed.

2. METHODS

2.1. Conditional-Independence Tests

Conditional independence (CI) is the key concept behind graphical models. Two sets of random variables A and B are conditionally independent upon a third set of random variables S if and only if $P(A, B|S) = P(A|S)P(B|S)$, and the conditional independence is denoted as $A \perp B|S$. A basic modular of the CI-based algorithms is to decide whether two random variables a and b are conditionally independent or not upon a third set of random variables S. This is usually implemented by hypothesis testing. For Gaussian continuous variables, partial correlation and t-test are the most widely used. To represent dependence in a more general sense, such as nonlinearity, here we do not assume Gaussian distributions or linearity, but discretize the originally continuous data into ordinal categorical data, and apply a variation of the Cochran-Mantel-Haenszel (CMH) test [12] defined as follows to test CI:

Let a, b and s denote three categorical variables, of I, J and K levels, respectively. a and b are further considered as ordinal variables, with their levels being assigned with ordered scores u_i (i = 1, 2, ... I) and v_j (j = 1, 2, ... J). An observation of the joint distribution of the three variables is a $I \times J \times K$ contingency table whose element n_{ijk} is the count of event $\{a = i \wedge b = j \wedge s = k\}$. Conditional independence $a \perp b|s$ can be tested with the M^2 statistic defined in Eq. (1), where $n_{i\cdot k}$, $n_{\cdot jk}$ and $n_{\cdot\cdot k}$ denote sums of n_{ijk} over the dimension(s) indicated by the subscript "$\cdot$". M^2 asymptotically follows a χ^2_1 distribution if $a \perp b|s$.

$$M^2 = \frac{\Big[\sum_{k=1}^{K}(\sum_{i=1}^{I}\sum_{j=1}^{J} u_i v_j n_{ijk} - E_k)\Big]^2}{\sum_{k=1}^{K} V_k}, \quad (1a)$$

$$E_k = \Big(\sum_{i=1}^{I} u_i n_{i\cdot k}\Big)\Big(\sum_{j=1}^{J} v_j n_{\cdot jk}\Big)/n_{\cdot\cdot k}, \quad (1b)$$

$$V_k = \frac{1}{n_{\cdot\cdot k}-1}\Big[\sum_{i}^{I} u_i^2 n_{i\cdot k} - \frac{(\sum_i^I u_i n_{i\cdot k})^2}{n_{\cdot\cdot k}}\Big] \Big[\sum_{j}^{J} v_i^2 n_{\cdot jk} - \frac{(\sum_j^J v_j n_{\cdot jk})^2}{n_{\cdot\cdot k}}\Big]. \quad (1c)$$

2.2. Discovery Algorithms

Discovery algorithms, such as the SGS algorithm, learn different types of graphs from CI relationships, but the skeletons of their final output graphs share the same rule of encoding CI relationships, *i.e.* if a and b are conditionally independent upon a certain set of variables S excluding a and b, then there is no edge between the vertices representing a and b. This rule can be implemented with algorithms of different computational efficiency, but all algorithms are to achieve the same goal: the encoding rule. To avoid distracting readers from the topic of error-rate control, we just present the most straight forward method that the SGS algorithm employs [11]: first, form an completely connected undirected graph G_f for the variable set V; second, for each pair of variables a and b, remove the edge between the vertices representing them if and only if there is a set of variables $S \subseteq V \setminus \{a, b\}$ such that $a \perp b|S$.

Testing $a \perp b|S$ with the CMH test will become inaccurate if the contingency table is sparse. Unfortunately this easily occurs in practice because the size of the contingency table increases exponentially as the number of the variables in S increases. One solution is to replace the CMH test with exact CI tests [12], however the intensive computation that exact tests demand restricts them from large scale applications. Thus, in practice we just test those CI relationships $a \perp b|S$ where the size of S is equal to or smaller than a certain threshold m, and simply assume conditional dependence if $|S|$ is larger than m. The algorithm we used is outlined as follows:

1. enumerate all the variable triples $\{a, b, S \mid S \subseteq V \setminus \{a, b\}$ and $|S| \leq m\}$.
2. test the conditional independence of all the triples enumerated in step-1 with the CMH test.
3. select a set of triples and regard them as conditional dependence with a FDR procedure (see Section 2.3) which controls the error rate of the selection lower than a certain threshold such as 5%.
4. remove the edge between the vertices representing a and b if there is a $S \subseteq V \setminus \{a, b\}$ such that $a \perp b|S$.

2.3. False Discovery Rate

When many hypotheses are tested simultaneously, the effect of multiple testing should be corrected accordingly. For example, if a researcher wants to control the probability that at least one of H independent hypotheses is falsely rejected to be less than 5%, (s)he should set the significance level of each individual test at $1-(1-5\%)^H$ but not at 5%. If (s)he sets the threshold at 5%, then the expected number of falsely rejected hypotheses will be $H5\%$, even when none of the hypotheses actually is wrong. The false discovery rate (FDR) is an error-rate criterion of multiple testing, defined as the expected ratio of falsely rejected hypotheses to all those rejected. Referring to Table 1, the FDR is formally defined as FDR = $P(\text{ TP + FP } > 0)E(\text{ FP } / \text{ TP + FP } | \text{ TP + FP } > 0)$. The FDR can be controlled to be lower than q by following the step-down procedure [9]. The minimum FDR threshold at which a hypothesis is rejected is called the test's q-value.

1. Sort the p-values of H hypothesis tests in ascendant order as $p_1 \leq \ldots \leq p_H$.
2. Find the largest k such that $p_k \leq \frac{k}{H}q$.
3. Reject hypotheses $1 \ldots k$.

test \ truth	positive	negative
positive	TP (true positive)	FP (false positive)
negative	FN (false negative)	TN (true negative)

Table 1. The counts of multiple testing's results, categorized according to the test results and the truth.

2.4. Group Analysis

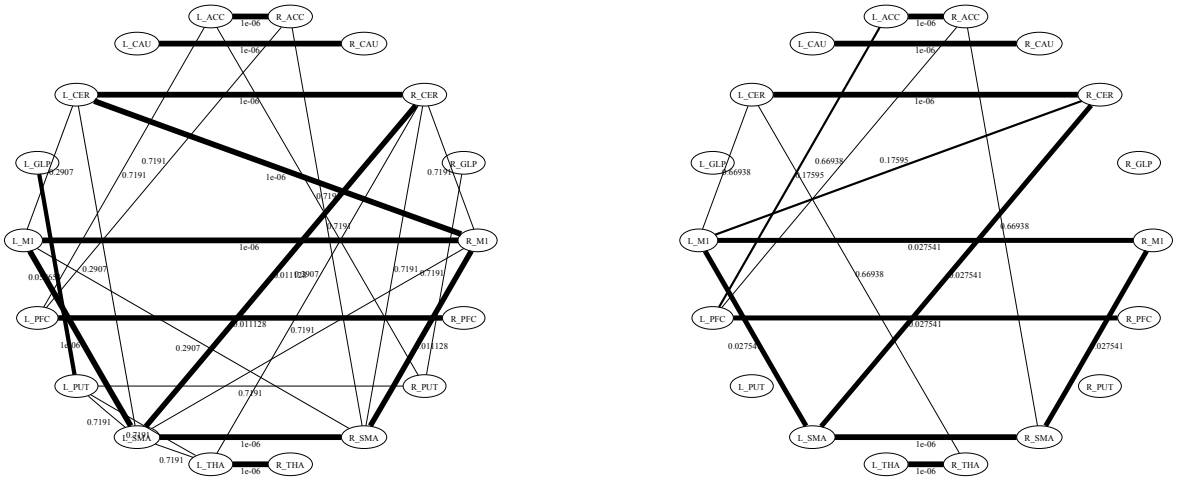
FMRI studies are usually performed to infer features shared by a group of subjects rather than subject-specific features. Whereas, we plan to learn a connectivity network for each subject individually to accommodate the differences among subjects, rather than one network for all subjects together. To extract the connectivity network that is consistently shared by group members, we propose the following group-analysis method.

Consistent sharing can be defined in at least two ways: absolute consistence and relative consistence. Suppose we have N undirected graphs $\{G_n\}$. If the appearance frequency of a certain connection among the N graphs is statistically significantly higher than a certain threshold, such as 80%, the connection is considered to be absolutely consistently recruited by the graphs. If the appearance frequency is statistically significantly higher than a result of randomness, the connection is considered to be relatively consistently recruited. Absolute consistence does not necessarily imply relative consistence, and vice versa. In most fMRI studies, the limited number of available subjects usually restricts the power of the inferences based on absolute consistence, so here we develop a group-analysis method based on relative consistence.

Suppose the N graphs contains C_n $(n = 1, \ldots, N)$ connections respectively and the total number of possible connections of a graph is C. If all the connections are randomly recruited with equal chance, then the number of a connection's appearances is a random number $Y = \sum_{n=1}^{N} X_n$ where $X_n \sim$ Bernoulli(C_n/C). The probability that a connection appears equal to or more than y times as a result of randomness is $P(Y \geq y)$. We applied the test on relative consistence to all the possible connections, and then adjusted the effect of multiple testing with the FDR procedure, and finally selected those whose q-values are lower than 5% as relatively consistent connections to compose a connectivity network at the group level.

3. APPLICATION TO PARKINSON'S DISEASE

Data Collection: The study was approved by the University of British Columbia ethics board and all subjects gave written informed consent prior to participating. Ten subjects with clinically diagnosed Parkinson's disease participated in the study. While in the fMRI scanner, subjects were instructed to squeeze a rubber bulb with their right hand at four frequencies (0.00Hz, 0.25Hz, 0.5Hz and 0.75Hz) in 30s blocks, arranged a pseudo-random order. The patients performance the experiment twice, once before medication and the other after medication. fMRI data of their brain activities during performing the task was collected with a Philips Achieva 3.0 T scanner. The following eighteen brain regions were selected as the regions of interest (ROI) in the study – the left and right: primary motor cortex (M1), supplementary motor cortex (SMA), lateral cerebellar hemisphere (CER), putamen (PUT), caudate (CAU), thalamus (THA), prefrontal cortex (PFC), anterior cingulate cortex (ACC), and globus pallidus (GLP).
Preprocessing: The raw fMRI time courses of the voxels within each ROI were averaged as the summary activity of the ROI. Then, the averaged time courses were detrended and normalized to unit variance. Finally, the continuous time courses were discretized into three ordinal categories of -1, 0 and 1 with thresholds at the lower and upper quartiles.
Results: The connectivity networks of the patients before and after medication are shown in Figure 1. The two figures are obviously not results of randomness, because most of the connections with their q-values less than 5% appear in both of the graphs. For example, the connections between the left and the right counterparts of ACC, CAU, CER, M1, PFC, SMA and THA, between L_M1 and L_SMA, between R_M1 and R_SMA, and between L_SMA and R_CER. Since the networks were learned from the data of the same group of patients, it was reasonable for the two networks to share certain similarity. This also suggests that controlling the error rate yields robust results, curbing the effect of randomness. The connections L_CER—R_M1 and L_GLP—L_SMA appear in Figure 1(a) with the q-values less than 10% but not in Figure 1(b) as their q-values are 1 in the analysis after medication.



(a) Before Medication (b) After Medication

Fig. 1. The connectivity networks of the Parkinson's disease patients, learned with $m = 4$. Edge labels are the q-values of the edges in the group analysis. Prefixes "L" and "R" before ROI names are short for "left" and "right" respectively.

The results confirm the compensatory recruitment of the contralteral motor circuit (right M1, left cerebellar hemisphere) in performing this right handed task in Parkinson's subjects off of medication. Additionally, they also demonstrate the effects of L-dopa medication: reduction in the contralateral motor circuit and re-emergence of the normal left M1, right cerebellar circuit.

4. CONCLUSIONS AND DISCUSSIONS

Graphical models have been increasingly investigated as an exploratory tool for discovering brain connectivity by using brain-imaging data. It is critical to control the error rate in the "discovered" connectivity network in real biomedical applications. Statistically rigorous methods are possible to be developed for this purpose. The method we developed based on conditional-independence (CI) tests and the false discovery rate (FDR) showed promising performance in a study on Parkinson's disease, where the learned connectivity networks were consistent with known biological knowledge about the effects of L-dopa medication in Parkinson's disease.

However, the proposed method does not control the error rate directly at the level of graphical features, such as the existence of certain edges, but indirectly at the level of CI relationships. Further improvement is achievable because according to the theory of graphical models the existence of a certain edge is a result of the rejection of a set of CI relationships, which is also a problem of multiple testing.

Finally, we note that the modular feature of this method provides the ability to easily integrate different data sources. CI relationships can be tested on data collected in different laboratories. All the p-values of these hypothesis tests can then be pooled together to learn a connectivity network.

5. REFERENCES

[1] A. R. McIntosh and F. Gonzalez-Lima, "Structural equation modeling and its application to network analysis in functional brain imaging," *HUM BRAIN MAPP*, vol. 2, no. 1-2, pp. 2–22, 1994.

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

[3] Xuebin Zheng and Jagath C. Rajapakse, "Learning functional structure from fMR images," *NeuroImage*, vol. 31, no. 4, pp. 1601–1613, July 2006.

[4] Jiancheng Zhuang, Stephen LaConte, Scott Peltier, Kan Zhang, and Xiaoping Hu, "Connectivity exploration with structural equation modeling: an fMRI study of bimanual motor coordination," *NeuroImage*, vol. 25, no. 2, pp. 462–470, Apr. 2005.

[5] Amos J. Storkey, Enrico Simonotto, Heather Whalley, Stephen Lawrie, Lawrence Murray, and David McGonigle, "Learning structural equation models for fMRI," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds., pp. 1329–1336. MIT Press, Cambridge, MA, 2007.

[6] W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston, "Comparing dynamic causal models," *NeuroImage*, vol. 22, no. 3, pp. 1157–1172, July 2004.

[7] F. F. Deleus, P. A. De Maziere, and M. M. Van Hulle, "Functional connectivity modelling in fMRI based on casual networks," in *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*, 2002, pp. 119–128.

[8] Jagath C. Rajapakse and Juan Zhou, "Learning effective brain connectivity with dynamic bayesian networks," *NeuroImage*, vol. 37, no. 3, pp. 749–760, Sept. 2007.

[9] Yoav Benjamini and Yosef Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, vol. 57, no. 1, pp. 289–300, 1995.

[10] J. T. A. Koster, "Markov properties of nonrecursive causal models," *ANN STAT*, vol. 24, no. 5, pp. 2148–2177, oct 1996.

[11] Peter Spirtes, Clark Glymour, and Richard Scheines, *Causation, Prediction, and Search*, chapter 5, pp. 82–92, 139–140, The MIT Press, 2001.

[12] Alan Agresti, *Categorical Data Analysis*, chapter 7, pp. 293–298, John Wiley & Sons, Inc., 2002.