

Discovering signals in fMRI data; a Bayesian nonparametric approach

Final project for STAT308

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

March 2, 2017

1 Introduction

In this paper we 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 tiny blocks 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. 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 hungry subjects are shown pictures of food.

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 hundreds of thousands 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 heteroscedastic. Identifying significant clusters (not just individual voxels) introduces an additional challenge.

In this paper, we explore a way of estimating significant clusters in fMRI using a Bayesian method.

2 Method

Let $\mathcal{P} = [0, 1]$ be our sample space and let \mathcal{D} represent location-time space, (for example $\mathcal{D} = \mathbf{R}^3 \times \mathbf{R}^+$). Suppose we have n data points $(p_1, d_1), \dots, (p_n, d_n)$ where $p_i \in \mathcal{P}$ and $d_i \in \mathcal{D}$. The p-value p_i represents our belief in d_i being a null feature. That is, our belief that the brain voxel at that point in time is not correlated with the stimulus.

Signals (features associated with the stimulus) in fMRI data typically tend to be sparse. That is, most of the ω_i are likely null signals (features not associated with the stimulus)

and so are drawn from a $\text{Uniform}(0, 1)$ distribution. But, there are typically a few number of unknown signal clusters, let's say k , which are clustered spatially in \mathcal{D} . These ω_i are drawn from some $\text{Beta}(1/\beta_j, \beta_j)$ distribution where $\beta_j > 2$ for $j = 1, \dots, k$.

We encode these beliefs in the following prior for our data:

1. (number of signal clusters) $k \sim \text{Truncated Poisson}(\lambda_i, 1, k_{max})$
2. (signal centers) $c_j \sim \text{Uniform}(\mathcal{D})$ for $j = 1, \dots, k$
3. (signal radius) $r_j \sim \text{Truncated Normal}(2, 10, 2, 10)$ for $j = 1, \dots, k$
4. (signal strength) $\beta_j \sim \text{Uniform}(2, 10)$ for $j = 1, \dots, k$.
5. (data in signal clusters) $x_i \sim \text{Beta}(1/\beta_j, \beta_j)$, where x_i is in cluster j
6. (data not in signal clusters) $x_i \sim \text{Uniform}(0, 1)$.

Note that overlapping clusters is okay theoretically and makes sense biologically.

Having established our prior, we sample from the posterior by using a method heavily inspired by the one in [Ste00]. That is, we construct a birth-death Markov chain with stationary distribution given by the posterior of the data.

3 Tests

3.1 Toy-data

3.2 Real data

3.3 Real data2

References

- [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.
- [Ste00] Matthew Stephens. Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. *Annals of statistics*, pages 40–74, 2000.