The Neural Basis of Loss Aversion in Decision-Making Under Risk

Chang, Siyao	Gong, Boying	Hsieh, Benjamin	Qiu, Brian	Zhu, Jiang
`changsiyao`	`boyinggong`	`BenjaminHsieh`	`brianqiu`	`pigriver123`

November 30, 2015

Abstract

Our paper is the *Neural Basis of Loss Aversion in Decision-Making Under Risk* [1]. The experiment investigates the phenomenon of loss aversion - where individuals decisions are influenced by the amount of potential loss more than they are by the amount of potential gains. The experiment involved giving 16 subjects 256 combinations of gain/loss of dollars with a 50/50 chance. The subject's decisions of whether to accept or reject of each of the proposed gambles were measured as well as their brain activity in the fMRI machine. The behavioral data file records the subjects responses and to which combination of values, and the BOLD data file records the subjects neurological reponse in 4-dimensional fMRI data.

1 Introduction

Most of the preliminary work we've done so far involves loading in the data and running basic summary statistics. Our group has also written functions to pull and graph the dvars (RMS of the signal derivatives) and the framewise displacement. We use that in conjunction with the mean signal of the BOLD data to identify outliers and use that in the near future to process our data further. We have also begun rough functions to calculate the betas and begin logistic regression.

2 Data

2.1 Overview

The study used 16 right-handed, healthy, English-speaking participants recruited through ads posted on UCLA. Out of 16 subjects, 9 were female and the mean age was 22 $\pm$ 2.9 years. [1]

2.2 Behavorial Data

The behavioral data consists of each subject undergoing 3 trial runs for the "gamble" task, in which each subject is presented with a combination of potential monetary gains and losses given a 50/50 chance of to win/loss. Each trial run consists of 86 different combinations of rewards/penalities spread out accross 474 seconds. Intervals between each onset of task range from 4 to 8 seconds. Subjects were given the 4 choices in reponse to each gambling proposal:

1. Strong Accept
2. Weak Accept
3. Weak Reject
4. Strong Reject

The choices are recorded by denoting reponse numbers 1, 2, 3, and 4, respectively. Furthermore, the response time for each gambling decision was recorded in seconds.

2.3 BOLD Data

Blood-oxygen-level dependent (BOLD) imaging data were collected from each subject as he/she performed the gamble tasks. 240 time scans were done on each run with a time between each scan of 2 seconds. So total scanning time is 480 seconds. Each scan consists of a snapshot consisting of a64 by 64 by 34 image matrix.

There are also 4 model conditions, with events corresponding to

1. Task
2. Parametric Gain
3. Parametric Loss
4. Distance from Indifference

3 Methods

3.1 Models and analysis

We use linear models to find the relationship between behavioral and nueral loss aversion cross participants as well as how participants react to different loss and gain level. Below we illustrate our model using simple multiple linear regression form. We may implement a mixed-effects model treating partipants as a random effect. Moreover, we may use the robust regression to reduce the influence of outliers.

3.1.1 Behavioral analysis

We fit a Logistic regression model on the behavioral data to examine how the response of individuals relates to the size of potential gain and loss of a gamble. Following is the model:

$$logit(Y_{resp}) = \beta_0 + \beta_{loss} * X_{loss} + \beta_{gain} * X_{gain} + \epsilon \tag{1}$$

where X_{loss} and X_{gain} are the potential loss and gain value seperately, Y_{resp} is a categorical independent variable representing the subjects' decision on whether to accept or reject the gambles:

$$Y_{resp} = \begin{cases} 1 & \text{If the subject accepted the gamble.} \\ 0 & \text{If the subject rejected the gamble.} \end{cases}$$

Then we calculate the the behavioral loss aversion (λ) for each subject as follows, note that for simplicity, we collapse 3 runs into one model for each participant.

$$\lambda = -\beta_{loss}/\beta_{gain} \tag{2}$$

We use λ as the metric for the degree of loss aversion for each participant. We have used R to fit the Logistic model, just as what the authors did in the paper, and we achieved almost the same results as the paper presented.

3.1.2 Linear Regression on BOLD data

For each voxel i, we fit a multiple linear model:

$$Y_i = \beta_{i,0} + \beta_{i,loss} * X_{loss} + \beta_{i,gain} * X_{gain} + \epsilon_i \tag{3}$$

where Y_i is the BOLD data of voxel i. For each voxel, we calculate the neural loss aversion η_i:

$$\eta_i = (-\beta_{loss}) - \beta_{gain} \tag{4}$$

Using the voxelwise neural loss aversion, we do a region-specific analysis on BOLD data for each participant. That is, we plot a heat map of η_i and $\beta_{i,loss}$, $\beta_{i,gain}$ for each participant to find out the regions with significant activation and regions which show a significant positive or negative correlation with increasing loss or gain levels.

3.1.3 Whole brain analysis of correlation between neural activity and behavioral response across participants

We then apply the above model on the standard brain to analysis the neural activity and behavioral response across participants. For each participant, we pick up several regions with highest activation level, calculate the mean neural loss aversion $\bar{\eta}$ within these specific region. Thus we could examine the relationship between neural activity and behavioral using the following regression model:

$$\lambda = \alpha_0 + \alpha_1 * \eta + \epsilon \tag{5}$$

where the sample size is the number of participants(16).

3.1.4 Cross-validation

We fit linear models for each voxel for each participant. For each linear model, we do a k-fold cross-validation. Since the sample size for each linear regression model range from 80-90, we choose to use 10 fold cross-validation, which means the original sample is randomly pertitioned into 10 equal sized subsamples.
In the behavioral analysis using Logistic regression, since the responce variables are binary, we calculate the misclassification error rate to summarize the fit. In the neural linear regression model using BOLD data, we use the mean squared error to summarize the errors.

3.1.5 Inferences on Data

After fitting regression models on our BOLD and behavioral data, we would try assessing and validating our models. In order to do this, we would calculate for the residual sum of squares for our model. We have to do three tests for the model. The first one is that we calculate the t-statistics and p-value for our beta coefficients to check whether our beta parameters are statistically significant at a significance level of 5%. The second one is that we calculate the residuals of this linear model and check whether it follows a normal distribution. The third one is that we calculate the R-Squared value and the adjusted R-squared value to see whether the values are good for the linear regression model.

3.2 Explanation on model simplification

3.2.1 Use of Data

First of all, for simplicity reasons, we are not using all the regressors the paper used. The model in the paper performed regression on the BOLD data with gain, loss and euclidean distance to indifference. In our model, we are leaving out the regressor euclidean distance to indifference. The paper and its supplement material didn't document the exact way the authors calculated this parameter; we are having a hard time reproducing this parameter. Therefore, we decide to leave out this parameter when doing our own regression.

3.2.2 Simplification of regression on BOLD data

We plan on simplifying the model on neural data. In the original data analysis, the authors performed a mixed effect model when regressing the potential gain and loss values against the BOLD data across runs, since there are three different runs for each subject and the authors were trying to incorporate all three runs into one model. The mixed effect model adds a random effects term, which is associated with individual experimental units drawn at random from a population. In this case, it measures the difference between the average brain activation in run i and the average brain activation in all three runs.

We are simplifying the model because it is much easier to perform a simple linear regression in python. In addition, we do not have a great deal of understanding of fMRI data, so simple linear model would suffice when we are only performing exploratory data analysis and looking for obvious pattern in the data.

After looking at the initial result from our linear regression model, we can decide whether we want to further explore the relationship between the dependent variable (BOLD data) and the independent variables (gain and loss) and whether we want to continue to fit a mixed effect model.

3.3 Issues with analyses and potential solutions

3.3.1 Selecting specific regions to further explore correlation between neural and behavioral activity

Since we have no knowledge on the sections of brain that might experience large difference in activation, it is hard for us to pick the regions to deeper explore the correspondence between neural and behavioral loss aversion.

There are two potential ways to deal with this issue. The first one is to read more paper and related articles to learn which parts of the brain are likely to react in our given scenario – faced with potential gain and loss combinations. Another way to deal with the issue to to fit a regression for every part of the brain and look for the areas with higher correspondence (higher slope). Then, we select and graph a few areas with the most significant positive or negative correlation between the parametric response to potential losses and behavioral loss aversion (ln()) across participants.

3.3.2 Producing heat map

Another issue that we are facing during our project is finding the same region to plot for each participant. We see that each region of the brain has its own standard coordinates. However, without much knowledge of fMRI, we are not sure how to use these standard coordinates to locate the regions of the brain.

From our understanding, each subject's brain is mapped onto a standard brain and we then use the coordinates for the standard brain to extract data from the areas we are interested in. However, currently, we don't have the skill to perform this step.

3.3.3 Further Research

We fit a linear regression model combining behavioral and BOLD data to examine the relationship of correlation between neural activity and behavioral response, we use another method which is different from what is mentioned in the paper. We add the behavioral response to the regression model on BOLD data as a predictor. We use the original 4-level response as stated below.

Moreover, if the three tests we do for the linear regression model is bad. We can plot the independents and the dependents on plots to see whether they fit a model that is different from linear regression models. There may be another reason why the performance of linear regression models are bad which is that we simplify our model that we didnt try a mixed model as the researchers in the paper did.

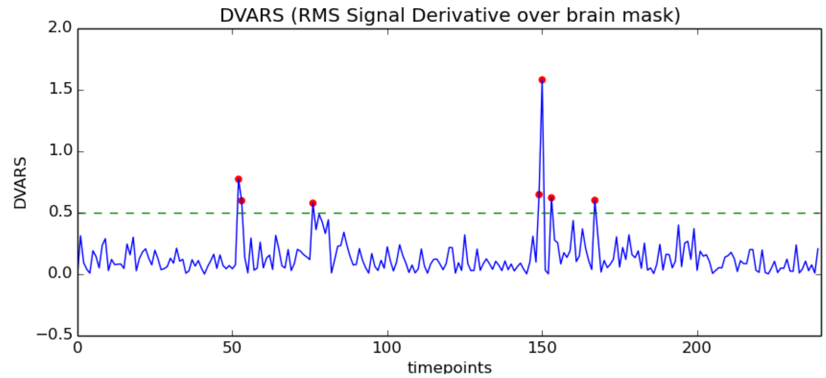
behavioral response	strongly accept	weakly accept	weakly reject	strongly reject
X_{behav}	1	2	3	4

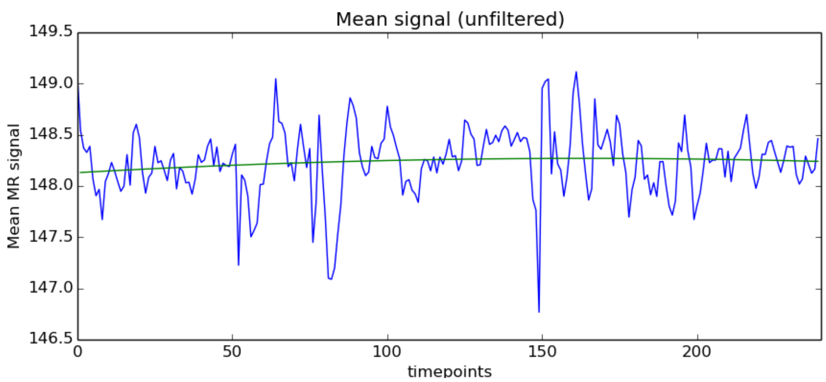
And the models are following:

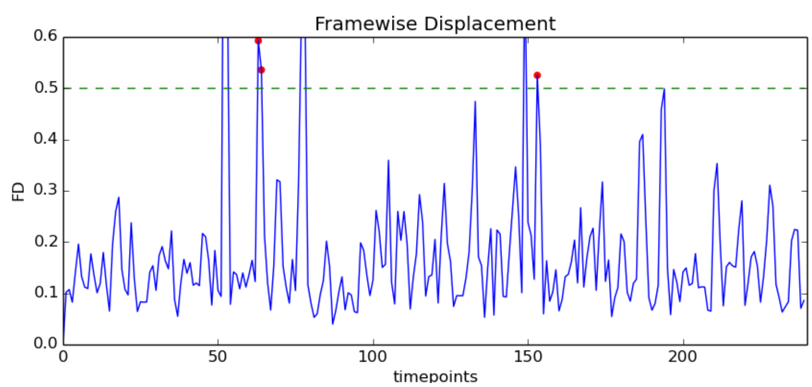
$$Y_i = \beta_{i,0} + \beta_{i,behav} * X_{behav} + \epsilon_i \tag{6}$$

However, since the response and level of loss and gain are potentially correlated, we might need to use stepwise regression to choose the best predictor from the regression model presented above.

4 Results







Red points signify outliers in all three figures.

5 Discussion of Challenges

One of the major challenges is trying to make this project as reproducible as possible while following guidlines on documentation, testing functions, and attempting to produce the results of the paper using our limited understanding of fMRI data. Travis CI bugs with various versions of python, coverage failures, and errors with directory/path locations often hinder the process of smooth workflows. Collaboration between five group members is no doubt difficult as we found it hard to come up with an attainable final goal that is still rewarding.

Technically, most of us are new to python programming and reseach using git workflows, thus we have only a preliminary understanding of the various python resources available for our use. Additionally, lack of statistical understanding of some aspects of the paper has urged us to do independent research. Yet the disconnent between theory and implementation has been a major obstacle as we try to put our knowledge into practice.

Some problems can be solved or alleviated by defined checkpoints and making the effort to read and re-read the paper and ask questions. Further, as we familiarize ourselves more and more with various python modules and toolkits, results can be easier to attain and interpret.

References

[1] S. M. Tom et al., *The neural basis of loss aversion in decision-making under risk*, Science, 315 (2007), pp. 515–518.