Collective Discovery of Brain Network with Unknown Groups

Xinyue Liu*, Xiangnan Kong*, Philip S. Yu#

*Worcester Polytechnic Institute  
#University of Illinois at Chicago
Network Representation of Brain

Each node = a brain region

Each edge = a functional connection
Use of Brain Network

Disease Diagnosis

Not Given!
Should be derived from neuroimaging data

Functional Study
The fMRI Data

The Data Matrix

\[ X \in \mathbb{R}^{m \times n} \]

Each column refers to a voxel of the brain
Each row refers to a snapshot of the brain
Brain Network Discovery Problem

Input

Output

Worcester Polytechnic Institute
Brain Network Discovery

(a) Edge Inference with Known Groups

Extracted from fMRI data

Provided by Neuroscientist
(b) Independent Inference
Collective Brain Network Discovery

(c) Collective Inference (Proposed)
Why Collective?

• [Edge/Link]: The predefined groups may be inferred anatomically and contain sub-regions that are each characterized by different functional connectivity patterns.

• [Node/Group]: In most brain network inference model, once the groups (nodes) are derived, it is difficult or impossible to improve it based on edges/links discovered in latter stages.
Given time series signal of neurons

One can infer the connections between pairs of neurons using Pearson correlation.

Covariance Matrix

Inverse Covariance Matrix

Indirect Connections

Direct Connections
Preliminary: Compute the Inverse

\[ S = \frac{1}{n} X^T X \]

empirical covariance matrix

Find the precision (inverse covariance) matrix

\[ \Theta \neq S^{-1} \]

Very likely to be singular
Preliminary: Graphical Lasso

\[
\text{minimize } \begin{cases} \Theta > 0 \end{cases} \quad -\log \det \Theta + \text{tr}(S\Theta) + \lambda \|\Theta\|_1
\]

- **positive definite constraint**
- **Negative Log Likelihood**
- **L1-norm Regularization**

Find a sparse positive definite matrix \( \Theta \) which has high likelihood to be the precision matrix of \( S \)
SGGL Model: Step 1
(Group Constrained Graphical Lasso)

\[
\arg \min_{\Theta \succ 0} - \log \det \Theta + \text{tr}(S\Theta) + \sum_{i,j} \lambda_{i,j} \| \{ \Theta_{G_i, G_j} \} \|_F
\]

- **Negative Log Likelihood**
- **Group-wise Sparseness Regularizer**

Solved by Spectral Projected Gradient Descent
SGGL Model: Step 2
(Spectral Clustering)

\[ \Omega(\hat{\Theta}) = \text{abs}(\hat{\Theta}) \]

use \[ \Omega(\hat{\Theta}) = \text{abs}(\hat{\Theta}) \]

as the affinity/similarity matrix for spectral clustering to update the group/node \((G_1, \ldots, G_N)\)
SGGL Model: Iterative Inference

Group Constraint
Graphical Lasso

Spectral Clustering

(c) Collective Inference (Proposed)
Synthetic Data

Generator

Ground Truth Precision Matrix

Add Noises

Draw from Multi-variant Gaussian Dist.

Worcester Polytechnic Institute
Synthetic Study (Illustration)

Data set 1
(a) Ground Truth
(b) GLasso
(c) SGGL

Data set 2
(d) Ground Truth
(e) GLasso
(f) SGGL
Synthetic Study (Edge Inference)

Compared Method:

- SGGL (proposed)
- Graphical Lasso
- K-means + GGL
Synthetic Study (Group Inference)

Compared Method:

- SGGL (proposed)
- Spectral Clustering
- K-means
Real fMRI Data

- ADHD-200
  - 20 healthy (TDC) subjects, 20 ADHD patients.
  - 3D brain images of size $61 \times 73 \times 61 \sim 180$ time steps

- All parameters are tuning using DMN.
Fig. 7. The connectivity of DMN of ADHD group discovered by SGGL. All regions in the DMN are strongly connected to each other, which is consistent with the essence of DMN.
Real fMRI data (DMN Recovery)

Fig. 6. Comparison of NMI scores on the DMN (Default Mode Network) of ADHD-200 Data.
Real fMRI data
(Entire Brain Group Inference)

(a) Groups of TDC inferred by SGGL ($k = 20$)

(b) Groups of ADHD inferred by SGGL ($k = 20$)

(c) Groups of TDC inferred by spectral clustering ($k = 20$)

(d) Groups of ADHD inferred by spectral clustering ($k = 20$)

more Interpretable results

Scattered, hard to see difference
Effect of Screening

The chart shows the time cost (s) for TDC and ADHD with and without screening. The chart indicates that about 40% time reduction is achieved.
Summary

• Problem Studied
  • Collective brain network discovery (nodes + edges)
• Proposed solution
  • Use Group Constrained Graphical Lasso to infer the edges
  • Use spectral clustering with the masked precision matrix as the affinity matrix to update the group
  • Use the updated group to further update the edges and to repeat previous two steps until converge
  • A set of screening rules is proposed to speed up the edge inference
• Conclusion
  • Experiments on synthetic data and real-world fMRI demonstrated the superiority of the proposed SGGL method