CS 289A Project Proposal

Jiaying Shi, Mo Zhou

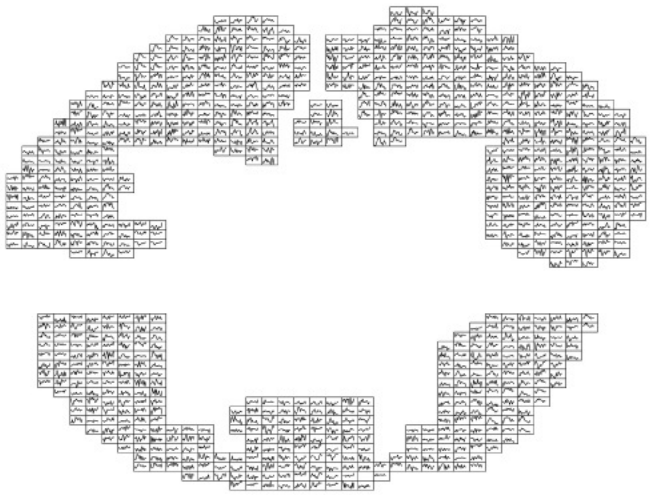
SID: 24978491, 21242515

Topic

Decode human brain activity from brain imaging data (fMRI)

Data Description

The data set is obtained from StarPlus fMRI data website[1]. The data set contains .mat files of seven human subjects. For the data of each subject, three variables are defined: info, data and meta. The variable 'info', which is a struct array, defines the experiment in terms of a sequence of a certain number of time intervals. It contains the information of each 'trial' such as the sentence presented to the subject, whether the picture is presented before or after the sentence and the action of the subject. During most of the time intervals, the human subject views a single picture and a single sentence and presses a button to indicate if the picture is correctly described by the sentence. The picture has two tokens where one is above the other. The possible tokens are star(s), plus(p) and dollar(d). The variable 'data' contains the raw observed data. The raw data is a sequence of images of the subject's brain activity collected over time. It is a cell array with one cell per 'trial'. The variable 'meta' presents information about the data set including the number of trials, the number of voxels, the total number of images and so on. A sample image is shown as below:



[1] fMRI data website: http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/

Project goal

We are interested in answering the following questions concerning brain reaction to images:

1. Does brain activity vary when the human subject is looking at a picture or a sentence? Would it be possible to train a classifier?

2. Does brain activity vary when the human subject is making decision? i.e. do we observe different activity when the human subject gives different answers to how well the sentence describes the image?

3. Does brain activity vary when the picture is showing food, buildings, animals, etc?

We aim to answer the above questions by developing classifiers and test on new data.

Consider the dimension of the problem, we would first use filters to reduce noise. Then we would apply dimensionality reduction methods to select the best features. After that, we would apply several classification techniques and use CV to compare their performance.

Possible methods we are considering:

PCA, SVM, Bayesian classifier, kNN, Lasso, Ridge