Modeling of Semantic Representation in the Brain Using fMRI Response

Sinha, Rishi
rishizsinha

Mo, Cindy
cxmo

Agrawal, Raj
raj4

Wang, Yuan (Aria)
ariaaay

Dong, Yucheng (Steve)
yuchengdong

November 12, 2015

Abstract

We attempt to use fMRI data to find correlations between brain activity and events occurring in an audio movie stimuli and make prediction about events happened in the movie. Previous study on semantic modeling has shown that with natural movie stimuli, it is able to predict the semantics categories only using the BOLD responses [?]. The data we used in this projects include a detailed transcript of an audio movie ('Forest Gump') as well as scenes description divided by time windows, and the BOLD response of 20 subjects listening to the audio movie [?]. There are two modeling goals for this project. The first one is to model semantic category in the audio description of the movie using BOLD response. We aim to predict what words/category a subject is listening to using the trained regression model on brain activity. The second one is to model the time-course activity using different scenes, aiming to predict what scenes was happening in the audio movie based on the BOLD response at the time. To accomplish this project, it requires preprocessing of BOLD response as well as the movie description, which involves NLP (Natural Language Processing) of the script using WordNet, as well as modeling method, such as general linear regression and ridge regression.

1 Introduction

Dustin E. Stansbury's article "Natural Scene Statistics Account for the Representation of Scene Categories in Human Visual Cortex" describes the method by which the human brain aggregate information about subjects to represent scene categories. Using statistical learning methods, researchers can learn categories of certain objects, and model fMRI brain signals when human subjects are viewed images of scenes using the learned categories. In our study, instead of images of scenes, we have audio descriptions of the scenes that are presented to human subjects. However, we hypothesize that the response would still stimulate different parts of the brain which can be modeled. We attempt to decode fMRI data by first taking the voxel responses and fit using a regression model to find decoder weights and then predict category probabilities. Categories can determined through WordNet, which classifies words based on hierarchical similarities. It groups English words into sets of synonyms called synsets, provides short definitions and usage examples, and records the number of relations among these synonym sets or their members (Wikipedia).

2 Data

2.1 Original Dataset

Our team obtained the "A high-resolution 7-Tesla fMRI dataset from complex natural simulation with an audio movie" dataset from OpenfMRI.org. The dataset belongs to a study conducted using high-resolution functional magnetic resonance (fMRI) to analyze stimulation due to visual stimulii. 20 participants recorded at high field strength (7 Tesla) during prolonged stimulation with an auditory feature film ("Forrest Gump").

The audio movie description was provided in a csv format, with each row containing a start and end time (in seconds) corresponding to the time in the movie, along with a description in German of the events occurring in the movie at that time. One noticeable feature of the dataset upon initial observation is that the description is not continuous – there exist around a 5 seconds gap in the audio events provided. The second CSV file includes the start and end times of each of the 196 distinct movie scenes. In addition, each table row contains whether a scene takes place indoors or outdoors. The last CSV file contains questionnaire responses from the 20 subjects regarding their backgrounds, but we have discarded them from our study.

As for the FMRI data, data for 20 subjects were provided, each of whom watched the movie while being scanned. The FMRI scan were taken every two seconds (TR=2s) for the duration of the movie. All of this data was in nii format, and thus compatible with nibabel libraries in Python. One thing we noticed is that between 8 runs of the fMRI scanning, there are around 6s of repetition of stimuli between each run. We thus discarded four volumes at the end of any preceding segment and at the start of the following segment for any transition between segments according to the instruction in http://studyforrest.org/annotation_timing.html. The other things is that these images have partial brain coverage mostly focused on the auditory cortices in both left and right hemispheres, including frontal and posterior portions of the brain. There is no coverage for the upper portion of the brain where large parts of motor and somato-sensory cortices are located. This case of brain data coverage would potentially affect the way that we are going to model and visualize the brain response.

For each of the 20 subjects, answers to pre-experiment survey questions were made available, asking a variety of questions (e.g. left or right handed).

2.2 Data Used in the Analysis

The survey data was used to select BOLD response from subject 004 to do a first pass modeling attempt. She was chosen partially because her fMRI data was one of the most complete and also for the arbitrary fact that she had perfect pitch. After we successfully implement first regression model, more subject(subject 014, 015) would be downloaded and used.

All of the movie description data as well as the scene description was used, as it is applicable to the FMRI time-course data for all subjects.

The survey data of emotion response was not used beyond this point.

3 Methods

In order to achieve the ability to predict objects and scenes from FMRI images, a predictive model is required. The project was thus divided into a few main stages for independent preprocessing of the movie description and FMRI images, model building, and prediction testing.

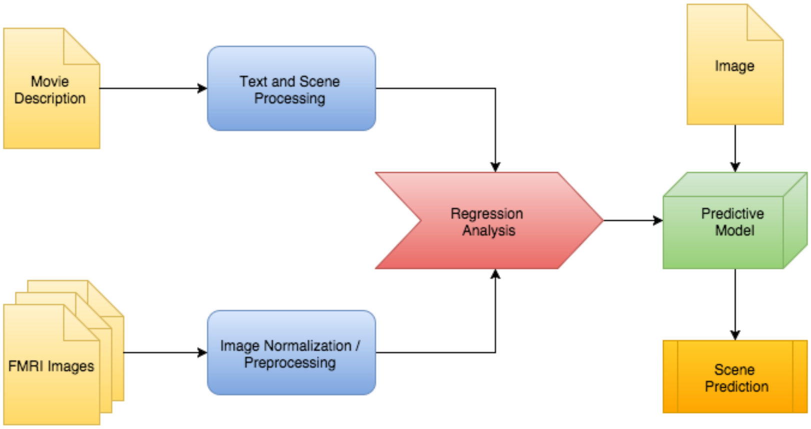

Figure 1: process flow of the project

3.1 Text Preprocessing

As the movie description as provided was in German, we first used Google Translate to translate the description to English before proceeding with any further processing. Though there may be translational errors, it was decided that it was best to transform the original descriptions (versus using other English descriptions) to maintain the original time stamps from the researchers to have optimal alignment between the description and FMRI data.

Due to grammatical differences, we decided to only keep nouns and verbs, and discarded the adjectives and other words including stop-words, which are commonly used words with little meaning or determinable context (e.g. 'and', 'to', 'him'). A publicly available list of stop-words from Princeton University was used for this task. Of the words that remained, a WordNet dictionary was built, which is a popular way to tag words according to a context-specific definition. In this way, words as stand-alone entities will have an unambiguous definition and deeper relationships can be derived from the correlations found.

After aggregating the set of all contextual definitions of words, we build a design matrix composed of the descriptive entries in the movie description. Thus each interval of time with a sentence description of the movie events was treated as a row, and each column represented a context-specific word as a feature. The matrix values are all binary, with a value of '1' indicating that the word is present in the sentence, while '0' indicates that it is not. This was illustrated as a rectangular image with a white square for '1' and black for '0'. As can be seen in Figure 2 below, this matrix is relatively sparse. Finally, in order to format the data such that it correlates explicitly to the FMRI images, the intervals were split into two-second intervals to create a one-to-one representation between description objects and images.

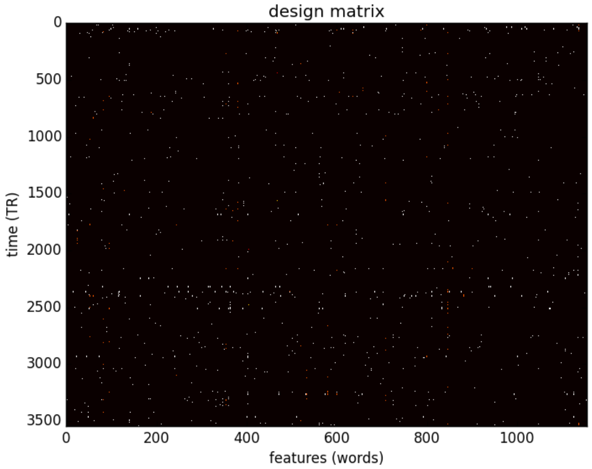

Figure 2: A visualization of design matrix for the semantic modeling

3.2 FMRI Preprocessing

We first looked at different measures of spread such as IQR, standard deviation, and RMS for each voxel across time. The following plots below show the standard deviation and RMS differences of the voxel-time courses. It is interesting that the outliers seem to have, to some degree, structure in their location; the outliers seem to clump together in certain places. In order to deal with this, we are going to run different regression models on the BOLD signal and do residual analysis. Hopefully, this will give us a better idea about the correlation structure in our data. We also had to deal with the overlap of movie scenes. More specifically, during the beginning of a new segment/run, researchers replayed the last six seconds of the audiotape of the previous run. To deal with this, we dropped the last four volumes for runs two through seven. We also experimented with PCA. It was hard to extract anything very informative since our dataset is so large. However, when we reach the analysis stage, there is a high chance that PCA will be used (the runtime of some packages may take too long to run on a data set of this size). If this is the case, PCA will potentially allow us to extract the relevant signal areas without sacrificing too much loss of information.

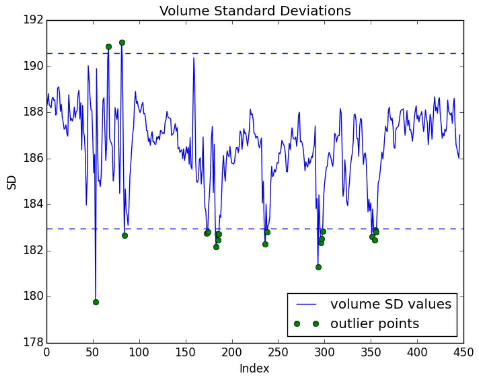

Figure 3: Standard deviation

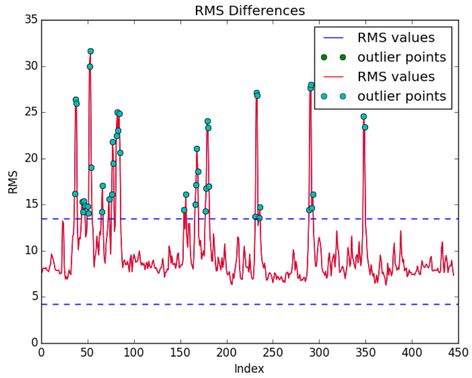

Figure 4: RMS(root mean squared differences)

3.3 Bold response modeling

3.3.1 Voxelwise modeling

We use ridge regression for voxelwise modeling of the BOLD response. The entire 8 runs of brain data are separated into training set (7 runs) and validations set (1 run). 7 runs of training set data are further split into 10 groups to do a 10 fold cross-validation on ridge parameters. And finally all the training set of data is used to model with the corresponding part of design matrix to build a voxel-wise ridge regression model.

3.3.2 Scene modeling

We were provided a cvs file containing the times at which different scenes occurred in the film. Our goal was to use these scenes as a variant of the the on/off neural task course. Here, we would choose say three scene categories as our stimulus which would comprise the on times in the time course for a specific run (all other scenes that occurred during this time interval would be considered off). With this, we could then check correlations between this time course and the voxel time course, perhaps revealing that certain scenes have different impacts on BOLD signals. We also hope that this will add another layer of comparison between our predictions of words based on BOLD activity.

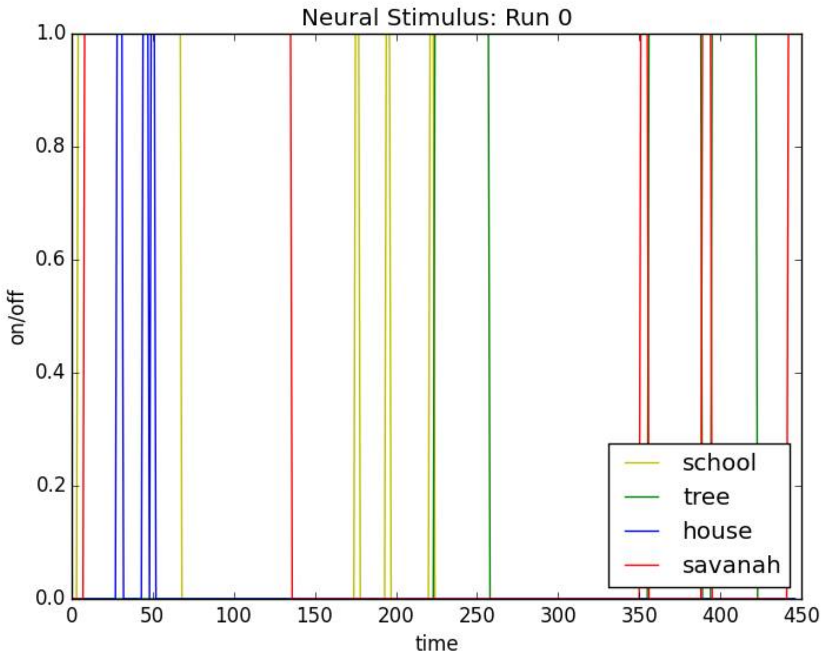

Figure 5: An example of scenes conditions across time

3.4 Predictive Testing

In voxel-wise modeling, correlations between predicted BOLD response and actual response are used to evaluate the model performance. Potentially, we could also identify areas that are predicted well by each categories and find the ”brain representation” of each semantic categories. Finally, the areas in the brain that are well predicted would be used to predict category labels from BOLD response.

4 Results

(To be continued.)

5 Discussion

There are places where we expected errors to be introduce to the project. For example, we use Google translate to conver the german description into English without supervising, which might induce random errors into the design matrix. And since not all words in the description has an one to one correspondence in WordNet, we took out some words in the process of generating the design matrix.

Furthermore, because the movie description we obtained from openfmri.org is not detailedly mark with each TR. It is not able for us to infer what words are said to the subjects inside a specific TR. We have to infer that brain activity are activated by the whole sentence within a certain time window instead of by each single words at a single time point.

References