Project Epsilon progress presentation

Min Gu Jo, Soazig Kaam, Zhuangdi Li, Timothy Yu, Ye Zhi

November 12, 2015

Background

The Paper

- The Neural Basis of Loss Aversion in Decision-Making Under Risk
- from OpenFMRI.org
- https://www.openfmri.org/dataset/ds000005
- ds005

The Study

- 16 subjects were presented gambling situations with 50% chance of winning
- Each of the 255 trials was associated with amounts of potential gains and losses (in $)
- Subjects ranked their level of willingness to accept or reject the gamble on 4 point likert scale
    1. Strongly accept
    2. Weakly accept
    3. Weakly reject
    4. Strongly reject

Our Progress

Initial work

- Downloaded data and used the checksums.txt to validate
- Behavior data: merged three runs for each subject and took out observations with -1 in "respcat"(maybe an error in the experiment)
- Bold data: unzip all files

Behavior data (1/3)

- Simple plots and summary statistics for each subject

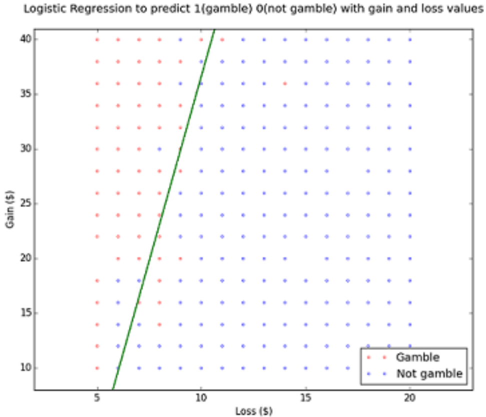
subject number 3 (total runs):

	onset	gain	loss	PTval	respnum	respcat
count	256.000000	256.000000	256.000000	256.000000	256.000000	256.000000
mean	234.851562	25.000000	12.500000	-59.416875	2.953125	0.230469
std	142.145719	9.237604	4.618802	32.531316	0.848153	0.431151
min	0.000000	10.000000	5.000000	-125.070000	0.000000	-1.000000
25%	109.000000	17.500000	8.750000	-86.240000	3.000000	0.000000
50%	232.000000	25.000000	12.500000	-59.415000	3.000000	0.000000
75%	364.000000	32.500000	16.250000	-32.592500	4.000000	0.000000
max	476.000000	40.000000	20.000000	6.230000	4.000000	1.000000

	RT	ratio
count	256.000000	256.000000
mean	1.143664	2.366260
std	0.335413	1.437319
min	0.000000	0.500000
25%	0.945250	1.333333
50%	1.105500	2.000000
75%	1.293500	3.000000
max	2.708000	8.000000

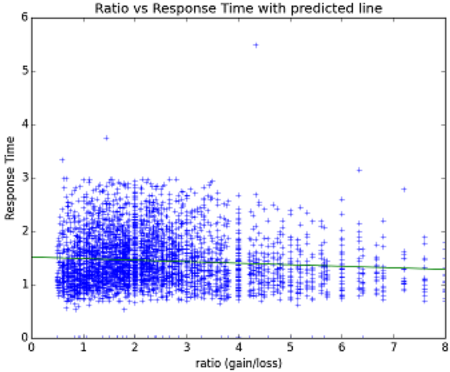
Behavior data (2/3)

- Logistic regression between Response(1/0 or Accept/Reject) and Gain/Loss *Scientific Question: If gain/loss would be significant for whether individuals would like to participate in the gamble* Result: According to our analysis, the decision to whether take the gamble of most of subjects, in general, is more affected by loss amount rather than by gain amount.



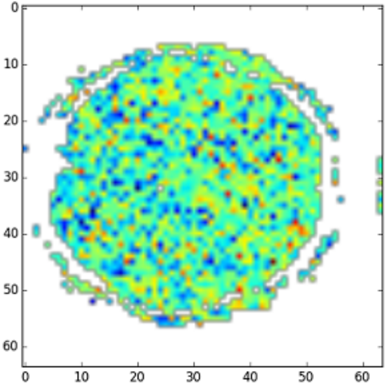
Behavior data (2/3)

- Linear Regression between Response time and Gain/Loss/Ratio *Result: Ratio is a significant predictor and people would actually care more about loss than gain



BOLD data

- Reproduced Quality Assurance Plots: mean, fd and dvars
- Calculated the correlation between task-on/task-off vectors and voxel time courses to identify the active region of the brain



Our research plan

Behavior data

- Use other classification methods than logistic regression to predict gamble/not gamble
- Explore correlation between neural activity (image data) and behavior data (survey data)
  - Can we use the image to predict behavior data ?

Modeling voxels for each participant

- Use convolved hemodynamic response and linear regression
- Train of the model using two randomly selected runs
- Validating the model using the third run
- Investigate different Gamma function shape parameters

BOLD images data analysis

- Identify brain activation region associated with decision making
  - Use K-means or other classification
  - Assess the sensitivity to gail and loss
- Compare across subject and runs (for now, only for 1 subject and 1 run)

Model validation

- Check assumptions of the regression models: normality, independance, equal variance
- Use cross validations to test our model accuracy

The problems we've faced

Understand the data

- Lack documentation of the data structure and meaning of variables:
  - spend much time reading through the paper and searching the website
  - still some problems unsolved: e.g. PTval in the behavior data of each run of each subject
- fMRI: technical field study difficult to understand
- Insufficient description of the analysis methods used to reproduce the work
  - e.g. QA section we can not match the scale/value of some variables
  - e.g. analysis of the individual logistic regressions for behavioral data

Coding

- Reproducibility:
  - For purpose of reproducible, we need to write a lot of functions, scripts and tests to travel through different folders, unzip and load the files – checksum.txt match hash and files
- Git:
  - Doing version control; merge results; branch management
  - Keep track of others' code – solved by adding more descriptions when making pull request – review

Data Analysis

- Image data analysis
  - Difficulties when working on data across subjects and runs
- Length of behavior data does not match the length of image data
  - Fill in 0 or NA?
- Convolution
  - How to create a convolution matrix
  - Can not determine the parameters of the gamma distribution (Tim sent a graph HRF vs time)
    - Using the common 2 Gamma functions (shape parameter:6 and 12) not reasonable
    - Find other sources (literature review)

Inference from data: future attempt to validate your model

- Assumption check
  - Analysis of residuals
  - Normality
- p-values
- Validation using R2
- Cross-validation

Our Process

Challenges

- Understanding the paper and the fMRI data
- Workflow on git and version control management (some problem with branch management and Travis CI)
- Difficulty with code review process
- Conflicting schedules but manage to set up a weekly meeting
- Overcome the overwhelming amount of data/work by working in smaller groups

Improving reproducibility

- Adopt a systematic method for code organization (functions, script, test) and writing (PEP 0008)
- Improve our process of code review
- Generate the support documentation to improve usability of our project
- Exchange more with groups working on the same topic

Feedback on the class

Feedback

- A good model with 3 (or 4) supervisors with their own expertise
- Would like more exposure to machine learning techniques (also the Basic linear model is one example)
- Lectures on git workflow and collaboration were very useful but fast-paced.
- Lecture on linear algebra was a good refresher but fast-paced and too theoretical
- Would like more linear regressions course focusing on the implementation

Ideas for improvement

- Supporting the lecture with slides or handouts with the fundamentals (e.g. git command for collaborative work) we can refer to after class
- Provide a support document with the mostly used statistics definitions for a good analysis design and interpretation