Generative Models for Decoding Real-Valued Natural Experience in FMRI

Greg J. Stephens
Center for Brain, Mind and Behavior
gstephen@princeton.edu

David M. Blei
Computer Science Department
blei@cs.princeton.edu

Princeton University
Princeton, NJ 08544USA

Functional Magnetic Resonance Imaging (FMRI) provides an unprecedented window into the complex functioning of the human brain, typically detailing the activity of thousands of voxels for hundreds of time points. The interpretation of FMRI is complicated, however, because of the unknown connection between the hemodynamic response and neural activity, and the unknown spatiotemporal characteristics of the cognitive patterns themselves.

Recent work has exploited techniques from machine learning to find patterns of voxel activity related to brain processes (see e.g., [1]). Many of these techniques involve *decoding*, inferring the value or category class of a stimulus $\vec{S}$ given a pattern of voxel activations $\vec{V}$. Decoding can generally be split into two approaches, discriminative and generative [2]. With a discriminative model one learns the conditional distribution $P(\vec{S}|\vec{V})$ directly by minimizing a loss such as minimum classification error. Alternatively, the generative approach obtains this conditional probability through Bayes rule; one posits and fits models for $P(\vec{S})$ and $P(\vec{V}|S)$ instead. Both approaches can reliably establish the existence of sufficient decoding information.

While decoding in FMRI is itself important (e.g. in lie detection [3]) there is also interest in using decoding techniques to reveal patterns of brain activity relevant to cognition. In this case, the low complexity of most stimulus paradigms makes the neural interpretation of many discriminative decoders difficult. For example, a neural network for a simple categorization task may find a multitude of models, characterized by different sets of weights and voxels, all with acceptable performance. Traditional model selection by predictive performance is misguided as it seems unreasonable to assume that the brain is an optimal classifier. Thus, we are left with no principled method of choosing among the many models.

Generative approaches are transparent in their approximations of patterns of brain activity and offer an additional dimension of interpretability to model slection. Also $P(\vec{V}|\vec{S})$ can be directly manipulated experimentally: while stimuli can be controlled, voxels fluctuate according to their own rules. Here, we outline progress in building a class of generative models for decoding real-valued natural stimuli.

We model the conditional probability with Bayes rule,

$$P(\vec{S}|\vec{V}) \propto P(\vec{V}|\vec{S})P(\vec{S}). \tag{1}$$

This equation has two ingredients: (i) a model of how voxels *encode* the stimuli $P(\vec{V}|\vec{S})$ and (ii) a model of the prior distribution of stimuli $P(\vec{S})$. We can consider

successively more realistic approximations of each of these components, ultimately resulting in a decoding model that incorporates reasonable aspects of brain function.

To begin our exploration, we consider the problem of predicting a stimulus from a subject who is watching television. Specifically, we analyze a subset of the data from the Experience Based Cognition (EBC) competition [4]: the rating for the prevalence of faces while a subject is watching an episode of "Home Improvement."

A natural starting point for the encoding distribution is the assumption that voxels are conditionally independent,

$$P(\vec{V}|S) = P(V_1|S)P(V_2|S)...P(V_n|S). \tag{2}$$

This is known as the naive Bayes assumption. While often applied to discrete classification problems, Conditionally independent models are less explored when predicting real values. We model each voxel with a Gaussian linear filter. The mean of the ith voxel at time t is

$$\mu_t^i = \sum_{j=1}^{T} \alpha_j^i V^i(t-j) + \beta^i S(t) \tag{3}$$

The parameter β^i models each voxel's response to the stimulus; the parameters α_j^i model the dynamics of blood flow, which is T-order Markovian. For each voxel model, these parameters are learned from the training data. The ensemble decoding estimate of the stimulus that minimizes the RMS error is a conditional mean.

Voxel selection is an important component to FMRI decoding because many voxels may not respond to the stimulus. A common approach is to choose those voxels that are maximally correlated with the stimulus during training, effectively picking the best linear filters. Some voxels, however, may be poor filters and yet strongly related to the stimulus. Thus, we choose voxels for which the stimulus provides better predictability.

In detail consider model A with the stimulus effect parameter β set to zero, and model B with a non-zero stimulus effect parameter. We consider the difference of jack-knifed residuals at each time point. (Jack-knifed residuals are computed from models that do not include the data point being predicted in the fit.) We choose voxels such that the median of that difference favors model B. This codifies our intuition that the voxels better predicted with the stimulus are those to include in the decoding model.

We discuss exploratory results of feature selection and stimulus decoding within our conditionally independent approximation and compare to common discriminative approaches. Our results are only a first step in a systematic exploration of more complex models and natural generalizations include accounting for the strong local spatial correlations among voxels. While understanding the human brain is a complex endeavor, we hope that our generative approach will strengthen the connection between our knowledge of brain function and the FMRI decoding of natural stimuli.

References

[1] Haynes, J.D. & Rees, G. (2006) Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience* **7** :523-534.

[2] Ng, A. Y & Jordan M. I. (2002) On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. *in NIPS* **14**.

[3] Mohamed FB, Faro SH, Gordon NJ, Platek SM, Ahmad H, Williams JM (2006) Brain mapping of deception and truth telling about an ecologically valid situation: functional MR imaging and polygraph investigation–initial experience. *Radiology* **238**(2):679-88.

[4] http://www.ebc.pitt.edu