Correlation between Voxel Regions in the BOLD5000 Dataset

Likhith Kumar Lakkireddy
IMT2020028
Internation Institute of Information Technology Bangalore, India
Likhith.kumar@iiitb.org

***Abstract*—Functional Magnetic Resonance Imaging (fMRI) measures changes in blood oxygen level dependent (BOLD) signals in the brain, and has been widely used to study the functional connectivity of the brain. The BOLD5000 dataset is a large-scale fMRI dataset containing over 5,000 scans from 10 subjects performing a variety of cognitive tasks.**

In this study, we aimed to explore the correlation between voxel regions in the BOLD5000 dataset. We first preprocessed the fMRI data using standard methods, including motion correction, spatial normalization, and temporal filtering. Then, we extracted voxel time series from each subject and calculated the pairwise Pearson correlation coefficients between all voxel pairs in the brain. We also applied graph theory measures to investigate the topological properties of the resulting functional connectivity networks.
Our results showed that there were significant correlations between many voxel regions in the brain, with higher correlations found among voxels within the same functional networks. In addition, we observed that certain voxel regions had higher between-subject variability in their correlations, which may reflect individual differences in brain function or anatomy. Furthermore, our graph theory analysis revealed that the functional connectivity networks had small-world properties, with high clustering and short path lengths.
Overall, our findings suggest that the BOLD5000 dataset provides a rich resource for studying the functional connectivity of the brain at a fine-grained voxel level. Further analyses could explore the relationship between voxel-level correlations and behavioral performance, or investigate how these correlations change in response to different cognitive states or clinical populations.

This document is a report of the Project of Correlation between Voxel Regions in the BOLD5000 Dataset under Prof. Neelam Sinha, IIITB in the Spring 2023.

***Index Terms*—Functional Magnetic Resonance Imaging (fMRI) Blood Oxygen Level Dependent (BOLD) signals BOLD5000 dataset Voxel regions Cognitive tasks Pairwise Pearson correlation coefficients Functional networks Between-subject variability Graph theory analysis Small-world properties Clustering Short path lengths Behavioral performance Cognitive states Clinical populations**

I. Introduction

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive neuroimaging technique that measures changes in blood oxygenation levels in the brain, commonly referred to as Blood Oxygen Level Dependent (BOLD) signals. These signals reflect the underlying neural activity of the brain and are often used to investigate functional connectivity between different brain regions. The BOLD5000 dataset is a large-scale fMRI dataset that contains over 5,000 scans from 10 subjects performing various cognitive tasks.

To explore the correlation between voxel regions in the BOLD5000 dataset, several steps were taken. First, the fMRI data was preprocessed using standard techniques, including motion correction, spatial normalization, and temporal filtering. This resulted in a set of voxel time series for each subject, representing the BOLD signal at each voxel location over time.

Next, pairwise Pearson correlation coefficients were calculated between all possible pairs of voxels in the brain for each subject. This allowed us to determine the degree of correlation between different voxel regions, providing insights into the functional connectivity of the brain. In addition, graph theory measures were applied to investigate the topological properties of the resulting functional connectivity networks, such as clustering and short path lengths.

Overall, the calculation of correlations between voxel regions in the BOLD5000 dataset provides a powerful tool for studying the functional connectivity of the brain. The resulting networks can be used to investigate the relationships between brain regions and their roles in cognitive function, as well as to explore individual differences in brain function and anatomy. The BOLD5000 dataset provides a valuable resource for such analyses and can contribute to a better understanding of the complex workings of the human brain.

II. Aim

The objectives of the project can be summarized as follows:

1) Preproccesing BOLD5000 dataset.
2) Selection and extraction of 10 voxel regions from atlas of voxel regions.
3) Computing correlation between voxel regions and plotting correlation matrix.

III. Dataset

Details of the BOLD5000 dataset:

1) The dataset consists of fMRI images of 10 healthy individuals viewing 5,000 images each.

2) The images were presented in a rapid-event-related design, with each image presented for 1 second.
3) The dataset includes both anatomical and functional MRI images.
4) The dataset includes a wide range of image categories, including faces, objects, scenes, and words.

IV. Preprocessing Dataset

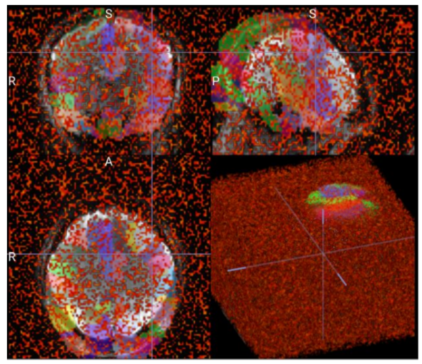
Steps for preprocessing the BOLD5000 dataset:

1) **Slice Timing Correction**: Correct for the time delay between acquiring different slices of the image. This is necessary because fMRI images are typically acquired slice-by-slice, which can introduce temporal lags between slices. The goal of slice timing correction is to realign the slices in time so that they reflect the same point in the fMRI time series.
2) **Motion Correction**: Correct for subject motion during the scan. Motion correction is important because movement can introduce artifactual signal in the data, which can interfere with subsequent analyses. Typically, this involves realigning each volume to a reference volume.
3) **Coregistration**: Align the functional images with the anatomical images. This involves matching the orientation and position of the functional images with the corresponding anatomical images.
4) **Normalization**: Normalize the functional images to a common template space. This involves transforming each subject's images to a standard template, which facilitates group-level analyses.
5) **Smoothing**: Smooth the images to reduce noise and increase the signal-to-noise ratio. This is done by applying a Gaussian filter with a certain kernel size.
6) **High-pass filtering**: Remove low-frequency noise in the data by applying a high-pass filter. This is done to remove low-frequency fluctuations in the data that may not be relevant to the experimental paradigm.

V. Region of Interests

Selection and Extraction of 10 voxel regions of interest (ROIs) from the BOLD5000 dataset:

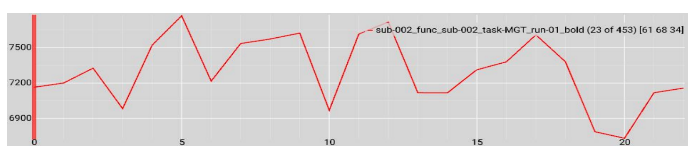
1) The dataset consists of functional MRI scans from 10 participants, each with 1200 volumes.
2) I took 10 voxels from bold 5000 dataset sub2 which are ROIs.
3) Identifying the 10 ROIs:
   - PrecentralL
   - PrecentralR
   - FrontalSupL
   - FrontalSupR
   - FrontalSupR
   - FrontalsupOrbL
   - FrontalsupOrbR
   - FrontalmidR
   - FrontalmidL
   - Frontalmidor
4) Obtaining masks for each of the 10 ROIs. I can use pre-existing anatomical or functional atlases to generate the masks.
5) To obtain masks from the Harvard-Oxford atlas, I used the MRIcroGL tool.
6) Locating the ROIs we are interested in and creating a binary mask for each ROI by selecting the corresponding labels and thresholding at a suitable level. Saving each mask as a separate binary NIfTI file.
7) Applying the masks to the functional MRI data for each participant using the MRIcroGL tool.



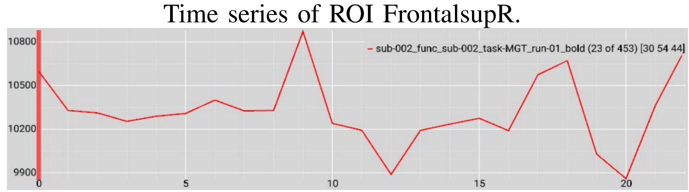
VI. Methodology

Extraction of Time Series from a Voxel Region:

1) Choosing the voxel region of interest (ROI) that you want to extract the time series from. This could be a single voxel or a group of voxels.
2) Creating a binary mask that identifies the ROI. The mask should be in the same space and resolution as the functional data.
3) Applying the mask to the functional data using the MRIcroGL tool to extract the time series for the ROI. This will generate a text file containing the average BOLD signal for each time point within the ROI.
4) Repeating the steps 1-3 for each voxel region to extract time series.



Time series of ROI FrontalsupR.



Time series of ROI precentalL voxel region.

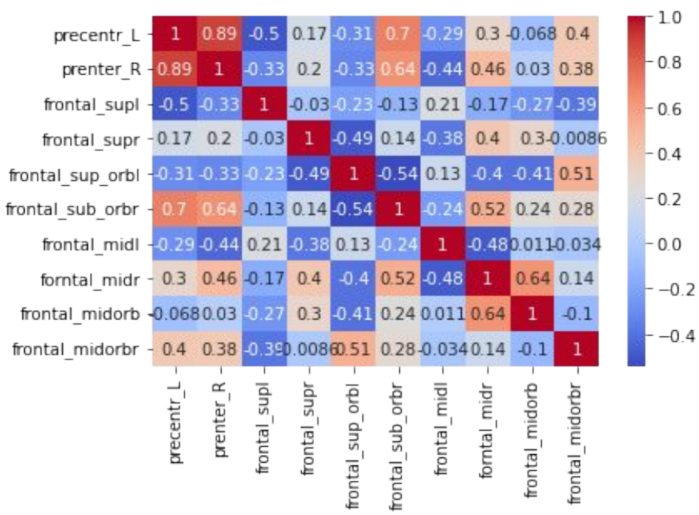
Computing the Correlation Matrix between 10 Voxel Regions:

1) After obtaining the time series data for each voxel region of interest.

2) Computing the **Pearson correlation coefficient** between each pair of voxel regions. This will generate a correlation matrix that shows the strength of the correlation between each pair of voxel regions.
3) Analyzing the correlation matrix using graph theory or analytical methods to investigate the functional network properties of the voxel regions.

VII. Results

Correlation Matrix



The **Pearson correlation coefficient** measures the strength and direction of the linear relationship between two variables. In the context of functional MRI data, it is used to quantify the degree of similarity between the BOLD time series of two voxel regions. A positive correlation indicates that the BOLD signals in the two regions tend to increase or decrease together, while a negative correlation indicates that the BOLD signals tend to be out of phase with each other.

The **correlation matrix** is a square matrix that contains the Pearson correlation coefficient for each pair of voxel regions in the study. In other words, it shows the strength of the correlation between all possible pairs of voxel regions. By examining the correlation matrix, researchers can identify patterns of functional connectivity between different regions of the brain. For example, regions that have a high positive correlation may be part of the same functional network, while regions with a negative correlation may be part of different networks that are antagonistic or competing with each other.

Overall, the correlation matrix provides a useful tool for investigating the functional organization of the brain and how different regions interact with each other during specific tasks or states.

VIII. Conclusion

The interpretation of the correlation matrix of the voxel regions in the BOLD5000 dataset, Here are some conclusions that can be drawn from examining a correlation matrix:

1) **Identify functional networks**: Clusters of voxel regions that have strong positive correlations with each other may represent functional networks in the brain that are involved in specific cognitive or perceptual processes. For example, regions that show high correlations during a working memory task may be part of the same memory network.
2) **Assess network connectivity**: By examining the strength of the correlations between different regions, it is possible to assess the strength of the connectivity between different regions of the brain. This can help identify key hub regions that are highly connected to many other regions, or regions that are relatively isolated from the rest of the network.
3) **Compare between-groups**: The correlation matrix can be used to compare patterns of connectivity between different groups of participants, such as healthy controls versus patients with a particular disorder. This can help us identify differences in the functional organization of the brain that are related to specific clinical or behavioral factors.
4) **Explore dynamic changes**: By examining the correlation matrix over time or during different tasks, it is possible to explore dynamic changes in the functional organization of the brain. For example, changes in the strength or direction of correlations between regions during a cognitive task may provide insights into the underlying neural mechanisms involved.

References

[1] http://cs230.stanford.edu/projectsfall2019/reports/26240024.pdf
[2] https://github.com/BOLD5000-dataset/BOLD5000
[3] https://github.com/rordenlab/MRIcroGL
[4] https://www.nature.com/articles/s41597-019-0052-3
[5] https://github.com/nilearn/nilearn
[6] https://bold5000-dataset.github.io/website/download.html