EN.580.694: Statistical Connectomics
Final Project Proposal

Heather G. Patsolic · April 2, 2015

Applying kernel-based hypothesis testing to a connectome-wide association study

Opportunity Many diseases seem to be correlated with functional and structural patterns in the brain. Therefore, it is of interest to study how these patterns in people with a particular disease differ from the connectivity patterns seen in people without the disease.

Challenge As was studied in Shehzad et al, I could use Multivariate Distance Matrix Regression (MDMR) (see [2]); this requires a distance matrix, similar to using a kernel, but without any of the advantages to kernel-based hypothesis testing. Since reproducing kernel Hilbert spaces have nice properties in hypothesis testing (as described in Harchaoui), it should be beneficial to use these methods instead of MDMR or univariate methods for finding phenotype/voxel-region associations.

Action Define a kernel map and use the kernel to determine closeness of connectivity patterns for voxels across individuals. *TODO: There is info from time-series about voxel connections. How should we use this?* Then, use this data to create a null distribution and use kernel-based methods for hypothesis testing, as described in Kernel-based methods for hypothesis testing by Harchaoui et al (2013), to determine association between voxel connectivity patterns and particular phenotypes [1]. *TODO: Get Data! (Wanting to test Parkinson's, if I can get the data for it.)* Hypothesis Testing: We say the null distribution represents connectivity patterns having no correlation with the particular phenotype being studied.

Resolution Using kernels allows us to do computations for analysis more easily, by allowing us the use of dot-products as a measure of distance. We also have the advantage of being able to use kernel-based methods for hypothesis testing as described by Harchaoui et al (2013) [1]. Hopefully, we will also be able to show that kernel-based methods work comparatively to what has been done before, if not better. *TODO: How can we take full advantage of using RKHS?*

Future In the future, we could test these methods with other data sets to determine associations between functional connectivity and other diseases or phenotypes. We could also determine associations between phenotypes and structural connectivity. I've been told that often we can apply similar methods from neural network data to social networks, so this could also be a future endeavor. We can also test whether kernels work better for connectome-wide studies or for more localized region of interest studies.

Statistical Decision Theoretic

There are n subjects. For each subject we associate a classifier of 1 or 0 as to whether the subject's Parkinson's status is positive or negative. We also keep the control group (non-Parkinson's) as similar to diagnostic group as possible by controlling for age and sex of the patients. With each patient there is also a time-series of vectors corresponding to blood oxygen level dependent values for a set of voxels that the patients all have in common. Suppose we look at m voxels at t different times per voxel (each voxel has same scan-time data as the others).

Sample Space $\{0,1\}^n \times \mathbb{R}^{n \times m \times t}$

Model Kernel or matrix of similarities, per voxel. *TODO: Figure this out in more detail.*

Action Space $\mathcal{A} = \{\text{reject}, \text{fail to reject}\}$

Decision Rule We reject the null when our p-value is less than .05 and fail to reject otherwise. Check the actual data against a null distribution.

H_0 is essentially that phenotype and voxel connectivity patterns are not correlated, so the phenotypic variable is not helpful in determining voxel patterns (and vice versa). The alternative, H_A, is that they are correlated (dependent) variables.

As in Harchaoui (2013), "the decision rule would be

$$T_n \in R(\alpha) \Rightarrow H_0$$
$$T_n \notin R(\alpha) \Rightarrow H_A$$

where T_n is a test statistic, $\alpha = \text{P(Type I error)}$ is the probability that we reject the null when it is actually true (false alarm), and $R(\alpha)$ is the critical region" [1].
TODO: check citation

Loss Function Potentially, the loss will be a function which yields a 1 if our evaluation is correct (that is, we correctly reject the null when null is false or fail to reject the null when it's true) and yields a 0 otherwise.

Risk Function Power of the decision rule (a.k.a. the probability we reject the null when the null is false). (Power = 1-P(type II error))

TODO: Check bibliography

References

[1] Z. Harchaoui, F. Bach, O. Cappé and É Moulines. Kernel-Based Methods for Hypothesis Testing. *IEEE Signal Processing Magazine* (2013 July), 87–97.

[2] Z. Shehzad, C. Kelly, P. T. Reiss, R. C. Craddock, J. W. Emerson, K. McMahon, D. A. Copland, F. X. Castellanos, and M. P. Milham. An Multivariate Distance-Based Analytic Framework for Connectome-Wide Association Studies. *Neuroimage* **93 Pt 1** (2014 June), 74-94.