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 13, 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 regions may have high readings, but not be statistically significant. Also, nearby readings are likely correlated. If researchers wanted to identify clusters of voxels which are significant over time, this introduces an additional challenge.

XXX explain what previous work has done. e.g. reference Rina's paper [FBR15]

For our purposes, we are going to assume that the intensity of the reading is given to us in the form of a p-value between 0 and 1. This corresponds to the hypothesis that at time i voxel j is significant . Just like Rina's paper.

This has been done in the literature

In our project, we will 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?

Here is an outline on how we'll tackle these two topics.

Adaptively identifying regions of interest Research has already worked on how to identify significant voxels of interest and also regions if those regions are already given to the statistician. One particular issue we want to focus on is identifying "regions of interest" adaptively, instead of using predefined ones. Here's one idea we have for doing so.

If we assume that the data is generated according to a Hierarchial Bayesian process and then identify the posterior distribution, we automatically answer the question without any multiple testing problems.

We assume that the data is generated in the following way.

1. For each location ζ_i, generate a prior distribution F_i which is drawn from a prior Φ on the set of prior distributions which have clusters.

2. For each location ζ_i generate the intensity according to the prior F_i

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

Second question: Resample from the data and see how "different" nonparametric algorithm is.

What datasets will we be working with? Do we have any baselines for comparison? First test will be Rina's data.

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.