Penalized Model-Based Clustering of FMRI Data

Andrew DiLernia

Division of Biostatistics
University of Minnesota

June 8, 2020

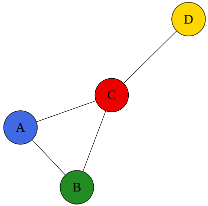
Background

- FMRI can describe functional connectivity (FC), temporal dependence of neuronal activity in regions of the brain[5].
- Alterations in FC associated with psychiatric disorders e.g. major depressive disorder[7], schizophrenia[3], neurodegenerative diseases such as Alzheimer's disease[2]

Background

- Graphical models useful for exploring relationships between regions of interest (ROI) to describe FC
- Under the assumption of the fMRI signals being Gaussian, ROI constitute vertices, and non-zero entries in the precision matrix, $\boldsymbol{\Omega}$, imply edges between vertices in a graph.

$$\boldsymbol{\Omega} = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \!\!\begin{array}{c} \begin{array}{cccc} A \quad & B \quad & C \quad & D \end{array} \\ \begin{pmatrix} 0.61 & 0.14 & 0.05 & 0 \\ 0.14 & 0.61 & 0.15 & 0 \\ 0.05 & 0.15 & 0.61 & 0.15 \\ 0 & 0 & 0.15 & 0.61 \end{pmatrix} \end{array} \longrightarrow$$



Motivation

- Want interpretable estimates of subject- and group-level FC for multi-subject data sets
- Also desire to group participants based on FC when classes are unknown
  - E.g. participants with varying degrees of schizophrenia or various stages of the progression of Alzheimer's
- Motivates penalized model-based clustering of fMRI

Motivation

- 2 main goals of proposed random covariance clustering model (RCCM):

  1. Clustering of subjects based on FC
  2. Sparse precision matrix estimation to describe subject- and group-level FC for multiple subjects

- Achieve simultaneously using penalized model-based clustering

Notation

Suppose we have fMRI data where:

- K: number of subjects
- p: number of ROI
- n_k: number of observations or time points for k^{th} subject
- G: number of clusters or groups

Model

- $\mathbf{y_{kt}} = (y_{k1t}, ..., y_{kpt})^T \sim \mathcal{N}_p(\boldsymbol{\mu}_\mathbf{k}, \boldsymbol{\Sigma}_\mathbf{k})$ are independent p-dimensional Gaussian random variables
- y_{kjt}: t^{th} observation or time point of the j^{th} ROI for the k^{th} subject
  - $k = 1, \ldots, K$ indexes subjects
  - $j = 1, \ldots, p$ indexes ROI
  - $t = 1, \ldots, n_k$ indexes observations or time points

Model

- $\mathbf{\Omega_k} = \mathbf{\Sigma_k}^{-1} \sim p(\mathbf{\Omega_k}; \{\mathbf{\Omega_{0g}}, \pi_g\}_{g=1}^{G}) = \sum_{g=1}^{G} \pi_g p_g(\mathbf{\Omega_k}; \lambda_2, \mathbf{\Omega_{0g}})$
  - Mixture distribution with G components
  - $p_g(\mathbf{\Omega_k}; \lambda_2, \mathbf{\Omega_{0g}})$: PDF of Wishart random matrix with degrees of freedom λ_2 and mean $\mathbf{\Omega_{0g}}$
  - $\mathbf{\Omega_{0g}}$: cluster-level precision matrix of cluster g
  - π_g: interpreted as proportion of subjects belonging to cluster g

Model

- Novelty lies in mixture Wishart distribution
  - Facilitates interpretation of each subject's FC being similar to their cluster-level FC, but not necessarily identical
  - Different from a Gaussian mixture model which clusters individual observations

Model

- Assuming centered data, model likelihood is:

$$L = \prod_{k=1}^{K} \prod_{t=1}^{n_k} \left(f_k(\mathbf{y_{kt}}; \mathbf{\Omega_k})\right) p(\mathbf{\Omega_k}; \{\pi_g, \mathbf{\Omega_{0g}}\}_{g=1}^{G})$$

- Induce sparsity in the precision matrices using ℓ_1 penalties:

$$-\log(L) + P_\lambda = -\log(L) + \lambda_1 \sum_{k=1}^{K} ||\mathbf{\Omega_k}||_1 + \lambda_3 \sum_{g=1}^{G} ||\mathbf{\Omega_{0g}}||_1$$

EM Algorithm

- Difficult function to minimize directly
  - Do not know which subjects belong to which group
- EM Algorithm with block-coordinate descent to optimize with respect to $\Theta = \{(\pi_g, \mathbf{\Omega_k}, \mathbf{\Omega_{0g}})\}$

M-Step

Block-coordinate descent:

1. Initialize $\mathbf{\Omega_k}^{(0)} = \widehat{\mathbf{\Omega}}_{\mathbf{k_{gl}}}$ for $k = 1, \ldots, K$ where $\widehat{\mathbf{\Omega}}_{\mathbf{k_{gl}}}$ is the individual GLasso estimate.

2. Initialize the cluster memberships for each subject

M-Step

Steps: (cont.)

3. Update $\{\boldsymbol{\Omega}_{0g}\}_{g=1}^{G}$ using coordinate descent approach for covariance graphical lasso[6].
4. Update the weights, and then $\{\boldsymbol{\Omega}_k\}_{k=1}^{K}$ using the GLasso algorithm[4].
5. Repeat steps 3 and 4 until convergence.

Simulations

Compared RCCM to a 2-step approach: Ward clustering & the group graphical lasso (GGL)[1]

- Ward clustering was done based on a distance matrix constructed from the sample precision matrices
- GGL conducts joint estimation of multiple precision matrices which encourages a shared sparsity structure

Edge Detection

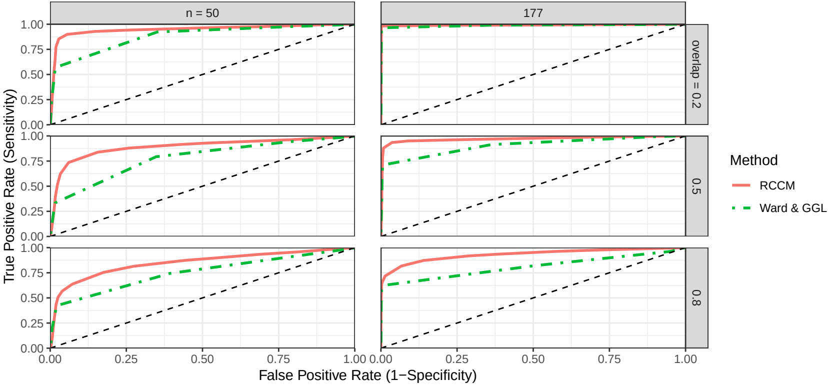


Figure 1: Group-level ROC curves for edge detection for 104 subjects belonging to one of $G = 2$ groups.

Data Analysis

- Also applied RCCM to a resting-state fMRI data set
- Data: 61 total participants with first-episode or chronic schizophrenia and 43 healthy controls

Data Analysis

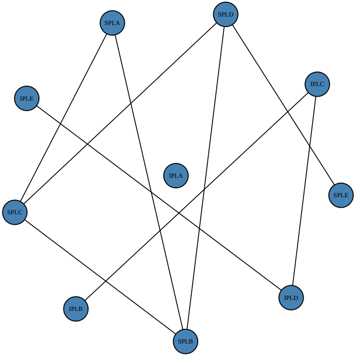


Figure 2: Group A (left) contained more participants with schizophrenia than B (right), and its estimated network had fewer connections than B, providing evidence for decreased FC among those with schizophrenia.

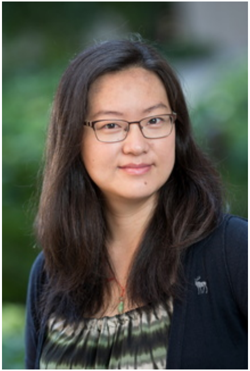
Future Directions

Potential extensions:

1. Account for autocorrelation present in fMRI data
2. Allow for time-varying connectivity
3. Extend for a supervised or semi-supervised approach

Acknowledgements

Many thanks to my brilliant and incredibly supportive advisor Dr. Lin Zhang.



Thank you

- Our proposed RCCM is available via an R package at github.com/dilernia/rcm.
- Thank you

References

[1] Danaher, P., Wang, P., and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society*, 76(2):373–397.

[2] Dennis, E. and Thompson, P. (2014). Functional brain connectivity using fmri in aging and alzheimer's disease. *Neuropsychology Review*, 24(1):49–62.

[3] Fornito, A., Zalesky, A., Pantelis, C., and Bullmore, E. T. (2012). Schizophrenia, neuroimaging and connectomics. *Neuroimage*, 62(4):2296–2314.

[4] Friedman, J., Hastie, T., and Tibshirani, R. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

[5] Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J. (1993). Functional connectivity: The principal-component analysis of large (pet) data sets. *Journal of Cerebral Blood Flow & Metabolism*, 13(1):5–14.

[6] Wang, H. (2014). Coordinate descent algorithm for covariance graphical lasso. *Statistics and Computing*, 24(4):521–529.

[7] Zeng, L., Shen, H., Liu, L., and Hu, D. (2014). Unsupervised classification of major depression using functional connectivity mri. *Human Brain Mapping*, 35(4):1630–1641.

Clustering Results

Table 1: Summary of clustering results for specifying $G = 2$ or 3 clusters. For $G = 2$, Group A had the highest proportion of participants diagnosed with schizophrenia, while healthy controls were more evenly distributed between Groups A and B. Results for $G = 3$ groups were not as clear, with most subjects being clustered into Group C.

G	Group	Control	1st Episode	Chronic
2	A	23 (53.5%)	14 (66.7%)	30 (75.0%)
	B	20 (46.5%)	7 (33.3%)	10 (25.0%)

University of Minnesota