Discovering signals in fMRI data; a multiple-testing, Bayesian nonparametric approach

Project Proposal for STAT308

Ahmed Bou-Rabee, Wanrong Zhu, Zheng Xu, Mo Zhou

February 15, 2017

Introduction

The goal of this project is to formulate and test a method which can be used to adaptively identify clusters of signals in functional magnetic resonance imaging (fMRI) data. Roughly, fMRI measures the change in brain blood flow associated with mental activity [HSM04]. The brain is divided into regions known as voxels, and the intensity of the blood flow over each voxel is recorded at evenly spaced time intervals while the subject is stimulated. The data is then in the form (voxel, time, intensity of reading). For example, suppose researchers wanted to identify regions of the brain associated with hunger or craving. To aid in this, fMRI readings can be taken while subjects are shown pictures of pizza and hamburgers.

An advantage of using fMRI is that it's a noninvasive procedure. Due to this, there are many publicly available datasets [PBM+13]. However, analyzing fMRI data poses many statistical challenges, one of them being the multiple comparisons problem. Any sampling procedure involves error. Because there are typically thousands of voxels, it's likely that individual voxels may have high readings, but not be statistically significant. Also, the noise of a reading is typically dependent on location, i.e., it is heteroscedasticic. Identifying clusters (not just individual voxels) introduces an additional challenge.

Project Goals

For the purposes of our project, we will assume, like [FBR15], that the data given to us is of the form (voxel, time, p-value). We will instead focus on two questions:

- Can we adaptively identify entire regions of the brain (not just voxels) which are associated with the stimulus while accounting for multiple testing error?
- There is no response variable in fMRI data, so how can we reliably test our algorithm?

Adaptively identifying regions of interest

Some existing work, e.g., [FBR15], has developed methods which identify significant voxels and predefined regions of interest. These methods presuppose that regions of interest have

already been identified by, say, biologists. We would like to focus on identifying these regions adaptively, rather than using predefined ones. One idea we have for doing so involves a non-parametric Bayesian procedure. Let's assume that the data is generated according to the following Bayesian process.

1. For each location and time (i,j), generate a Beta distribution $Beta(\alpha_{ij}, \beta_{ij})$, where (α_i, β_i) is drawn from a prior distribution on distributions Φ which incorporates time and spatial information.

2. For each location i, j generate the corresponding p-value by making a random draw from $Beta(\alpha_{ij}, \beta_{ij})$.

Then, given our set of data , we can identify, computationally, the maximum posterior likelihood and automatically answer this question without any multiple testing problems, while accounting for noise which is correlated with location.

XX

Once we have identified prior distributions, we identify clusters of high activity by picking clusters which have a distribution that is sufficiently far away from uniform. Kullback-Leibler divergence from uniform

Testing the algorithm

We will reimplement p-filter paper and test against it. We will also find 2 other datasets and test p-filter and our algorithm against it.

We will also do a resampling procedure and test the variance of our nonparametric algorithm.

References

[FBR15] Rina Foygel Barber and Aaditya Ramdas. The p-filter: multi-layer fdr control for grouped hypotheses. *arXiv preprint arXiv:1512.03397*, 2015.

[HSM04] Scott A Huettel, Allen W Song, and Gregory McCarthy. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, 2004.

[PBM+13] Russell A Poldrack, Deanna M Barch, Jason Mitchell, Tor Wager, Anthony D Wagner, Joseph T Devlin, Chad Cumba, Oluwasanmi Koyejo, and Michael Milham. Toward open sharing of task-based fmri data: the openfmri project. *Frontiers in neuroinformatics*, 7:12, 2013.