Group representative brain connectivity model of episodic encoding using large fMRI dataset

Supat Saetia, Natsue Yoshimura, Yasuharu Koike

Department of Information Processing, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8503, JAPAN

E-mail: saetia.s.aa@m.titech.ac.jp

Abstract. Episodic memory is autobiographical memory crucial to human cognitive function. Brain connectivity modeling is an essential tool for describing brain function of interest. While a large body of research has focused on analyzing functional brain connectivity, most studies has focused on a connectivity of an individual. Brain connectivity studies are usually based on functional Magnetic Resonance Imaging (fMRI) which often involved many subjects. Only a few researches have focused on group level connectivity. We employ a framework called Tigramite to estimate causal interaction and model episodic memory connectivity on Human Connectome Project (HCP) BOLD data. Regions of interest are identified by performing fMRI univariate group subtraction analysis on working memory task-fMRI data using subsequent memory test paradigm. From each resulting individual connectivity graphs, we construct a group representative graph by performing median aggregation on each corresponding edges and nodes of each individual graphs. However, there is no established method to measure a quality of the group representative graph. So, we propose the use of Levenshtein distance as a measurement to validate the group representative graph if it represents general characteristics of each individual graphs in the population.

1. Introduction

Writing is one of the greatest invention of humanity, it enables us to physically record our ideas or experiences we have encountered, then to pass it on to others or to the next generation. Modern day human has several means of storing information either physically, or even digitally. However, for most of humanity’s existence, it has been only a short period of time since the invention of writing. Before our means of recording information was invented, all human had was their brain, a biological storage which stores information called memory. The brain is our most primitive means of information storing, yet slightly we know of how information, our memories, are organized inside our brains.

Episodic memory is a memory of personal events times, places, associated with emotion and other contextual knowledge (autobiographical events). With episodic memory, a person can remember personal experienced events and consciously aware of the certain situation at a certain time of that events [1]. It is a part of a bigger category of long-term memory called declarative memory [2]. The episodic memory is different from other types of memory in that it often represented in form of visual information in order of event occurrence on personal timeline, and it needs to have perspective observer of the memory owner. The importance of episodic memory is that it enables human to project oneself backward
in time and recall many aspects of one previous experience. This ability becomes the source of self-awareness and induces intelligence which set human apart from other animals.

Observing how information are organized in human brain is a challenging task. However, in recent years, the progress in functional Magnetic Resonance Imaging (fMRI) technology has allowed the study of brain connectivity, the statistical dependencies between localized regional brain activity. In this context, we can model how information flows in the brain by analyzing brain connectivity during the memory-related cognitive task. Brain connectivity can be visualized in a graphical model [3] or time series graph [4].

Granger causality [5] is a mathematical framework commonly employed to model causality of the neuronal activity from fMRI BOLD data. The underlying assumption of this framework is that if \( X \rightarrow Y \) if and only if a change in \( X \) has an effect on \( Y \) [6]. However, all we can imply from observational data are statistical dependencies. It is controversial to infer effective connectivity (directed connectivity) due to the low temporal resolution nature of the BOLD signal as Granger causality is prone to undersampling signals. To avoid that issue in this study, we utilize time-lagged causal discovery framework called Tigramite (time-series graph-based measures of information transfer) [7]. This framework relies on different set of assumptions to identify a causal graph. It performs conditional independence testing using the assumptions of time-order, Causal Sufficiency, the Causal Markov Condition, and Faithfulness, among others [8].

While most conventional fMRI studies rely on pool of data collected across individuals, modeling of the connectivity from data across individuals is controversial [9]. Nonetheless, a method to establish a general model that represents brain connectivity across individuals is essential. We propose a method of constructing a representative graph based on graph structure by using median aggregation of the weight of corresponding edges and nodes of each individual graphs.

2. Method

2.1. Human Connectome Project

Human Connectome Project is a project conducted by the Washington University-University of Minnesota Human Connectome Project Consortium (WU-Minn HCP) [10]. We utilize this dataset because it provides access to large exceptional spatiotemporal resolution of well-characterized samples of healthy individuals.

This study made use of 376 subjects available in HCP data. Of all available 511 subjects, 52 subjects are excluded due to data incompletion, 83 subjects are excluded from the study due to biased responses in picture sequence memory test (subjects always respond to the stimuli in the test only either remember or know). Of the 376 subjects included in this study, 54 are between the ages of 22-25, 179 are between the ages of 26-30, 141 are between the ages of 31-35, and 2 are above 35 year-old.

2.2. Subsequent memory paradigm

The HCP dataset does not include a specific task related to episodic memory, hence, we employ a subsequent memory paradigm by combining HCP working memory task-fMRI and corresponding picture sequence memory test. Subsequent memory paradigm is a study of neural activity during an encoding phase of the stimuli that are later subsequently remembered in contrast to the stimuli that are forgotten [11].

In the encoding phase, a subject is given a chance to study a series of stimulus while their brain activity data are being collected. Later after the encoding phase is completed, the subject then performs a recollection test. In this test, the subject is shown a series of stimulus consists of both old stimuli presented during encoding phase and completely new unseen stimuli. During this phase, the subject is allowed to respond to the stimulus by indicating that they either remember, know, or seeing new stimulus.

The response is collected and evaluated for memory recollection quality. The correction and the confidence of the response is evaluated. If remember or know responses are given to the old stimuli, it
indicates the subject’s correct recollection. The remember response indicates subject’s confidence in their conscious recollection, and the know response indicates subjects doubt in their recollection, in which they may feel familiarity with the stimulus but fails to consciously recall it precisely. Then if the subject gives the new response to the old stimuli, it is considered that the stimuli have subsequently been forgotten by the subject, and if remember or know response is given to the new stimuli, that response is considered as a false recollection. If the new response is given to the new stimulus, it indicates that the subject remembers that the presented stimulus is not included in the set of stimuli previously shown during encoding phase.

2.3. Univariate group subtraction

The scanned images are processed and analyzed using FSL 5.0.7 (FMRIBs Software Library, www.fmrib.ox.ac.uk/fsl) software suite. The HCPs structural MRI and fMRI were preprocessed using FSL 5.0.6 software suite. The preprocessing pipeline is done by HCP using the pipeline derived from [12]. FEAT (FMRI Expert Analysis Tool, v6.00) was used to analyze first- and second-level analysis. The General Linear Model was used to model the event duration. The event-timing were specified in FSLs 3-column Explanatory Value file format. The explanatory file describes time when the interested events occurred, the duration of the events, and the value of the input during that time. The trial on which the stimuli are identified by the subjects as remember and know are selected for the analysis because this study focuses on comparing the condition where the subject encodes the memory of the stimuli that later be successfully recalled with one that failed to be recalled.

Additional Psychophysiological Interaction (PPI) Analysis is done using parahippocampal gyrus as a seed to identify additional memory encoding related region.

2.4. Group causal modeling

We model causal relation between 4 regions of interest (table 1) for each individual subjects using Tigramite causal time series analysis software package. Tigramite is a time-lagged causal discovery frameworks [8]. There are 2 free parameters for the algorithm, the maximum time lag $\tau_{\text{max}}$, and the significance threshold $\alpha$ in the condition- selection step. To determine maximum time lag $\tau_{\text{max}}$, we estimate lagged unconditional dependencies of the BOLD time-series and found the dependencies diminish beyond a lag of 8. The significance threshold $\alpha$ is set to 0.1. The $\alpha$ is a regularization parameter in model- selection techniques, and should not be seen as significance test level in the condition-selection step [13].

Then we construct a group representative causal model by median aggregation of connection weight of corresponding link of each individual connectivity graph. We use median to reduce the effect of outliers. However, interpreting a generalized representative connectivity model from data across individuals is controversial. So, we propose and perform a validation process to establish that the representative connectivity reflects common characteristics of connectivity in each individual subjects.

We use Levenshtein distance [14], a classic graph edit distance (GED), to measure similarity between connectivity graphs. The edit distance compares only structural aspect of the graphs and completely omits link weight. We validate the representative graph by comparing it with a representative graph of a sub-group of randomly selected 100 individual for 1,000 iteration. Then, we average the resulting distance from all iteration.

3. Results

3.1. Univariate group subtraction and regions of interest

The peak activated region yielded by fMRI group subtraction analysis is large (17,529 voxels) and cover wide area of the brain (figure 1). Harvard-Oxford cortical and subcortical structural probability atlases are used to identify the activated region. The major activated regions are selected as regions of interest. Additionally, the PPI shows that lingual gyrus has significant
correlation with parahippocampal gyrus. Therefore, 4 ROIs shown in table 1 are selected for connectivity modeling.

![Activated cluster from univariate group subtraction analysis of all 376 subjects on Successful recollection > Failed recollection (threshold at Z > 2.3, P < 0.05).](image)

**Table 1.** Selected regions of interest (Harvard-Oxford cortical structural probability atlases).

| Regions                      | Hemispheres       |
|------------------------------|-------------------|
| parahippocampal gyrus        | Left and right    |
| temporal occipital fusiform cortex | Left and right |
| occipital gyrus              | Left and right    |
| lingual gyrus                | Left and right    |

3.2. Causal modeling and group representative model

The resulting group representative graphical model and time series graph are presented in figure 2 and figure 3, respectively. In the representative model, only 2 significant connections, a connection from occipital gyrus to temporal occipital fusiform cortex, and from lingual gyrus to parahippocampal gyrus, are presented. Its average Levenshtein distance from our test is 6.35 (SD: 0.28).

The main concern of connectivity measure derived from fMRI data is that movement and physiological noise sources, which vary from subject to subject and session to session, can potentially induce spurious correlations between ROIs, which could confound the interpretation of the results by increasing chance of false positive [15]. The effect of aforementioned spurious correlations may contribute to the variation among individual graphs. The number of connections in the individual graph varies from only 1 to as many as 12 connections. Group model is calculated using median aggregation
to reduce the effect of outlier. The Levenshtein distance is introduced to measure the reliability of the resulting group model by showing that the model is in the average distance among the population.

4. Discussions and conclusions

4.1. Episodic memory encoding
In this study’s subsequent memory paradigm, brains activity is observed only during the encoding phase. Cognitive functions related to memory beyond memory encoding, such as, memory organization or memory recollection are not addressed.

According to previous studies, the resulting peak activated regions are known to have roles in memory formation. A study by Ofen et al. [16] found that middle temporal gyrus, fusiform gyrus, and parahippocampal gyrus play significant role in development of declarative memory formation development from child to adult. Occipital area and fusiform gyrus are known to play important roles in face recognition [17] and declarative memory forming [16]. The protocol used by HCP may contribute to the activation of this region since images of human face are used as stimuli for working memory task fMRI [18]. Lingual gyrus is a part of declarative memory system [19] and also contributes to visual memory forming [20].

The results of this study agree with previous studies and affirm that the prior knowledges hold in large population since most of the previous fMRI studies included only small number of subject due to technical difficulty of fMRI experiment protocol.

4.2. Causal modeling and group representative model
The goal of brain research is to describe biological mechanic of the brain that responsible for any particular cognitive function. Most researches identify brain function in forms of active region and pathway. Sensory pathways is one of the well-known example. However, the knowledge of higher cognitive function’s pathway, such as, consciousness or memory organization is limited since these functions lack physical interaction that can be observed. Causal interaction analysis on fMRI BOLD time-series is a common way to study said brain functions.

Granger causality [5] is a popular approach among numerous techniques developed to analyze causal interactions in fMRI data [21]. However, inferring causal relationship (effective connectivity or directed connection) using this framework on fMRI BOLD data is controversial because low temporal resolution nature of the BOLD signal may confound its statistical dependencies implication. Without the temporal order integrity, the framework shows only correlation (functional connectivity or undirected connection). Consequently, we use Tigramite framework [8] for our study to avoid temporal sensitivity issue of Granger causality framework.
In fMRI study, the localization of specific brain function is observed across individuals using statistical inference, and therefore the location can be generalized across population. However, fMRI statistical framework does not account for temporal order of activation. We have no statistical basis that guarantee temporal order of activation or activation time lag across individual to be able to infer general causal model of a group of individuals. Nonetheless, the general representative model is needed for drawing a general conclusion.

To address this issue, we compare difference between the resulting representative model with model of random sub-groups using Levenshtein distance. The average distant is 6.35 (SD: 0.28), with the maximum distance at 11, and the minimum at 1. While structure of individual subject’s model varies, ranged from a graph with only one connection to a complete graph, our test shows that the representative graph is balanced.

Therefore, we conclude that the graph is a general representative model of this sample population. However, it is important to emphasize that this method only concerns structural similarity, the connection weights are ignored.

References
[1] Tulving E 1993 Curr. Dir. Psychol. Sci. 2 67–70
[2] Ullman M T 2004 Cognition 92 231–270
[3] Lauritzen S L 1996 Graphical Models (Oxford: Oxford University Press)
[4] Eichler M 2012 Probab. Theory Rel. 153 233–268
[5] Granger C W 1969 Econometrica 37 424
[6] Pearl J 2009 Causality: Models, Reasoning and Inference 2nd ed (New York, NY, USA: Cambridge University Press) ISBN 052189560X, 9780521895606
[7] Runge J 2015 Phys. Rev. E 92 062829–18
[8] Runge J 2018 Chaos: An Interdisciplinary Journal of Nonlinear Science 28 075310–21
[9] Duggento A, Passamonti L, Valenza G, Barbieri R, Guerrisi M and Toschi N 2018 Sci. Rep. 8(1) 5571
[10] Van Essen D C et al. 2012 NeuroImage 62 2222–2231
[11] Dolcos F, LaBar K S and Cabeza R 2004 NeuroImage 23 64–74
[12] Glasser M F et al. 2013 NeuroImage 80 105–124
[13] Runge J, Nowack P, Kretschmer M, Flaxman S and Sejdinovic D 2017 arXiv.org (Preprint 1702.07007v2)
[14] Levenshtein V I 1965 Soviet Physics. Doklady 10 707–710
[15] Whitfield-Gabrieli S and Nieto-Castanon A 2012 Brain connectivity 2 125–141
[16] Ofen N, Kao Y C, Sokol-Hessner P, Kim H, Whitfield-Gabrieli S and Gabrieli J D E 2007 Nat. Neurosci. 10 1198–1205
[17] Fusar-Poli P et al. 2009 J. Psychiatr. Neurosci. 34 418–432
[18] Marcus D S et al. 2013 NeuroImage 80 202–219
[19] Menon V, Boyett-Anderson J M, Schatzberg A F and Reiss A L 2002 Cognitive brain res. 13 261–265
[20] Schankin C J, Maniyar F H, Sprenger T, Chou D E, Eller M and Goadsby P J 2014 Headache 54 957–966
[21] Margulies D S et al. 2010 Magma (New York, N.Y.) 23 289–307