Mind the Map: Exploring Task Similarities and Differences in the IBC Dataset

Anusha Nath Roy 2020101124
Freyam Mehta 2020101114
Srijan Chakraborty 2020115001

Abstract

The primary objective of this project is to conduct a comprehensive analysis of the IBC dataset, a valuable resource for researchers in the neuroscience domain to exploit. Through the analysis, the aim is to uncover hidden patterns and insights within the data that may contribute to our understanding of the complex workings of the human brain. The focus of the study is centered around a cross-comparison of various tasks recorded in IBC release 2, with the purpose of identifying similarities and differences across tasks.

To achieve this, we used the z-maps of individuals as a powerful tool to uncover these patterns. By conducting this study, we hope to gain valuable insights into the hierarchical organization of brain areas and their functioning. The potential impact of this study extends beyond the academic community and could contribute to the development of improved cognitive and behavioral therapies for individuals affected by neurological disorders.

1 Introduction

Cognitive neuroscience is a burgeoning field that is concerned with unraveling the complex neural mechanisms underlying the cognitive processes that underpin our everyday lives. The study of cognitive neuroscience is essential for understanding the nature of consciousness and behavior, and it has become an area of intense research focus in recent years. At its core, cognitive neuroscience seeks to elucidate the neural representations that give rise to cognitive functions such as perception, decision-making, and action.

The IBC dataset provides a valuable resource for cognitive neuroscientists seeking to gain insights into the organization of cognitive processes in the brain. The dataset contains a rich array of data, including fMRI images, behavioral measures, and task descriptions, from hundreds of individuals performing a range of cognitive tasks. By analyzing this data using a range of data analysis techniques, cognitive neuroscientists can uncover valuable insights into the neural basis of cognitive processes.

To analyze the IBC dataset, this project employs a range of data analysis techniques, including pattern analysis, hierarchical cluster analysis, dimensionality reduction, encoding and decoding, and statistical modeling. These techniques enable the research team to identify patterns of neural activity that are associated with particular cognitive processes, cluster similar cognitive tasks, reduce the dimensionality of the data, identify the neural code that underlies cognitive processes, and build statistical models that capture the relationships between the data.

Through the analysis of the IBC dataset using these data analysis techniques, this study aims to provide a deeper understanding of the mechanisms underlying cognitive processes. Finally, this project aims to provide a more comprehensive understanding of the neural basis of cognitive processes, which has the potential to revolutionize our understanding of the mind and behavior.

2 Review: What are some papers we have looked at?

2.1 Quantitative models reveal the organization of diverse cognitive functions in the brain

The first paper by Nakai and Nishimoto presents a quantitative modeling approach to understand the organization of diverse cognitive functions in the brain [1]. The authors used a large-scale dataset to develop a computational model that describes how different cognitive functions are organized in the brain, using different visualisations using regression weights.

Our project aims to extend this work by employing a range of data analysis techniques, such as pattern analysis, hierarchical cluster analysis, and statistical modeling, to explore the IBC dataset further. By analyzing the z-maps of individuals, we aim to identify similarities and differences across various tasks recorded in IBC release 2. This will allow us to gain insights into the organization of cognitive processes in the brain, and how different tasks are related to each other.

2.2 Individual Brain Charting dataset extension, second release of high-resolution fMRI data for cognitive mapping

The second paper by Pinho et al. describes the second release of the IBC dataset, which includes high-resolution fMRI data for cognitive mapping [2]. The dataset includes data from hundreds of individuals performing a range of cognitive tasks, making it an invaluable resource for cognitive neuroscience research. The is directly related to our project, as it describes the second release of the IBC dataset that you are using for your analysis. The paper presents detailed information about the data acquisition process, which is important for understanding the characteristics of the dataset. The high-resolution fMRI data from the IBC dataset can be used in our project for conducting pattern analysis, hierarchical cluster analysis, dimensionality reduction, encoding and decoding, and statistical modeling. Furthermore, the extension of the IBC dataset provides an opportunity to replicate and validate previous findings, as well as to explore new research questions in cognitive neuroscience. .

3 Methodology and Experiments

Our methodology involves using various data analysis techniques to gain insights into the organization of cognitive processes in the brain using the IBC dataset. Using the z-maps which have been publicly uploaded to the neurovault website, IBC Dataset Extension, we have used several different methods for data analysis and visualisation. These techniques include pattern analysis, hierarchical cluster analysis, dimensionality reduction, encoding and decoding, and statistical modeling. By applying these techniques, we aim to uncover patterns and relationships within the data that can provide valuable insights into the functioning of the human brain. We have run these visualisation and analysis techniques on 500 z-maps, collected for one user, over different sessions (00, 04 and 07) and for several different tasks. We have 100 labels for our z-maps.

1. **Data Acquisition and Preprocessing**: The first step in our methodology involves acquiring and preprocessing the IBC dataset, which is available on OpenNeuro. We also download z-maps available on NeuroVault and perform preliminary preprocessing/labelling.
2. **Pattern Analysis**: In this step, we generate mean z-maps for each task and plot the Representational Similarity Analysis (RSA). We also apply Dictionary Learning and generate RSA using the sparse codes. Furthermore, we visualize the spatial distribution of the dictionary components across the brain.
3. **Hierarchical Cluster Analysis**: We used dendrograms to visualize the clustering results and to identify groups of tasks that show similar neural patterns. We applied hierarchical clustering to both the RSA maps created from the z-map voxel activation data and the model weights used during decoding. This helped us to explore the clustering results using different perspectives.
4. **Dimensionality Reduction**: We perform dimensionality reduction to 2D and 3D using both Principal Component Analysis (PCA) and Multidimensional Scaling (MDS). This is carried out

using the mean z-maps for each label. We then analyze it by running k-means cluster with different k values to check which tasks group together in this reduced dimensional space.

5. **Decoding and Prediction**: To gain a better understanding of the neural activity patterns associated with each task, decoding models were created using the mean z-maps for each task. These models were trained to predict the task label based on the voxel activation patterns. To obtain more detailed information about the neural representation of each task, separate decoding models were also trained for each individual task. The coefficients of the models were recorded for each label, which provide a measure of the importance of each voxel for predicting the task label.

6. **Statistical Modeling**: We used regression analysis to predict the tasks based on the brain activity patterns. Specifically, we created models to predict each task individually and recorded the coefficients of the model for each label. We then used these coefficients to create an RSA matrix, which allowed us to explore the similarity or dissimilarity between the tasks based on their coefficients. Finally, we applied hierarchical clustering to the RSA matrix to visualize the clustering results and to identify groups of tasks that show similar patterns of coefficients. This approach helped us to identify the most important brain regions and the patterns of brain activity that are most informative in predicting the cognitive tasks being performed.

4 Results and Discussions

We obtained some initial findings and observations. We use these results as our baselines for further analyses and comparisons.

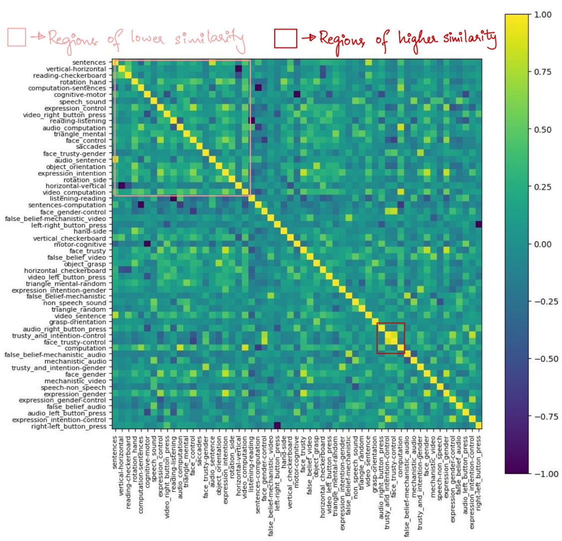

Figure 1: RSA Matrix for Baseline

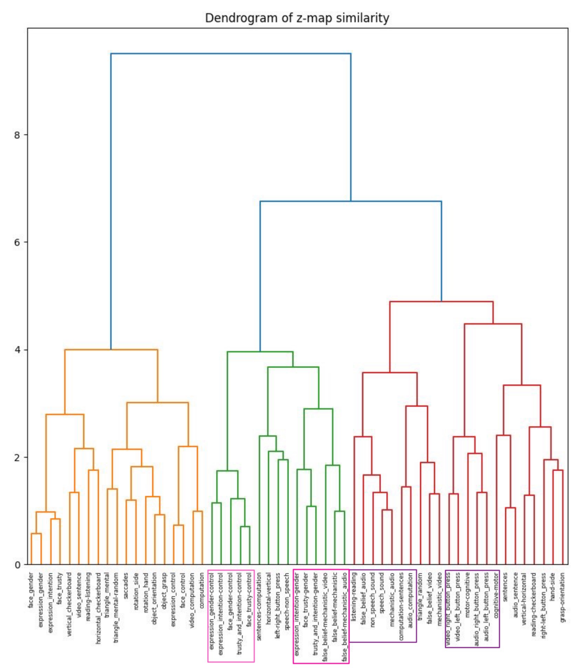

Figure 2: Dendogram for hierarchical clustering for Baseline

As we can see, the baselines do not have a lot of similarity, with a very small region (as indicated by the red box in the RSA matrix) and a large region with relatively lesser similarity (as indicated by the pink box in the top left side of the RSA matrix). Even the dendogram visualising hierarchical clustering are not very informative, with only few regions of high functional similarity. This goes on to show that while the baselines used in the analysis provide a preliminary understanding of the functional similarity between neural representations of stimuli, it is important to note that they may not provide a complete picture. As noted earlier, the RSA matrix and dendrogram provide limited information about the functional similarity between the baselines being compared.

Therefore, it may be necessary to use more powerful tools and methods to further analyze the neural representations of stimuli and better understand the differences and similarities between the baselines. For our project, we use dictionary learning and decoding models that can be used to identify specific patterns of neural activity that are associated with specific stimuli or tasks.

Using these more powerful tools can help to uncover more nuanced similarities and differences between neural representations of stimuli, allowing for a deeper understanding of the underlying cognitive processes involved. Therefore, while baselines can provide a useful starting point, more sophisticated analyses are often needed to fully elucidate the neural mechanisms underlying perception and cognition.

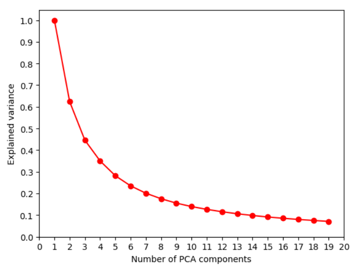

Figure 3: Variance versus PCA Curve

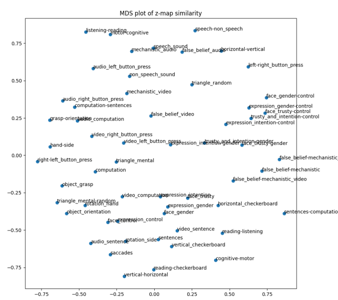

Figure 4: MDS

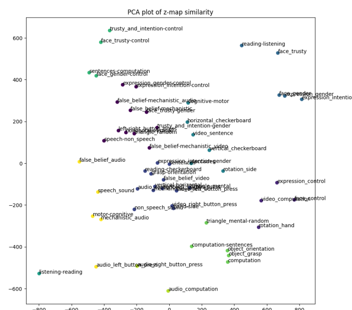

Figure 5: PCA

We have also used dimensionality reduction techniques to visualise the high dimension data to a more understandable 2-D or 3-D space. PCA allows us to identify the principal components that capture the most significant sources of variation in the data. These components represent the combinations of brain activity patterns that explain the majority of the variance in our dataset. MDS is particularly useful for visualizing and clustering brain activity data, as it enables us to visualize the data in two or three dimensions, making it easier to identify clusters or patterns of similarity between tasks or conditions. We see that these results aren't very helpful and attribute this to the fact that the brain activity is highly similar across the tasks.

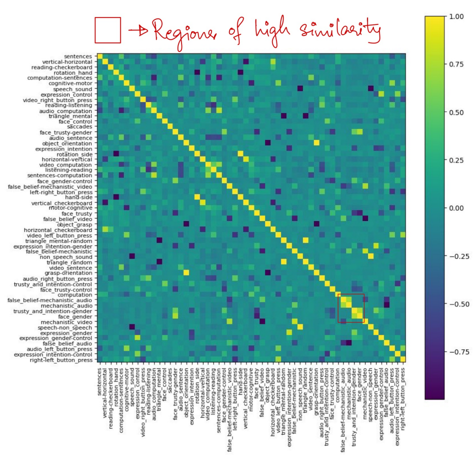

Figure 6: RSA Matrix for Statistical Modeling

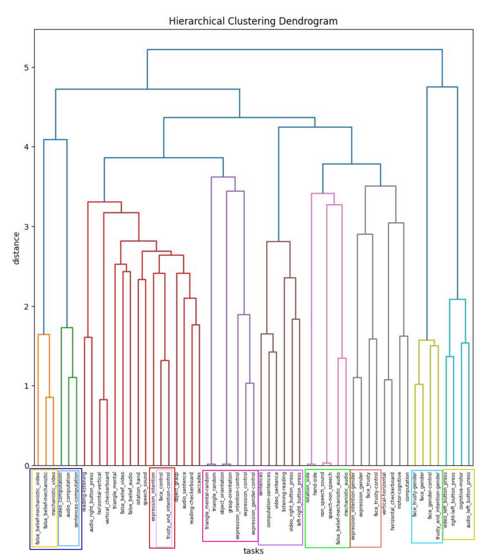

Figure 7: Dendogram for hierarchical clustering for Statistical Modeling

The use of regression weights in the analysis has significantly enhanced the results, particularly in relation to the RSA (Representational Similarity Analysis) matrix. The RSA matrix provides a measure of similarity between different tasks or conditions based on neural activation patterns. By incorporating the regression weights that were used to train the decoding model, the RSA matrix becomes more informative and allows for a more comprehensive comparison of tasks.

One notable improvement is the increased number of higher similarity regions observed in the RSA matrix. This indicates that tasks that are truly similar exhibit stronger neural activation patterns in common regions, while tasks that are dissimilar show distinct activation patterns. With the regression weights taken into account, the RSA matrix provides a more nuanced understanding of how different tasks can be compared and differentiated.

Additionally, the use of regression weights has had a significant impact on the dendrogram, which represents the hierarchical clustering of the tasks based on their similarity. Prior to incorporating the regression weights, the dendrogram may have been less clearly defined, with clusters potentially appearing less distinct. However, after incorporating the weights, the dendrogram becomes more refined, and the clusters are better defined.

It is definitely interesting to see that some tasks have been grouped together in the hierarchical clustering. Similar tasks have been clustered together in a box, and similar clusters are further encompassed by a bigger box in each of the dendogram visualisations. We see that most of the audio and sound related tasks have been clustered close together (indicated by the green box), similar to the expression and face-control or the motor tasks. These indicate that clustering based on the regression weights is outperforming the others in terms of identifying similar groups, and is a very insightful method for understanding brain functions.

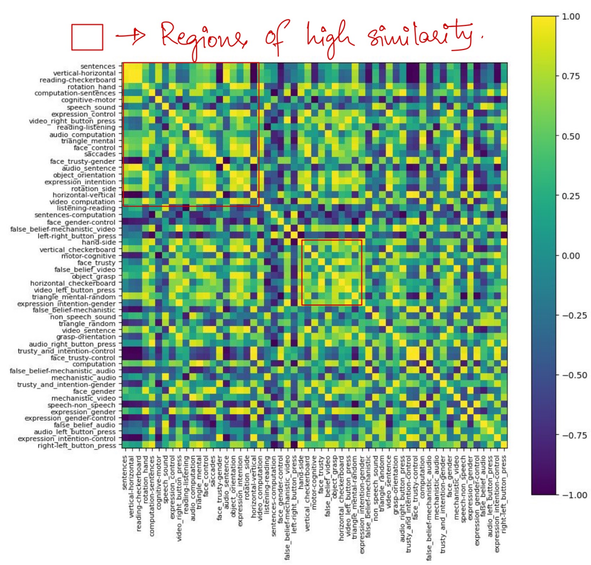

Figure 8: RSA Matrix for Dictionary Learning

The application of the dictionary learning method to our model yields remarkable improvements in mapping higher similarity among brain functions. This indicates that there are specific components or patterns of neural activity that are consistently employed by the brain across different tasks. The dictionary learning method allows us to identify and extract these shared components, shedding light on the fundamental computations carried out by the human brain.

In technical terms, dictionary learning is a technique used to decompose complex data, such as neural activation patterns, into a set of basis functions or components. These basis functions form a dictionary that represents the underlying structure of the data. By applying dictionary learning to our model, we are able to identify the most relevant components that are common across different tasks.

The remarkable improvement in mapping higher similarity suggests that the four most frequently utilized components by the brain exhibit a high degree of similarity. This finding implies that these components are essential for various cognitive processes and are consistently engaged regardless of the specific task being performed. We can identify the shared components that play a crucial role in multiple tasks, as well as those that are task-specific. This analysis provides insights into the fundamental building blocks of cognitive processes and helps us understand the neural mechanisms involved in various tasks.

5 Future Work

There are several potential directions for future work on a project that involves analyzing brain activity using techniques like RSA, regression weights, and dictionary learning. Here are a few possible avenues for further investigation:

1. **Refining and expanding the analysis:** We can further refine the analysis pipeline by incorporating additional advanced techniques and algorithms. Furthermore, we can explore alternative methods for RSA or regression analysis that may provide even more precise measures of similarity or improved differentiation among tasks. Additionally, we can consider expanding the analysis to include more brain regions or larger datasets to obtain a more comprehensive understanding of neural representations.

2. **Validation and interpretation:** We can validate the findings and interpretations obtained from the analysis using independent datasets or alternative experimental designs. Also, we can consider conducting behavioral experiments or psychological studies to validate the identified neural representations and assess their functional significance.

6 References

1. Nakai, T., Nishimoto, S. Quantitative models reveal the organization of diverse cognitive functions in the brain. Nat Commun 11, 1142 (2020). https://doi.org/10.1038/s41467-020-14913-w

2. Pinho, Ana Amadon, Alexis Gauthier, Baptiste Clairis, Nicolas Knops, André Genon, Sarah Dohmatob, Elvis Torre Tresols, Juan Ginisty, Chantal Becuwe-Desmidt, Séverine Roger, Séverine Lecomte, Yann Berland, Valérie Laurier, Laurence Joly-Testault, Véronique Médiouni-Cloarec, Gaëlle Doublé, Christine Martins, Bernadette Salmon, Eric Thirion, Bertrand. (2020). Individual Brain Charting dataset extension, second release of high-resolution fMRI data for cognitive mapping. Scientific Data. 7. 10.1038/s41597-020-00670-4.