DATA

KN†GHTS

Project Proposal

Sleep Deprivation Classification using BOLD fMRI Data

Alex Fung, Patrick Osborne, Tony Lee, Viswesh Krishnamurthy
8-7-2020

Project Proposal: Sleep deprivation classification using BOLD fMRI Data

ALEX FUNG, York University – School of Continuing Studies

PATRICK OSBORNE, York University – School of Continuing Studies

TONY LEE, York University – School of Continuing Studies

VISWESH KRISHNAMURTHY, York University – School of Continuing Studies

Background

The specific dataset and deliverables for this project were arrived at in consultation with David Rotenberg, Operations Director, Krembil Centre for Neuroinformatics, CAMH and Dr. Erin Dickie, Scientist and Assistant Professor, Krembil Centre for Neuroinformatics, CAMH. The goal in selecting a problem and dataset was to provide both useful academic conclusions to the CAMH scientists and to provide a challenging and educational project to the team.

Dr. Dickie studies brain connectivity in people with complex brain disorders using fMRI imaging. [1]Considering this, we have chosen a dataset that contains BOLD (blood-oxygen-level-dependent) functional Magnetic Resonance Imaging to capitalize on her area of expertise and to provide results that are useful and relevant to her studies.

The Stockholm Sleepy Brain Study open dataset was chosen because it contains real-world BOLD fMRI scan data with metadata on the level of sleep deprivation (among other factors) of each subject. Dr. Dickie expressed interest in identifying whether there are visible changes in the brain (and in the fMRI images) when a subject is sleep deprived vs not sleep deprived. We plan to use Machine Learning techniques to explore the BOLD fMRI data and to build a classifier based on the sleep deprivation metadata for each subject. The success or failure of this classifier will shed light on whether identifiable changes exist in the brain (or at least in the BOLD fMRI images) when a subject is sleep deprived. These insights can be applied to broader work at CAMH involving the correlation of sleep quality/disruption with mental illness.

The Dataset

The selected dataset for this project is "*The Stockholm Sleepy Brain Study: Effects of Sleep Deprivation on Cognitive and Emotional Processing in Young and Old*".[2] It is available from the OpenNeuro neuroinformatics database under a creative commons CCO license (public domain).

It contains BOLD fMRI imaging data for 86 subjects, with 2 sessions each. In one of the sessions the participant was sleep-deprived (only 3 hours of sleep out their regular cycle). In the other session the participant was not sleep deprived. In each session the participant was asked to perform 7 separate tasks. BOLD fMRI scans were made of each participant during each task and session.

In total, there should be 1204 BOLD fMRI scans in the dataset (86 subjects x 2 sessions x 7 tasks). Of these, 602 should represent sleep-deprived subjects and 602 should represent non-sleep deprived participants. This makes the dataset evenly balanced for our purposes.

Figures 1 and 2 show that the dataset is not uniform in shape. fMRI parameters for some tasks were different resulting in a "zoomed" image with cropped Z-axis which, from our knowledge, allows the MRI to capture more data in the more restricted region of interest. A mask will need to be applied to the whole dataset to restrict the Z-axis.
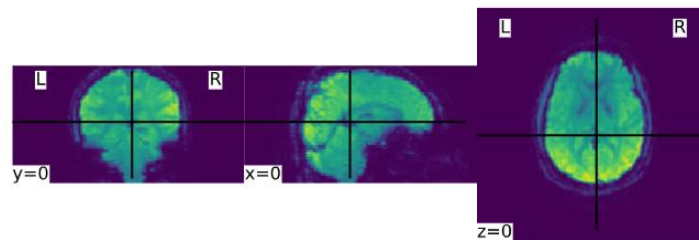

Figure 1: Orthogonal slice of BOLD fMRI image at origin (0, 0, 0)
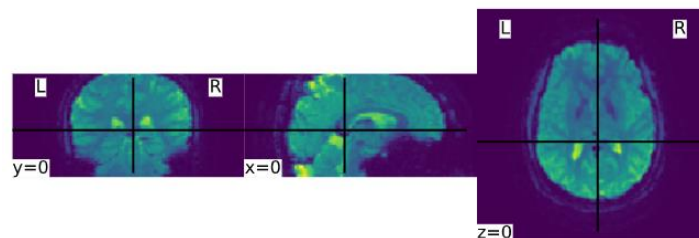

Figure 2: (Masked image, x and y axis reduced) - Orthogonal slice of BOLD fMRI image at origin (0, 0, 0)

Extensive pre-processing has been completed by Dr. Dickie to ensure that all of the fMRI scans are aligned spatially and temporally (as the subjects move their heads and have different cranial structure). Pre-processing of fMRI images is a field in and of itself in neuroimaging so it was necessary to outsource some of this work to focus on our specific deliverables. Significant data processing and feature engineering is still required on this dataset.

Problem Statement

The problem we are looking to solve centers around Dr. Dickie's goal of investigating whether there are measurable changes in the brain that correspond to sleep deprivation/sleep states and whether these changes can be identified using focused feature engineering and machine learning modelling.

The specific deliverables of the team to Dr. Dickie as are as follows:

1. Develop a binary classifier that identifies whether a subject is sleep deprived or not based on the BOLD fMRI scans.
   a. Pre-process/clean the data and engineer any necessary features to support this.
   b. Run both traditional and deep learning machine learning algorithms and choose the best classifier based on specific scoring metrics.
2. Engineer frequency-domain features via a Fourier transform on the time series of BOLD fMRI scans. Develop and assess a binary classifier using these features to determine whether certain specific frequencies (noted in the "BOLD signatures of sleep"[3] paper) correlate with different sleep states.

Scope

- Only the dataset from OpenNeuro.org, called "The Stockholm Sleepy Brain Study: Effects of Sleep Deprivation on Cognitive and Emotional Processing in Young and Old" will be used. Specifically, the dataset with OpenNeuro Accession Number: ds000201
- Data used for the project will be preprocessed at the sponsor's end. Any additional data preprocessing can mean a significant extension of the project deadlines and hence, not in scope

Risks & Constraints

- The size of the dataset is significant – 100+ GB imposing storage and compute resource constraints on local machines, requiring cloud or compute infrastructure at CAMH
- Learning curve with compute infrastructure at CAMH is steep.
- Compute scheduling constraints unknown. There may be numerous ongoing projects that vie for the same compute resources which may delay the project.
- Extensive domain knowledge required to understand and work with the dataset. Significant time may be spent discovering how to work with the data and using unfamiliar Python packages.
- Approximately 4 weeks on a part-time basis, is a very tight time constraint to implement the full machine learning workflow of data cleaning, feature extraction, feature engineering, feature selection, algorithm selection, training and evaluation of model, hyperparameter tune, and otherwise adjust model from "good enough" to "best effort".

Timeline / Milestones

Date	Milestone
8th Aug 2020	Project proposal submission
15th Aug 2020	Feature engineering (first pass) – CAMH Project progress & solution motivation – YORK SCS
29th Aug 2020	Primary goal, baseline models (first pass) – CAMH Peer review & critic responses – YORK SCS
5rd Sep 2020	Secondary goal – first pass
10th Sep 2020	Final project results
13th Sep 2020	Report out & project submission

Links

Project Repository: https://github.com/edickie/ds000201_preproc

---

[1] "Erin Dickie (She/Her) Scientist, Assistant Professor, Krembil Centre for Neuroinfomatics, Centre for Addiction and Mental Health, University of Toronto" https://ohbm.github.io/osr2020/speakers/erin_dickie.html

[2] "The Stockholm Sleepy Brain Study: Effects of Sleep Deprivation on Cognitive and Emotional Processing in Young and Old" - https://openneuro.org/datasets/ds000201/versions/1.0.3

[3] "BOLD signatures of sleep" - https://www.biorxiv.org/content/10.1101/531186v1.full