A principled approach to decoding

Charles Zheng and Yuval Benjamini

March 16, 2015

Abstract

In functional MRI (fMRI) studies, one presents a sequence of T (possibly repeated) stimuli parameterized by features $x^{(1)}, \ldots, x^{(T)}$, where each $x^{(i)}$ is a p-dimensional vector. The time-varying MRI image is processed to yield corresponding response profiles $y^{(1)}, \ldots, y^{(T)}$, where each $y^{(i)}$ is a vector of V voxel-specific responses. The goal of these studies is to understand the relationship between $x^{(t)}$ and $y^{(t)}$: this goal can be subdivided into the subgoal of learning an *encoding model*, which predicts the response y given the simulus, and the subgoal of learning a *decoding model*, which reconstructs the stimulus given the response y. One could interpret both models as multivariate regression problems, with encoding fitting a model of the form $Y = f(X) + \epsilon$ and decoding fitting a model of the form $X = g(Y) + \varepsilon$. However, the regression formulation is not the only interpretation of the encoding/decoding problem. Notably, Kay *et al* treat the encoding problem as a linear model, but pose the decoding problem as one of *identification*: that is, given stimuli-response pairs $(x^{[i_1]}, y^{(1)}), \ldots, (x^{[i_j]}, y^{(j)})$ where the unobserved $x^{[i]}$ lie in a known set of stimuli $S = \{x^{[1]}, \ldots, x^{[|S|]}\}$, correctly recover the labels $i_1, \ldots, i_j$ given only the responses $y^{(1)}, \ldots, y^{(j)}$. Furthermore, Kay *et al* quantify the quality of the decoding model by the classification rate for the identification problem when S is selected randomly from a larger database of images $\mathcal{S}$ (Kay 2008, Vu 2011). This approach is more suited for the goal of identifying *natural images* from fMRI responses, and has been adopted by numerous fMRI studies (Chen 2013). Such studies usually use a combination of multivariate linear or nonlinear models and feature selection to implement the decoding model. However, such studies have not explicitly motivated their decoding models based on the criterion of maxmizing correct classification for random stimuli subsets. We proposed a principled approach to decoding, wherein we formulate a decoding model which optimally maximizes the identification perfomance of the model. Our approach is based on a theoretical analysis of the identification perfomance of a linear model, resulting

in an approximate measure of identification performance which can be tractably optimized in training data.

1 Introduction

2 References

- Kay, KN., Naselaris, T., Prenger, R. J., and Gallant, J. L. "Identifying natural images from human brain activity". *Nature* (2008)
- Vu, V. Q., Ravikumar, P., Naselaris, T., Kay, K. N., and Yu, B. "Encoding and decoding V1 fMRI responses to natural images with sparse nonparametric models", *The Annals of Applied Statistics*. (2011)
- Chen, M., Han, J,. Hu, X., Jiang, Xi., Guo, L. and Liu, T. "Survey of encoding and decoding of visual stimulus via fMRI: an image analysis perspective." *Brain Imaging and Behavior*. (2014)
- Schoenmakers, S., Barth, M., Heskes, T., van Gerven, M., "Linear reconstruction of percieved images from human brain activity" *NeuroImage* (2013)