ATI Biobank Project cheatsheet

Alex Navarro

June 21, 2016

1 Data

The driving idea behind this section is to have quick access to terminology referring to the dataset as well as some additional pointers for other things related to the main document.

1.1 Main directories

- Data file: `/vols/Data/HCP/BBUK/workspace3b.mat`
- Steve's original readme: `/home/fs0/steve/BB_README`
- Data description and lookup: `/vols/Data/HCP/BBUK/SMS/old/ukb6225.html`

1.2 List of main acronyms

General nomenclature

- ID: subject identity
- IDP (Imaging-derived phenotypes): Scores derives from brain imaginging ()
- QC (Quality control): Metrics relating to the "goodness" of the data.
- dMRI (diffusion RMI): A *structural* imaging technique which shows
- rfMRI (resting-state functional MRI): *functional* or cognitive technique, fMRI imaging with no stimulus
- tfMRI (task functional MRI) : *functional* or cognitive technique, fMRI while performing some specific task (e.g. recognizing faces)

Processing nomenclature

- SIENAX (Structural Image Evaluation using Normalisation of Atrophy): Technique described in in Smith 2002. It sums up to a measure of the volume of the brain structure adjust the structural image by its atrophy
- FIRST (FMRIB's Integrated Registration and Segmentation Tool): A tool by Patenaude 2011 for segregating the brain into 15 subcortical structures

ALEX: Check this against Stephen's paper, might be reading too much into the name.

1.3 Data description

Using the UK Biobank Imaging Documentation (UKBID) —which I found and added to the group folder on Mendeley— we compared the nomenclature to the data stored within the Matlab variables and Steve's readme description.

The complete list of variables and their related codes can be found on `/vols/Data/HCP/BBUK/SMS/old/ukb6225.html`. This file can be open, e.g. using firefox remotely.

`IDPnames` variable helps to relate the variables to the terminology used in the UKBID. It divides variables in the following groups: Original numver of subjects: 5847

ALEX: Cool, but what does it look like in a 3D shape of the brain? Can we easily localize where to visualise it?

Variable type	Matlab Range	Type	Physical Meaning
ID	1	None	
QC	2:17	Usability score?	
IDP T1 SIENAX	18:28	Structural	Volumes for brain regions
IDP T1 FIRST	29:43	Structural	Volumes for brain regions
IDP SWI T2*	44:57	Structural	Microbleeding count?
IDP tfMRI	58:73	Functional	
IDP dMRI TBSS FA	74:505	Structural	
IDP dMRI ProbtrackX FA	506:748	Structural	
rfMRI amplitudes (ICA 25) node	749:769	Functional	
rfMRI amplitudes (ICA 100) node	770:824	Functional	
rfMRI connectivity (ICA 25) node	825:1034	Functional	
rfMRI connectivity (ICA 100) node	1035:2519	Functional	

Table 1: Type and range for each type of data.

2 Statistical and Machine Learning Methods

The objective of this section is to standardize the terms the group uses to avoid problems when discussing ideas and algorithms. Also, this is meant as a quick guide to the methods we will be using.

2.1 Nomenclature

- Confounding variable - a variable that when conditioned on induces independence on other variables (e.g. $p(x, y) \neq p(x)p(y)$ but $p(x, y|z) = p(x|z)p(y|z)$)

2.2 Models

- ICA (Independent Component Analysis): for a simple introduction, see [1] pages 407–416.

References

[1] K. P. Murphy, *Machine learning: a probabilistic perspective*. Cambridge, MA: The MIT Press, 3 ed., 2012.