EN.580.694: Statistical Connectomics
Final Project Proposal

Kristin Gunnarsdottir – March 31, 2015

Gender classification using networks based analysis of the human connectome

Opportunity

It is well known that men and women have different characteristics and studies have for example shown that males have better motor and spatial abilities, whereas females have high memory and social cognition skills. Literature suggests that while there are many similarities in brain structure, function and neurotransmission in men and women, there are important differences that discriminate male brains from female brains. Understanding the sex differences in the human brain can provide important information and while differences in total brain size are less meaningful, size differences in smaller brain structures are important for both normal behaviors and diseases and may even reflect the pathology or frequency of disease in men and women [1]. Using the human connectome to analyze the brain as a network provides a great opportunity to gain additional knowledge about the organization and interactions between brain structures and demonstrate if gender difference shows altered connectivity patterns between brain regions.

Challenge

Studies show sex differences in human brains (such as size and shape of specific brain structures), but currently the reasons for the connectivity difference are not as well understood. Additionally, because of a high variability between individuals in brain anatomy and brain activity, a certain degree of difference can be expected between the data of different subjects. The sample size is also rather small, with only 21 subjects (two scans per subject so a total of 42 samples), which could potentially make it harder to build a reliable classifier.

Action

In this project I will attempt to discriminate between male and female brains based on their brain region connectivity patterns. The dataset used are the adjacency matrices (small graphs) of the KKI-42 dataset, including scans from 21 individuals [2]. I intend to make a Likelihood Ratio Classifier based on node-wise measures computed from the graphs. Specifically, I will compute the eigenvector centrality, the cosine similarity and the degree of each node, as well as the weighted degree and/or the local clustering coefficient of each node. For each measure I will only observe the nodes that show the most significant difference between the two groups and use them as an input to the classifier.

Resolution

Hopefully I will be able to observe which of the features mentioned above are significantly different between the same regions in the two groups. Furthermore, I hope to be able to detect the features and regions that give the most valuable information about the gender difference and thereby gain further insight into the difference of connectivity patterns between males and females.

Future Work

In this project I only chose a few node-wise measures to compute from the network in order to observe if there was a significant difference between men and women. Other node-, edge- or network-wise measures can possibly yield better classification accuracy. Moreover, the classifier will only use one feature at time, but potentially a better accuracy can be obtained using more than one of those measures simultaneously as an input to the classifier. Finally, the dataset used was very small so using a larger dataset will provide more significant and conclusive results.

Statistical Decision Theoretic

Sample Space

$\mathcal{G}_n = (\mathcal{V}, \mathcal{E}, \mathcal{Y}, \mathcal{C})$

Where

$\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ is the set of vertices

$\mathcal{E} = \{e_{11}, \dots, e_{22}\}$ is the set of potential weighted edges

$\mathcal{Y} = \{0,1\}^m$ is the class label for each attribute $(0 = female, 1 = male)$.
$m\ is\ the\ number\ of\ samples\ (m = 42)$

$\mathcal{C} = \{c_1, c_2, \dots, c_k\}$ are the attributes computed from the graphs. Here $k = 5$

Model

$\mathcal{G}_n = (\mathcal{V}, \mathcal{E}, \mathcal{C})$

Action Space

The set of class labels

$\mathcal{A} = \{\mathcal{Y} = \{0,1\}^m\}$

Decision Rule Class

$f{:}\,\mathcal{C} \to \mathcal{Y}$

$$f = \begin{cases} 0 & \dfrac{P(\mathcal{G}_i|y=0)P(y=0)}{P(\mathcal{G})} \geq \dfrac{P(\mathcal{G}_i|y=1)P(y=1)}{P(\mathcal{G})} \\ 1 & \dfrac{P(\mathcal{G}_i|y=0)P(y=0)}{P(\mathcal{G})} < \dfrac{P(\mathcal{G}_i|y=1)P(y=1)}{P(\mathcal{G})} \end{cases}$$

Loss Function

The loss is determined by the number of times a graph was labelled correctly, i.e.

$$\ell: \mathcal{G}_n \times \mathcal{A} \to \mathbb{R}_+$$

$$\ell = \sum_{i=1}^{m} \mathbb{I}\{\hat{y}_i = y_i\}$$

Risk Function.

The risk function is the expected loss

$$R = E[\ell]$$

References

[1] Cosgrove, K., Mazure, C., & Staley, J. (2007). Evolving Knowledge Of Sex Differences In Brain Structure, Function, And Chemistry. Biological Psychiatry, 62:847-855.

[2] Open Connectome Project. Retrieved April 2, 2015, from http://openconnecto.me/data/public/MR/MIGRAINE_v1_0/