Project Name

FMRI Study Analysis by the meaning of Nouns

4th year 1st semester
Session: 2015-2016

Submitted by:
Roll: 15 Asif Zaman
Roll: 35 Marzan Binte Abid
Roll: 64 Esha Ferdous

Date of Submission: 25/04/2019

Submitted To:
Dr. Moinul Islam Zaber
Associate Professor, Dept. of CSE, DU

Contents

1.Introduction

Functional **M**agnetic **R**esonance **I**maging **(fMRI)** is one way to look inside your body without having to cut open it. An MRI machine takes pictures of places in your body that contain water, and the detail in these images comes from the ways that different tissues interfere with the electromagnetic waves coming from water molecules.

In this way, fMRI captures an image of human body parts at different states and later analyze for further findings. Functional magnetic resonance imaging (fMRI) plays an important role in medical imaging analysis, biological and neuroscience research and practice. In this project, we worked with FMRI data to study neurological activity associated with the meanings of noun.

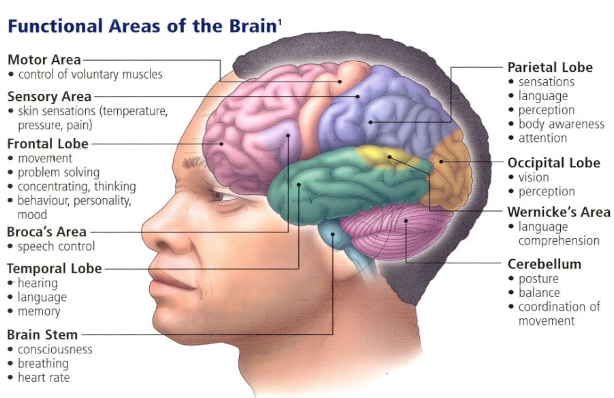


Fig 1: Different functional areas of the brain and their activity

We studied the paper **Predicting Human Brain Activity Associated with the Meanings of Nouns, Tom M. Mitchell, Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang,**

Vicente L. Malave, Robert A. Mason, Marcel Adam and it's dataset. Then we studied some papers related to this reference paper and tried to find their techniques for the work. Finally, we implemented a machine learning algorithm and evaluated their performance.

2. Dataset Study and Analysis

The source of the dataset is:

http://www.cs.cmu.edu/afs/cs/project/theo73/www/science2008/data.html

Nine right-handed adults from the Carnegie Mellon community (provide students with opportunities to enhance their personal growth and development as they progress through their college years). participated in this experiment.The stimuli were line drawings and noun labels of 60 concrete objects from 12 semantic categories with 5 exemplars per category. The entire set of 60 stimuli was presented 6 times during the scanning session, in a different random order each time. Participants silently viewed the stimuli and were asked to think of the same item properties consistently across the 6 presentations. Each stimulus was presented for 3s, followed by a 7s rest period, during which the participants were instructed to fixate on an X displayed in the center of the screen. There were two additional presentations of the fixation, 31s each, at the beginning and at the end of each session, to provide a baseline measure of activity.

Each of the participants' data contains in a mat file that holds six variables.

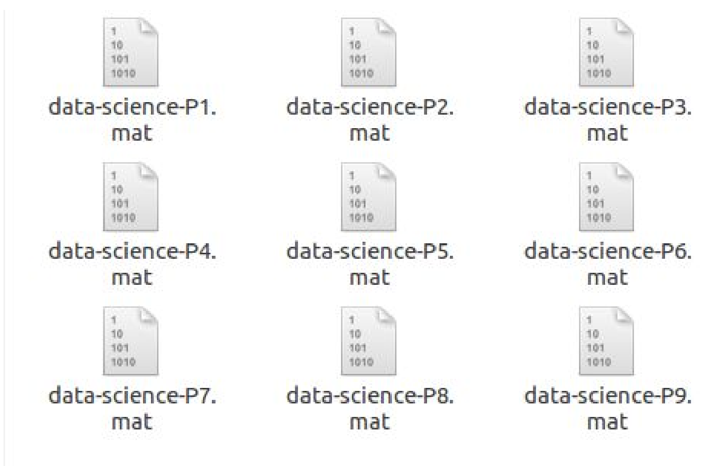


Fig 2:Snapsort of mat files

2.1. Variables:

- Header
- Version
- Globals
- Meta
- Info
- Data

Header:

Header Provides information about file type, platform and creation date and time. Example: b'MATLAB 5.0 MAT-file, Platform: GLNXA64, Created on: Tue Oct 20 19:12:48 2009'

Version:

Version denotes the version of the file. Example: 1.0

Globals:

Globals represents the list of global variables in the file which is empty for the dataset files.example : (empty list)

Meta:

This variable provides information about the data set, contains these fields below:

- Study: Gives the name of the fMRI study.
  Example: study: 'science'
- Subject: Gives the identifier for the human subject.
  Example: subject: 'P1'

- Ntrials : Gives the number of trials in this dataset.
  Example: ntrials:360
- Nvoxels : Gives the number of voxels (3D pixels) in each image.
  Example: nvoxels:21764
- Dimx : Gives the maximum x coordinate in the brain image. The minimum x coordinate is x=1.
  Example: dimx:51
- Dimy : Gives the maximum y coordinate in the brain image.
  Example: dimy:61
- Dimz: Gives the maximum z coordinate in the brain image.
  Example: dimz:23
- colToCoord(v,:) : Gives the geometric coordinate (x,y,z) of the voxel

corresponding to column v in the data.It is a 21764x3 array.

- coordToCol(x,y,z) : Gives the column index (within the data) of the voxel whose coordinate is (x,y,z).

A fraction of coordToCol and their corresponding column index in data is given below:

X	Y	Z	Column Index
3	30	8	7875
3	30	9	9347
3	31	7	6603
3	31	8	7910
3	32	7	6623
4	19	7	6261
4	19	8	7576
4	20	7	6300
4	20	8	7604

Info:

This variable defines the experiment in terms of a sequence of 'trials'.'info' is a 1x360 struct array, describing the 360 trials, contains these fields below:

- Cond: Gives the condition presented during the trial.
- Cond_number: Gives numeric index of the condition presented during this trial, ranges from 2 to 13 because there are twelve different categories.
- Word: Gives the word presented during this trial. For instance, the word 'hammer' is presented during this trial.
- Word_number: Gives the numeric index of the word presented during this trial, ranges from 1 to 5 because there are five words per condition.

- Epoch gives the number of times this word has been presented. For instance, epoch=1 denotes this is the first time the word 'hammer' is presented.

Example:

The 100th entry corresponds to ['furniture','7','table','5','2'] so for the particular trial we have

- 'furniture' as a condition
- '7' corresponds to the condition number of 'furniture'
- 'table' falls under 'furniture' condition as one of the five words of the condition
- '5' corresponds to the word number of 'table'
- '2' corresponding to epoch indicates that the word 'table' has been presented two times

Data:
This variable contains the raw observed data. The fMRI data is a sequence of images collected over time. It is a [360x1] array, with one cell per 'trial' in the experiment. Each element in this cell array is an 1xV array of observed fMRI activations. The element
data{x}(1,v) gives the fMRI observation at voxel v within trial x. The full image at time t within trial x is given by data{x}(1,:).For Peron 'P1' we have a 360x21764 shaped array
with data containing [0-1].

3 . Papers Studies

3.1 . Paper Title: Support Vector Machine for Analyzing Contributions of Brain Regions During Task-State fMRI
Authors: Mengyue Wang, Chunlin Li, Wenjing Zhang, Yonghao Wang, Yuan Feng, Ying Liang, Jing Wei, Xu Zhang, Xia Li and Renji Chen

Introduction:
In this paper, a linear support vector machine is used as a classifier to classify two types of tasks- math and story tasks. Then using svm they compared them with the activated brain regions of an SPM statistical analysis. 13 regions used for classification in SVM have activated regions out of 25 regions and 12 were non-activated regions. Right, Paracentral Lobule and right Rolandic Operculum contributed most to the classification which belongs to non-activated regions.

Dataset and Experimental Paradigms:
1046 healthy subjects were obtained from the open source database where 560 were female and 486 were male. There were two types of tasks-comprehension questions and math problems.Each participant was asked a question about a comprehension and math problem.Then data were acquired using 32 channel head coil on a modified 3T Siemens Skyra. After acquiring data, it was preprocessed using FSL and FreeSurfer.

SPM Statistical Analysis:

GLM analysis is used in order to see the difference between two tasks and to see functional activation.The result of this analysis was used to compare with the results of machine learning and to analyze the activation of brain functions

Classification Model:
The spmT file was generated for each of the two experimental conditions after the SPM processed individual data. The feature vector was: math task 1046 × 90, story task 1046 × 90 for a total of 1046 participants.The characteristics of 800 subjects were selected as a training set and 246 subjects were used as the test set.A z-score was used to normalize the training set before classification and for feature selection, the Lasso regression algorithm was used.They use a linear support vector machine with 10 fold class cross validation. The test set was used to obtain the actual classification and the accuracy of the prediction result

Result and accuracy:

For math task: Mean Reaction time and accuracy: 3.79 ± 0.38 s, 83.28% For story task: Mean Reaction time and accuracy: 3.50 ± 0.39 s and 92.57% They found that math task had a slower reaction time and lower accuracy compared to the story task.Comparing the classified brain region contribution results and the group analysis activation region results, it was found that 13 of the 25 characteristic brain regions overlapped with the group analysis activated brain regions. Among the 13 brain regions, there were 11 brain regions that overlapped with a different activation map between the math task and the story task. They show the relationship between the penalty coefficient C of the linear support vector machine and the correct rate of the prediction result under different alpha values in the following graph.When α = 0.002, C = 0.09, the highest classification accuracy rate was 87.60%.

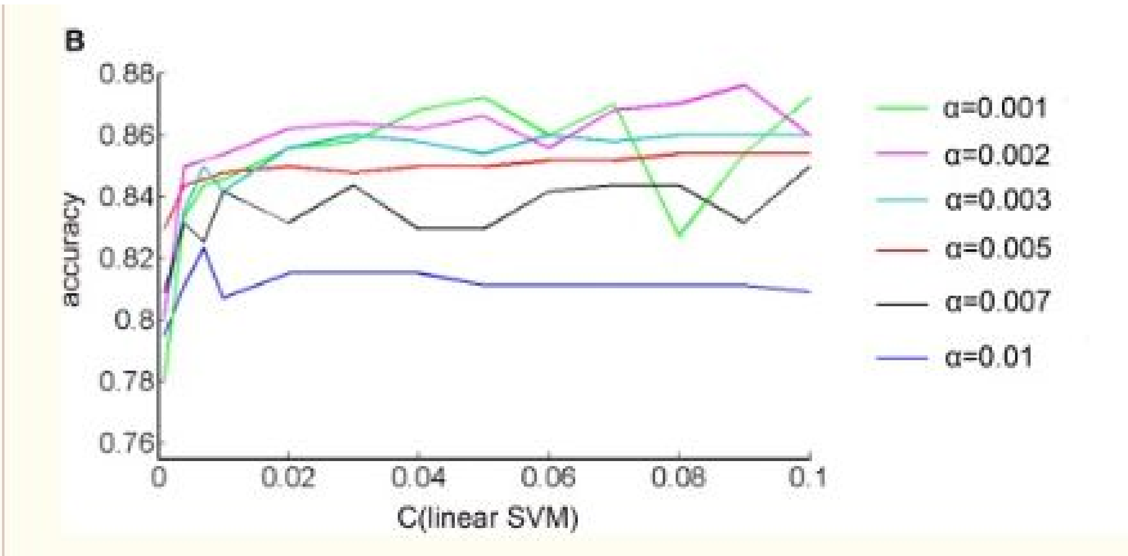


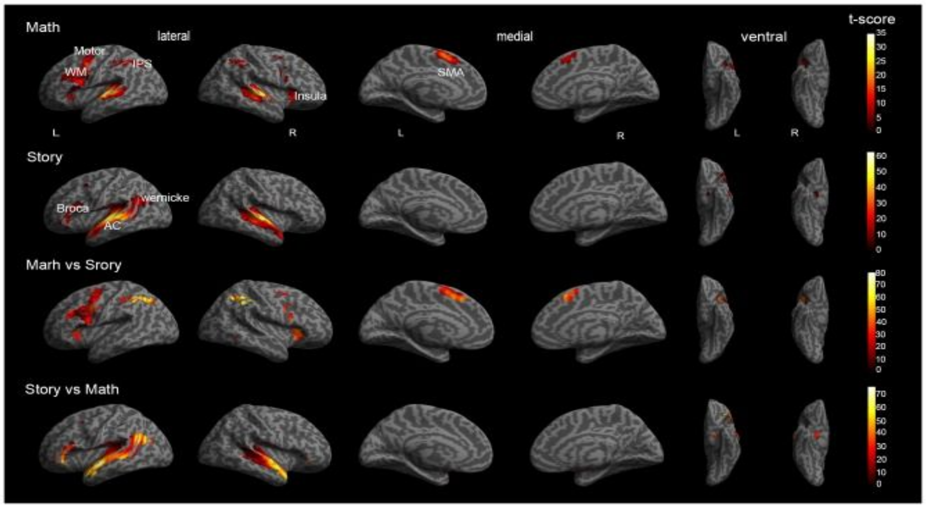
Image Analysis:

For each task activate brain regions are analyzed and found similarities and differences.

The activations of different regions due to math task, story vs math(activated region of a story which has higher activation intensity than math task), math vs story(activated region of math which have higher activation intensity than story task):

- Activation of math and story: Both the left and right temporal lobe
- Activation of math: Superior Parietal Lobule and Inferior Parietal Lobule
- Activation of story: the Parahippocampal Gyrus Amygdala on the left and right sides
- Story vs math: the left Inferior Temporal Gyrus, Superior Temporal Gyrus and the Middle Temporal Gyrus
- Math vs story: the left Inferior Frontal Gyrus, left Inferior Parietal Lobule and the left Superior Parietal Lobule

The image of these four conditions is given below.This is collected from the paper where we can see the activated areas(red/different color). This picture gives us an idea about the difference between the activation.



Contribution:

- Use different techniques to preprocess data.
- GLM analysis is used to evaluate the difference between two tasks which is an important task to understand which brain regions active for which task and the result were compared with the results of machine learning.
- For the classification tasks, the lasso regression algorithm was used for feature selection.After that svm was applied with 10 fold class validation.
- Penalty coefficient C and regularization parameter α, weights were set for correct classification.
- They also showed how regularization parameter α contributed to the accuracy in a graph.

Conclusion:

In this paper the linear support vector machine was used for classification, The Lasso regression algorithm was used for feature selection and the result was compared with the SPM group analysis activation result.The images were analyzed in order to see overlapping and non-overlapping brain regions.There were two regions-coincident brain regions and non-coincident brain regions: The activation intensity was strong for coincident brain .regions and was mostly the difference between tasks to activate the brain regions. Non-coincident brain regions are that regions that contributed to significant classification, right Paracentral Lobule, and right Rolandic Operculum. Therefore this paper includes, analyzing the difference between the two states of brain activity and finding important brain regions with no statistical difference and finding overlapping and nonoverlapping regions.This is an important analysis for negative activation of brain regions in brain research.

3.2. Paper Title: Beyond mind-reading: multi-voxel pattern analysis of fMRI data
Authors: Kenneth A. Norman, Sean M. Polyn, Greg J. Detre and James V. Haxby

Introduction: The Paper represents the technique of processing multiple voxel information in cumulative manner. Rather than individually selecting each voxel which can be inefficient process, a set of candidate voxels is selected. The characteristics in the voxel set is a way of representing metal state of the brain map onto patterns of neural activity.
The entire process of selecting a group of voxels and applying a pattern analysis algorithm for determining neurological activity is called multi-voxel pattern analysis (MVPA).Characterizing neural coding using the MVPA method and processing information in domains ranging from visual perception to memory search is the motivation of the paper fashioned work.

Feature Selection: For Feature Selection and dimensionality reduction multi-voxel pattern analysis is followed which evaluates sets of voxels, based on the strength and informativeness of patterns of activity expressed over those voxels. The difficulty which may occur in this method is exhaustive search in a large voxel space set is very time and space costly. The issue was resolved by This issue can be addressed by constraining the search to sets of spatially adjacent voxels.

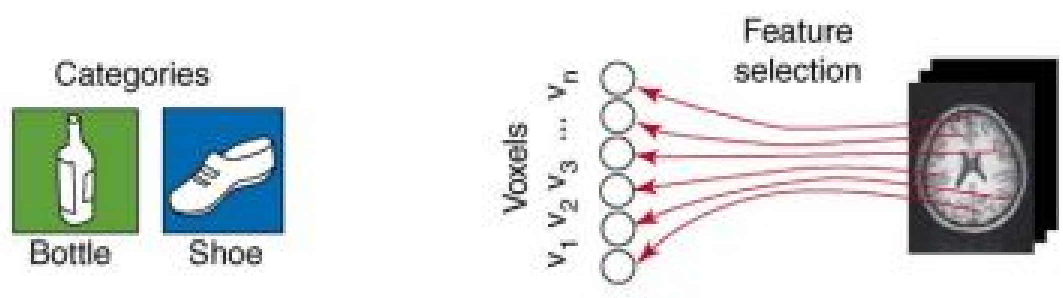


Fig 3 : Object Category and Feature Selection

Contributions :

- The achievements will be building a bridge between cognitive variables and individual brain voxels (volumetric pixels).
- Another contribution of this paper increasing sensitivity by applying MVPA methods instead of individual-voxel-based methods which was considered as the primary strategy advantage.

Questions of Interest:

- Can MVPA method be used for improving human ability to recognize cognitive states (from fMRI) in real-time? If the objective is met can this method be used to provide ‘cognitive biofeedback’ to help subjects learn
to control their thoughts or external devices?
- Can the MVPA method be used for studying how representations change as a function of learning?
- Which methods can prove most effective for mapping between brain states and points in cognitive space in a continuous fashion?

3.3 . Paper Title: Functional magnetic resonance imaging of human brain activity in a verbal fluency task.

<u>Authors:</u> R Schlösser, M Hutchinson, S JoseVer, H Rusinek, A Saarimaki, J Stevenson, S L Dewey, J D Brodie

Introduction

Magnetic resonance scanners provide functional brain mapping data with a determined sensitivity. In this paper verbal fluency was used as a task, to ascertain the consistency of fMRI and determining the anatomic substrate of language between subjects and between sexes. Verbal fluency tasks activate the dorsolateral prefrontal cortex as well as variable deactivation in temporal areas and in the posterior cingulate, which can be investigated by using functional MRI (fMRI).

Feature Selection

While doing verbal tasks like word generation, lower activation in the left frontal cortex and

higher activation in the left temporal areas have been noticed. A priori assumptions about the loci of activation is required for this test.
The first aim of the present study, therefore, was to determine the feasibility of whole-brain fMRI for language mapping on a clinical scanner.

Subjects and methods

SUBJECTS Six male and six female right handed, native English speaking normal volunteers were recruited from the student population of New York University Medical Center. The subjects were homogeneous for age 22 to 26. A 1.5 Tesla SIEMENS Magnetom (Vision) was used for imaging sequences and head coils.

- During each scan, a series of 100 sequential acquisitions were obtained.
- The whole-brain (16 slices) images were collected every five seconds.
- Each study was divided into 10 epochs of 50 seconds each, resulting in a total study length of eight minutes and 20 seconds.
- The five baseline and five activation conditions were arranged in an alternating sequence, starting with the baseline.
- During the baseline condition, subjects were asked to count forward, starting with the number 1, at a rate of about one a second.
- The data were analyzed with statistical parametric mapping (SPM).
- A global normalization of the data was rendered through an analysis of covariance.
- A delayed boxcar function modeled the expected response function of the fMRI signal over time. The condition effects were estimated according to the general linear model at each voxel. The resulting set of voxel values for each contrast constitutes a statistical parametric map of the t statistic SPM(t). The SPM(t) was transformed to the unit normal distribution (SPM{ Z}) and thresholded at Z=3.09 (uncorrected).

Contributions

- Activation in the left prefrontal and right cerebellar regions have been noticed.
- No gross differences in the pattern of activation between male and female subjects.
- Decremental responses in mesial and dorsolateral parietal cortex bilaterally during the activation task.

Questions of Interest

- determine the feasibility of whole-brain fMRI for language mapping on a clinical scanner.
- to investigate the consistency of activation in terms of cortical localization
- whether sex differences exist in the functional organization of the brain or not.

Problems Faced

The lack of direct control for task performance is a problem common to many fMRI studies, in which direct feedback could compromise the quality of the acquired data set. Task compliance in this study was confirmed by debriefing all subjects immediately after the study and by showing that the region-specific acquired signal in all subjects followed the boxcar design of the instructions.

Conclusion

Demonstrating a robust response to a verbal fluency task in each of 12 normal subjects, they showed the potential for reliable brain mapping of the functional organization of language on a conventional MRI scanner. Their study also shows that Activation in the primary visual cortex was detected only when efforts were made to constrain the head in flexion-extension, the movement of the head due to respiration.

4. Experiments Conducted

The first phase of our experiment was getting to know with the dataset and reading the reference paper. The research paper along with online resources of the paper such as review paper, dataset description, link, information about data format, etc was studied for primitive knowledge acquisition about the topic. After that, we prepared our dataset for applications such as vectorization so that learning algorithms can be performed on this data. Then we have conducted some basic statistical tests. Through the timeline, we occasionally visualized data for better understanding of the behavior of the data. For our experiment, we began the process by preprocessing the data followed by multivariate feature selection. After that, we ran some classification algorithms on 84.44 % of our data for training and the remaining 16.66 % were used for performance evaluation which is the test set. After Getting the accuracy along with a prediction vector we measured the performance in terms of different contexts.

Fig 4 : Two samples of data reading for the corresponding voxel

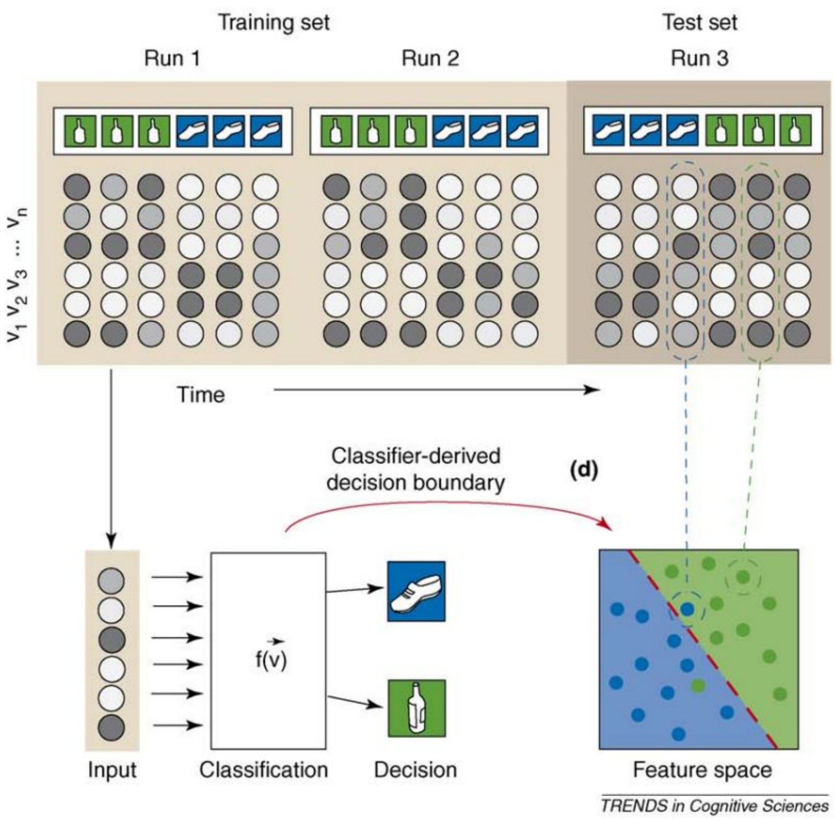


Fig 5 : Classification process followed by paper 2

5. Problems Faced

While conducting experiments on FMRI data several issues regarding voxel readings and positional information were found. First, the voxel values at a particular time frame do not remain consistent as a result of any kind of assumption for a particular word trial is will be invalid. Again pattern matching or clustering techniques sometimes clips of a significant amount of voxels that were supposed to be considered for stimuli activity observation.

6. Performance Analysis

We evaluated the performance of this algorithm in terms of Linear Support Vector Machine as it has the highest prediction accuracy. We Calculated accuracy score, balanced accuracy score, f1 score, precision score, and recall score.

Table 1 : Performance analysis for different person

	Accuracy Score	Balanced Accuracy Score	F1 Score	Precision Score	Recall Score
P1	0.36666666666	0.36666666666666	0.3467231842	0.35330687830	0.366666666
P2	0.26666666666	0.26666666666666	0.2533039329	0.30114087301	0.266666666
P3	0.18333333333	0.18333333333333	0.1655959780	0.16087962962	0.183333333
P4	0.4	0.4	0.3757820882	0.39166666666	0.4
P5	0.16666666666	0.16666666666666	0.1657647907	0.18703703703	0.166666666

7. Conclusion

FMRI data is a versatile resource for data scientists that has multiple functionalities of extracting information regarding the thinking process, stimulus activity, multiple brain organ correlation. In this project, we studied and analyzed the dataset, conducted some experiments for setting up some learning algorithms and finally benchmarked performance and accuracy. Although our implementation did not perform well as reference paper, our achievement from the project is acquired knowledge about dataset analysis and experimentation.

8. References

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6414418/?fbclid=IwAR0ipvG15RMXHZFwRUrokI6hlh3PhoJITQVzNuVk3XLK4q-B9ay7HtwRbfo

2. https://www.sciencedirect.com/science/article/pii/S1053811903000491?via%3Dihub&fbclid=IwAR37tKHZPCd1N8bzj7OtZZJ2lCeX_qVZ5ZKK2SLEyCOxu14mauxp4F0cnZs

3. https://www.verywellmind.com/the-anatomy-of-the-brain-2794895