Brain Functional Connectivity Analysis Using Mutual Information

Zhe Wang, Ahmed Alahmadi, David C. Zhu, Tongtong Li

Department of Electrical & Computer Engineering

Michigan State University

East Lansing, Michigan 48824, USA.

IEEE GlobalSIP 2015, Orlando, Florida, USA.

Outline

- Introduction: Pearson vs. mutual information (MI)
- Existing work on MI estimation
- Proposed approach for MI Estimation
- Numerical Results
- Conclusions

Introduction

- Connectivity analysis based on functional Magnetic Resonance Imaging (fMRI) data helps to reveal insights on brain functioning and disease analysis.

- For decades, an important metric in measuring functional connectivities has been the **Pearson correlation coefficient** defined as:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

- In recent years, **mutual information** (MI) has been applied as an alternative metric since it measures not only linear dependence between two time series but also <u>non-linear</u> relationships.
- For brain functional connectivity this means that MI can confirm *known functional connections* as well as discovering *new connections* [1].
- In [2], the authors used MI to build the brain activation map. They showed that MI was robust in quantifying the relationship between any two fMRI temporal response waveforms.
- In [3], MI has been applied in the activation detection, and showed that MI is a more sensitive metric than the Jaccard overlap metric.
- In [4], the authors applied MI in the decoding algorithm in selecting features from high dimensional data, and showed that MI was efficient in selecting very few but strongly informative voxels.

Existing Approach on MI Estimation

- The mutual information between two random variables X and Y is defined as:

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y) log \frac{f_{XY}(x,y)}{f_X(x) f_Y(y)} dxdy$$

- In words, MI is the information successfully transmitted through a channel.

- In literature, a dominant approach for MI calculation has been the k nearest neighbor (kNN) estimator.

- Given n samples with dimension d, the k nearest neighbor (kNN) estimator calculates the probability density function (pdf) as:

$$f(\overrightarrow{x}_i) = \frac{1}{2}\frac{\Gamma(d/2+2)}{\pi^{d/2}}\frac{1}{r(\overrightarrow{x}_i)^d}$$

where $r(\overrightarrow{x}_i)$ is the Euclidean distance from $\overrightarrow{x_i}$ to its kth nearest neighboring points, and $\Gamma(*)$ the Gamma function.

- There are several concerns related to the kNN estimator:

1. The choice of k is not a well-defined problem and usually solved by heuristic techniques [5].

2. The choice of k is highly empirical. It involves trade-off between the estimation bias and variance.

3. The algorithm cannot be guaranteed to converge to the true value when the data samples are not independently distributed.

The Proposed Approach for MI Estimation

- To address the limitations above, we propose a novel approach to estimate MI and measure brain functional connectivity.

- The major steps in the proposed method are:

1. Decorrelate the data segments.

2. Apply Kernel-based estimation to estimate the probability density function.

3. Use Monte Carlo Integration for MI estimation.

Decorrelation of Data Segments

- The data segments from the fMRI data are generally correlated with each other, which introduces skewness in the distribution [6].
- We propose to adopt the whitening transform on the data segments before estimating the probability density function.
- After decorrelation, the covariance matrix of the transformed data is close to the identity matrix.

- Given a d-dimensional data vector $\overrightarrow{x}$, with mean $\overrightarrow{\mu}$ and covariance matrix $\Sigma = E(\overrightarrow{x} - \overrightarrow{\mu})(\overrightarrow{x} - \overrightarrow{\mu})^T$.
- The covariance matrix can be decomposed as $\Sigma = \phi\Lambda\phi^{-1}$, where Λ is a diagonal matrix, with the eigenvalues as its diagonal elements, and ϕ the eigenvectors of the covariance matrix.
- The whitening transform is defined as:

$$\overrightarrow{x}' = A^T\overrightarrow{x}, A = \phi\Lambda^{-1/2}$$

- After the transform, the new data vector will have mean $A^T\overrightarrow{\mu}$ and covariance I. That is:

$$E(A^T\overrightarrow{x} - A^T\overrightarrow{\mu})(A^T\overrightarrow{x} - A^T\overrightarrow{\mu})^T = I$$

Kernel-based Estimation of Probability Density Function

- The basic idea is to calculate the average of kernel functions K on each point that falls into a pre-specified kernel window.

- Given a set of data points $\{\epsilon_i | i = 1, \cdots m\}$, the kernel estimation for the probability density function at any point x is given by:

$$\hat{f}(x) = \frac{1}{m} \sum_{i=1}^{m} K\left(\frac{x - \epsilon_i}{h}\right)$$

  in which h is the bandwidth for the kernel function that needs to be optimized.

- For any d-dimensional data vector $\overrightarrow{x} = [x_1, ..., x_j, ..., x_d]$, the estimation function can be extended as:

$$\hat{f}(\overrightarrow{x}) = \frac{1}{mh_1h_2...h_d} \sum_{i=1}^{m} K_h(\overrightarrow{x}, \overrightarrow{\epsilon}_i)$$

  where

$$K_h(\overrightarrow{x}, \overrightarrow{\epsilon}_i) = \prod_{j=1}^{d} K\left(\frac{x_j - \epsilon_{ij}}{h_j}\right)$$

- The kernel function K for a continuous variable x is often chosen as the Gaussian function [7]:

$$K(h, \epsilon, x) = \frac{1}{2\sqrt{\pi}} e^{-\frac{(x-\epsilon)^2}{4h^2}}$$

Optimal Kernel Bandwidth Estimation

- The bandwidth h for the kernel function has a significant influence on the estimation accuracy.
- h is chosen to maximize the *leave-one-out* log likelihood function given by:

$$L = \sum_{i=1}^{m} log \hat{f}_{-i}(\overrightarrow{\epsilon}_i)$$

where

$$\hat{f}_{-i}(\overrightarrow{\epsilon_i}) = \frac{1}{(m-1)h_1 h_2 ... h_d} \sum_{j \neq i} K_h(\overrightarrow{\epsilon}_i, \overrightarrow{\epsilon}_j)$$

- To optimize the likelihood function L, we applied downhill simplex method in [8].

MI Estimation Through Monte Carlo Integration

- The Monte Carlo integration method is used here to calculate the MI after the probability distribution function has been obtained.

- Let $i(\overrightarrow{x}, \overrightarrow{y}) = f_{XY}(\overrightarrow{x}, \overrightarrow{y}) log \frac{f_{XY}(\overrightarrow{x}, \overrightarrow{y})}{f_X(\overrightarrow{x}) f_Y(\overrightarrow{y})}$. Then, to calculate the MI $I = \int\limits_{R^d} \int\limits_{R^d} i(\overrightarrow{x}, \overrightarrow{y}) d\overrightarrow{x} d\overrightarrow{y}$, the algorithm uniformly samples a finite space with a volume of V, and generates m samples $\{(\overrightarrow{x}_i, \overrightarrow{y}_i), i = 1, 2, \cdots, m\}$.

- The mutual information, then, can be estimated as:

$$I_m \approx \frac{V}{m} \sum_{i=1}^{m} i(\overrightarrow{x}_i, \overrightarrow{y}_i)$$

- Since the definition domain of a Gaussian function is infinite, sampling on the whole space is impossible.

- We limit the sampling space within the interval $[\mu - 3\sigma, \mu + 3\sigma]$, where μ and σ denote the mean and standard deviation of the Gaussian function, respectively.

- *Convergence*: It can be shown by the Law of Large Numbers (LLN): as m goes to infinity, the approximation above will converge to the real value of I, i.e., $\lim_{m \to \infty} I_m = I$.

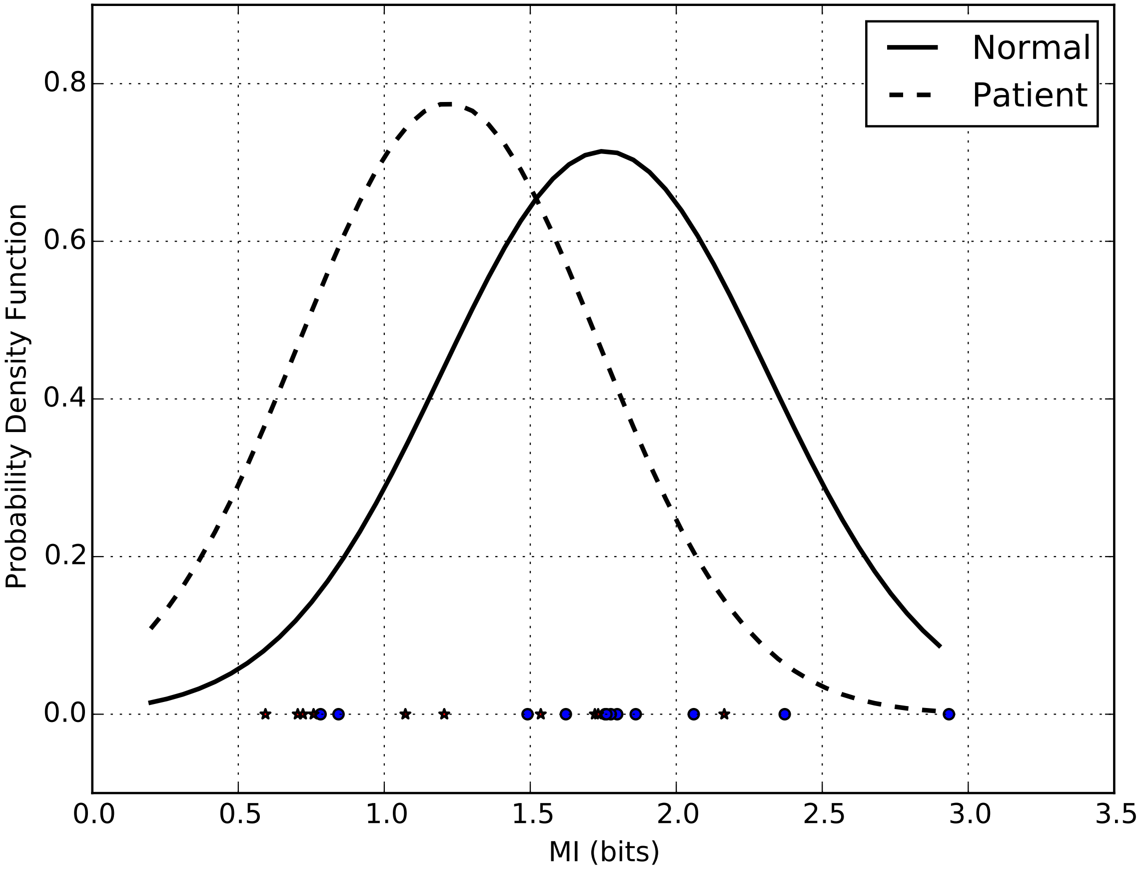
Numerical Results

- Here, we apply the proposed approaches to resting state fMRI data collected from both Alzheimer's Disease patients and normal control subjects.

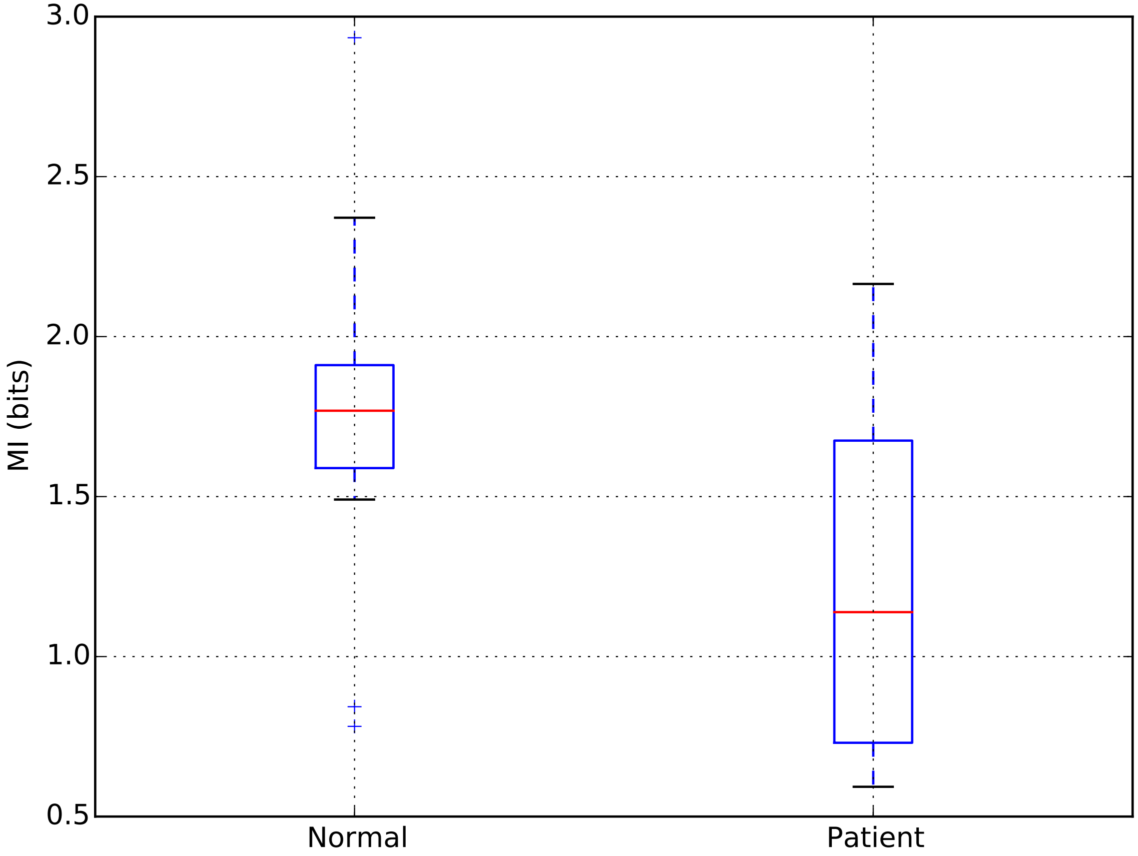
- Brain networks operate in a cohesive manner of connections between nodes. A progressive weakening trend of functional connectivities has been observed in the default mode network (DMN) in AD patients [9]. In the following, we will evaluate MI between two regions of DMN, the posterior cingulate cortex (PCC) and superior frontal gyrus (SFG).

- In the data collection process, eleven patients with mild-to-moderate probable AD and twelve healthy normal control subjects were recruited to participate in this study.

- Figure 1 shows the probability distribution function of calculated mutual information for connections between PCC and SFG.
- As expected, the connections experienced a decrease in AD patients compared to NC subjects.



- Figure 2 shows the boxplots for the two groups of calculated mutual information for connections between PCC and SFG.
- It can be seen that the median suffered a 35.6%'s decrease in AD patients compared to NC subjects.



Conclusion

- In this paper, we considered the measurement of functional brain connectivities using mutual information.
- We proposed a novel approach for the estimation of MI, which was composed of three major components: decorrelation, kernel based estimation of probability density function and Monte Carlo Integration for MI estimation.
- The analysis results obtained using the proposed method were consistent with clinical observations in the AD data sets.

References

[1] B. Chai, D. B. Walther, D. M. Beck, and L. Fei-Fei, "Exploring functional connectivities of the human brain using multivariate information analysis," in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds. Curran Associates, Inc., 2009, pp. 270–278.

[2] A. Tsai, I. Fisher, JohnW., C. Wible, I. Wells, WilliamM., J. Kim, and A. Willsky, "Analysis of functional MRI data using mutual information," in *Medical Image Computing and Computer-Assisted Intervention MICCAI99*, ser. Lecture Notes in Computer Science, C. Taylor and A. Colchester, Eds. Springer Berlin Heidelberg, 1999, vol. 1679, pp. 473–480.

[3] B. Afshin-Pour, H. Soltanian-Zadeh, G.-A. Hossein-Zadeh, C. L. Grady, and S. C. Strother, "A mutual information-based metric for evaluation of fMRI data-processing approaches," *Humman Brain Mapping*, vol. 32, pp. 699–715, 2011.

[4] V. Michel, C. Damon, and B. Thirion, "Mutual information-based feature selection enhances fMRI brain activity classification," in *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008.*, May 2008, pp. 592–595.

[5] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Miscellaneous Clustering Methods, in Cluster Analysis*, 5th ed. Chichester, UK: Wiley, 2011.

[6] Q. Wang, S. Kulkarni, and S. Verdu, "Divergence estimation for multidimensional densities via k-nearest-neighbor distances," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2392–2405, May 2009.

[7] J. S. Racine, "Nonparametric econometrics: A primer," *Foundations and Trends in Econometrics*, vol. 3, no. 1, pp. 1–88, 2008.

[8] T. Weise, *Global Optimization Algorithms – Theory and Application*. Germany: it-weise.de (self-published), 2009. [Online]. Available: http://www.it-weise.de/projects/book.pdf

[9] D. C. Zhu, S. Majumdar, I. O. Korolev, K. L. Berger, and A. C. Bozoki, "Alzheimer's disease and amnestic mild cognitive impairment weaken connections within the default-mode network: a multi-modal imaging study," *Journal of Alzheimer's Disease*, vol. 34, no. 4, pp. 969–984, 2013.