Improving Cognitive State Detection Using Supervised Learning

Jennifer Alch

February 16 2015

1 Motivation and Background

Functional Magnetic Resonance Imaging (fMRI) is a powerful tool used for measuring brain activity that's capable of detecting changes in blood oxygenation. fMRIS are especially useful in the realm of mental health because they are one of the only tools that can safely provide insight into cognitive processes such as memory formation, language acquisition, and emotional development. Despite the strong benefits of fMRIs such as their high spatial and temporal resolution and the fact that they do not rely upon radiation technology, critics have chastised the foundation upon which further statistical analysis relies. Moreover, the analysis of fMRIs is based on a Generalized Linear Model approach. In order to accurately utilize this approach, a number of assumptions about the data must be met, and if not met, any analysis and inference made could be significantly biased. Each image contains 5000 voxels (volume and pixel) showing activation across different portions of the brain.

2 Tools

I will be working with brain imaging data provided by Carnegie Mellon's Center for Cognitive Brain Imaging, which contains a time series of brain activation using functional magnetic resonance imaging. One image is produced every 500 milliseconds while 12 human subjects perform 40 sentence to picture description techniques. Subjects must read a sentence and examine a corresponding picture with the task of deciding whether or not the sentence describes the contents of the picture or not.

3 Methods

My goal is to improve the quality of potential analyses and inferences drawn from fMRIs, and in order to do so, I will be concentrating on three major areas of contention: Dimensionality reduction, feature selection and Bayes network classification. Dimension reduction is necessary in order to improve the classification of the fMRI data. Because of the high dimension of inputs (i.e. 5000 voxels per each of the eight images for each of 12 subjects), in order to minimize potential error we must determine a way to find a lower dimensional representation that still captures enough of the variance in the data. I will be comparing the merits of applying principal component analysis versus t-distributed stochastic neighbor embedding to the data in order to determine which provides a better alternative for dimensionality reduction. Next I will handle the task of feature selection. The brain imaging data gives us information as to the activity levels of various parts of the brain during the reading of the sentences and subsequent reflection upon the sentences and their correspondence with the pictures. Feature selection is crucial as it may improve the performance of any classification algorithm. I will be using methods based on both greedy forward selection and greedy backward elimination and seeing which contributes more to the overall classification accuracy. Lastly I will be developing a new model for classification by training a Bayes network. As stated above, fMRIs have long been criticized for the conditional independence assumptions their statistical analyses rely upon. In formulating Bayesian network classifiers by utilizing a tree structure, I hope to improve the accuracy of the overall results to determine when a subject is reading a sentence versus looking at a picture as opposed to linking the two.

4 Resources

Brain Imaging Data: http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/

Cognitive Classification Articles:

https://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-73/www/papers/XueruiReport-10-2002.pdf

http://papers.nips.cc/paper/2449-training-fmri-classifiers-to-detect-cognitive-states-across-multiple-human-subjects.pdf

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0001394

Feature Selection: http://www.stat.rutgers.edu/home/tzhang/papers/nips08-foba.pdf