

The provided scripts can be used to predict the neuroimaging data using behavioural data as inputs. For each given voxel, the neuroimaging data for a specific modality is predicted (using a leave-one-out process) from the full set of behavioural data. This is done by constructing a linear regression model (using support vector regression) at each voxel, where the neuroimaging data is the dependent variable and the behavioural data are independent variables.

The inputs:

- 1) MAT files that are produced by our pre-processing pipeline (one MAT file per subject);
- 2) An Excel spreadsheet that contains the behavioural data that will be used as predictors.

The outputs:

- 1) The NIFTI brain map (which can be opened as an overlay in MriCro software) with voxelwise values of prediction accuracy (Pearson's correlation coefficient between the actual and predicted values of neuroimaging data, computed across participants);
- 2) The NIFTI brain map of p values corresponding to each voxel's correlation value from the prediction accuracy map (these values can be used for thresholding);
- 3) A set of NIFTI brain maps, one for each behavioural predictor, which show the importance of this predictor for predicting the neuroimaging data. They are Z-scored beta maps, computed independently for each voxel.

The preparation

1. Locate the folder with the MAT files. Decide on the modality that you would like to analyze; make sure you have its exact name (you can get the list of modalities by running NiiStat). The MAT files that don't contain the given modality will be ignored in all subsequent analysis.
2. Locate the Excel spreadsheet with behavioural data. The file must contain a sheet with the name "to_analyze"; the other sheets will be ignored. Make sure that there are no missing values. Only numerical scores that will be

used for analysis; anything containing strings will be ignored. Numerical scores that don't vary across participants will also be ignored.

3. Locate the folders with NiiStat and SPM (we will use some of their functions). Make sure these folders are in the Matlab path; alternatively, add the paths to NiiStat and to SPM to the code of `process_chunk.m` and `combine_chunks.m` (currently they are commented out).
4. Install this repository on your computer, either by downloading the zipped file (https://github.com/grigori-yourganov/voxelwise_prediction/archive/refs/heads/main.zip) and then unpacking it, or by cloning it in the terminal ("git clone https://github.com/grigori-yourganov/voxelwise_prediction.git").

The stages of the process:

1. Extracting the relevant modality from subject-specific MAT files (this could be done on a laptop/desktop)
2. Voxelwise prediction of the neuroimaging modality from the behavioural data: this is quite slow and ideally should be done on a cluster. In order to make it more manageable, the brain is separated in eight "chunks" which are analyzed separately. After a chunk is processed, the program produces voxelwise maps of prediction accuracy, the associated p -value, and Z-maps for each behavioural score.
3. After all chunks are processed, we combine the maps from all chunks into whole-brain maps.

Detailed procedure for each stage

1. Extracting the relevant modality from subject-specific MAT files

This is a preliminary stage, its purpose is to speed up the voxelwise prediction. Here, we use the program `loaddata.m` to load the subject-specific MAT files, extract the relevant modality, and save it as a single MAT file.