CSE 512 Machine Learning

Learning Cognitive states using Brain fMRI

MidWay Project Report
Sriganesh, Navaneethakrishnan(109928706)
Vasudev, Balasubramaniam (109928706)

April 16, 2015

1 Introduction

Functional Magnetic Resonance Imaging (fMRI) has emerged as a powerful tool to collect vast quantities of data from the human brain. The data produced from the fMRI image is 3 dimensional. The data represents the neural activity level of different regions of the brain. This local neural activity is measured using blood oxygenated level. The 3 dimensional image of the brain is made up of voxels. The voxels represent cells or group of cells which are activated due to a cognitive response. The blood oxygenated level of each voxel is recoreded in the fMRI image. There are approximately 15000 voxels collected once per second. This yields tens and millions of observations. This project aims at using Machine Learning techniques to analyze data from the fMRI to study the cognitive state of the human brain.

2 Objective

The objective of the project is to apply machine learning techniques on the fMRI data and classify and study the different cognitive activity based on the BOLD (blood oxygenated level dependent) response of the voxels. The study involves analysis of various machine learning techniques including feature selection on the data and understand why different methods yield different results.

3 Data

The Data is obtained from http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/ . The data consists of fMRI BOLD responses from 6 different subjects. Each subject data consists of 55 trials. Each trial consists of fMRI BOLD responses of voxels. The number of voxels recorded varies based on subjects. The number of responses spans to 55 images, each taken with an interval of 55 seconds each. During the trial the subject is presented with a sentence or a picture. This stimulus is presented in two phases. The first phase occurs during the first 4 seconds followed by no stimulus during the next 4 seconds. This consists of first 16 records in each trial. The neural activity of the voxels is retained after the stimulus has been withdrawn and it lasts for 9-12 seconds. The period without the stimulus is thus considered to have similar neural activity as that of the previous stimulus. The second phase is the following 16 records consisting of neural activity of the brain for next 8 seconds. In the two phases the subject is given a stimulus of either a picture or a sentence. This provides the labelling necessary for classification

4 Feature Selection

The data processing involved segregation of the records based on the stimulus. The approach used for classifying the records is based on the neural activity of the voxel. This is represented using BOLD levels. The number of voxels thus provided the number of features. This is a very high dimensional data, consisting of 5000 voxels per record and 1500 records per subject. This varied significantly based on the subject. Feature selection made use of ROI. ROI identify specific anatomically defined Regions of Interest inside the brain based on Neural activity of the voxels. The ROI are collected by a procedure which involves segregating the voxels based on a scheme called parcellation. This involves identifying overlapping areas in the structural and functional MRI's. The ROIs are made up of a group of voxels. Feature selection was employed based on the suggestions provided on the original paper. Three different types of features selection was used.

- **Based on Average BOLD responses in ROI** Average BOLD response in ROI consists of super voxels. This level indicates the mean activation of the voxels within that ROI
- **Based on the activity level of each individual voxels**
- **Based on select ROI of voxels** Voxels consisting of certain Regions of interests.

5 Classification

The main challenges for this project include,

1. The amount of computation power required to run the base algorithm or the state of the art algorithm has to be assessed and has to be executed to give us a feel of how hard the computation is.
2. Any additional software or computing tools has to be picked out and installed properly for the set up.
3. The data set that represents a structured and real life graph has to be collected and transformed into a machine readable form for analysis
4. Once the setup is done. Different algorithms with different inputs (sampled inputs) have to be evaluated and studied for any improvement.

6 Deliverables

The deliverables for this project include,

1. Sample data on which the algorithms will be run.
2. Results of the existing algorithms on the sample data.
3. Source of enhanced fast algorithm.
4. Results of the improved algorithm on sample data.
5. Project Report.

References

[1] Vladimir Ufimtsev, Sanjukta Bhowmick *An Extremely Fast Algorithm for Identifying High Closeness Centrality Vertices in Large-Scale Networks*, 2014 4th Workshop on Irregular Applications: Architectures and Algorithms

[2] N. Halko, P.G. Martinsson, and J.A. Tropp, *Finding structure with randomness: Probabilistic algorithms for constructing Approximate matrix decompositions*, 2011, SIAM

[3] Ahmet Erdem Sarıyüce, Erik Saule, Kamer Kaya, and Ümit V. Catalyürek, *Shattering and Compressing Networks for Betweenness Centrality*, SIAM Data Mining Conference (SDM), 2013

[4] U Kang, Spiros Papadimitriou, Jimeng Sun, Hanghang Tong, *Centralities in Large Networks: Algorithms and Observations*, 2011, Citeseer