Dimensionality Reduction in the acquisition of fMRI brain graphs and its Impact on Discriminability

Eric Bridgeford (ebridge2@jhu.edu), Tanay Agarwal (tagarwa2@jhu.edu)

November 4, 2016

1 Abstract

The twenty first century has seen the rapid ascent of technology into everyday life. From high performance desktops, to smart phones, to super computers, many industries rely on the power and consistency of computing. The psychiatry field, for instance, continues to rely on psychological examinations almost exclusively when determining factors such as rate of development, mental disorder, and overall mental health in millions of patients each year. Standards of care and diagnosis vary widely between psychiatrists, and no quantitative tools exist to aid in their tasks. Machine learning has furthered their applicability, as computers can be trained to recognize nearly invisible (to the human eye) patterns, yet many medical fields continue to analyze psychiatric subjects with the human eye. Throughout the course of a scanning session, individual scans may be corrupted by a wide variety of noise variables with respect to the brain activation at a point in time; notably, scanning related factors (head motion, scanner error, etc), and processing-imposed factors (assumptions made throughout the course of analysis that add bias or variance). Through this project, the investigators will expand upon a summer project by Eric Bridgeford, the FNGS (Functional Neurodata Graphs Service) pipeline, and examine the impact of dimensionality reduction on the reliability of the graphs obtained.

2 Methods

2.1 Machine Learning Methods

1. PCA (Principal Component Analysis): can concentrate the majority of the signal of a particular scan into the first few principal components, and later components can be removed to reduce the dimensionality (and correspondingly, since the later components have less signal, they can be thought of as higher noise components)

2. CompCor (Component-based Noise Correction): major components are calculated from regions of the brain known to be higher noise (such as the cerebrospinal fluid) and removed from the scan, since these regions will contribute very little relevant signal to the timeseries

2.2 Analysis Statistics

1. Discriminability: The researchers will use a statistic developed by Wang et. al [1], the discriminability, to compare the scan-scan reliability of the fMRI graphs obtained with and without dimensionality reduction. The discriminability separates this study as largely novel; while previous investigations have studied PCA and CompCor for fMRI processing on an individual graph basis, the discriminability allows a dataset-wide comparison of the degree to which individuals scanned repeatedly best resemble themselves (essentially, fingerprinting of brain scans).

3 Resources

1. The project will add a module to the FNGS pipeline, which is a python library for fMRI processing developed by Eric Bridgeford. The pipeline is written in python, and utilizes many python libraries; particular libraries that will be explicitly necessary for this module will be numpy and scipy (acquisition of the brain graphs). Post processing will be done using R (analysis of brain graphs and the discriminability).
2. The researchers will use several existing multi scan fMRI datasets (part of the CoRR [2] collection), notably, the KKI dataset, the NKI24 dataset, the HNU1 dataset, the DC1 datset, and the SWU4 dataset.
3. Simulated data developed from the randomts R package [4].

4 Milestones

4.1 Must achieve

12/12/16: project due

12/1/16: algorithms implemented and tested with bugs removed

4.2 Expected to achieve

11/20/16: PCA implemented, may have bugs

11/27/16: CompCor implemented, may have bugs

12/5/16: general framework for paper (ordered bullet points; sections; etc), with figures generated

12/7/16: rough draft of overview completed

12/10/16: final paper completed and checked

4.3 Would like to achieve

Everything in the "expected to achieve" section shifted forward a week

5 Final Writeup

The final writeup will feature the following organization:

1. fMRI, their relevance to the psych community, what's been done so far with fMRI processing, FNGS pipeline in existing state
2. explanation of the relevance of PCA and CompCor to fMRI processing, and how they reduce noise.
3. Simulation results (data simulated using randomts package in R, and then exported to csv for use with python).
4. Real-data results (data from the CoRR dataset).

   * Results for simulated and real data will be reported as a distance matrix (shows the distance between individual graphs in the dataset), and a density plot of the intra and inter subject relationships (intra subject = same subject, different scan; inter subject = different subject). Expect to have the distances between subjects be lower (see here for an example).

6 Bibliography

https://en.wikipedia.org/wiki/Dimensionality$_r eduction Advantages_o f_d imensionality_r eduction$

https://www.ncbi.nlm.nih.gov/pubmed/17560126

https://en.wikipedia.org/wiki/Principal$_c omponent_a nalysis Dimensionality_r eduction$

link to relevant article about PCA

link to relevant article about compcor

Discriminability tutorial link

CoRR dataset link

NDMG

Randomts