EN.580.694: Statistical Connectomics
Final Project Proposal

Michelle Chyn ·April 2, 2015

Kernel based approach for Connectome Wide Association Studies

Opportunity Currently there are many separate studies on estimating connectomes of specific disease states or phenotypes such as ADHD, developmental age, IQ, etc.. The next step would be to find voxel connections that are indicators of specific phenotypes and to see what voxelwise connections differ between phenotypes for connectome wide association studies.

Challenge The method used most commonly involved univariate analysis, which was computationally expensive. The novel concept presented by Shezhad et al. in 2014 was to use multivariate distance matrix regression [**?**], but this method employed the use of a kernel like matrix without taking advantage of the toolbox offered by using kernels.

Action We propose to create a kernel based distance matrix similar to in MDMR and employ kernel based methods for hypothesis testing. This method will take a multivariate data set and compute statistics through dot products in higher dimensional spaces. These results will be compared against what is expected from meta-analysis [**?**], and against another data sets to ensure different clusters are being called significant.

Resolution This method will improve on current methods of comparing voxel connections associated with connectomes of specific phenotypes and also take advantage of kernel based methods, which do not need information such as the distribution of the data.

Future Work The original dataset for the MDMR study came with time series data. Possible improvements on the results of this project would be to incorporate time dependence before apply kernel based testing, to apply kernel based testing to connectome wide association studies of more phenotypes, or for structural connectomes.

Statistical Decision Theoretic

We will take the kernel distance between voxels and form $n, v \times v$ matrices. Then we will compare the distance between voxels between patients. Finally, we will use permutation testing to create a null distribution which we can compare against.

Sample Space $\Omega = T \times N \times V \times P$, where T is the number of time steps, N is the number of participants, V is the number of voxels, and P is the number of phenotypes.

Model $V \times D^{n \times n}$ Distance Matries where $d_{i,j}$ represents kernel distance.

Action Space $\mathcal{H}_0 : P_1 = P_2$, $\mathcal{H}_A : P_1 \neq P_2$ [?].

Decision Rule Class Create a null distribution using permuation testing.
If $\mathcal{T}_n \in R(\alpha)$, decide $\mathcal{H}_0$, if $\mathcal{T}_n \notin R(\alpha)$, decide $\mathcal{H}_A$, where

$$\mathcal{T}_n = \frac{n_1 n_2}{n_1 + n_2} (\hat{D}_2 - \hat{D}_1)^T \hat{\Sigma}_W^{-1} (\hat{D}_2 - \hat{D}_1)$$

and $\hat{\Sigma}_W^{-1}$ is the within-sample pooled covariance matrix

$$\hat{\Sigma}_W^{-1} = \frac{n_1}{n_1 + n_2} \hat{\Sigma}_1 + \frac{n_2}{n_1 + n_2} \hat{\Sigma}_2 \ [? \ ].$$

Loss Function
$\mathcal{P}_{FA} = \mathbb{P}(\text{decide } \mathcal{H}_A | \mathcal{H}_0 \text{ is true }) = \alpha = \text{Type I error,}$
$\mathcal{P}_{MD} = \mathbb{P}(\text{decide } \mathcal{H}_0 | \mathcal{H}_A \text{ is true }) = \pi = \text{Type II error.}$

Risk Function $R = E[L]$

References

[1] Harchaoui, Z., Bach, F., Cappe, O., & Moulines, E. (2013). Kernel-Based Methods for Hypothesis Testing: A Unified View. *IEEE Signal Processing Magazine*, 87-97.

[2] Neurosynth. (n.d.). Retrieved April 1, 2015, from http://neurosynth.org/

[3] Shehzad, Z., Kelly, C., Reiss, P., Craddock, R., Emerson, J., Mcmahon, K., . . . Milham, M. (2014). A multivariate distance-based analytic framework for connectome-wide association studies. *NeuroImage*, 74-94.