Revealing the functional specialization of the brain using classification

by Alejandro de la Vega

Introduction

A primary goal of cognitive neuroscience is to map spatially distinct regions of the brain to the cognitive functions they perform. However, the brain is a massively complex organism, and despite have gained a broad perspective on the functional organization of the brain, the functional specialization of subregions is still hotly debated. A primary way to test models of functional organization has been to test and compare various theoretical models. However, the sheer complexity of the brain makes this approach difficult, as there may be various idiosyncratic ways the brain is organized that are not obvious a priori. Moreover, progress has been hampered by low replicability of single fMRI studies. While single studies certainly provide useful signal and insight towards brain organization, low power limits conclusions that can be made.

However, the recent development of large scale databases of neuroimaging studies, such as Neurosynth, have enabled a new breed of data driven approaches that promise to reveal novel insights into the brain. These databases provide the necessary power to apply exploratory methods and provide the ability to estimate the robustness of such findings. In addition, these databases contain a large variety of cognitive processes, bypassing the manual selection of studies performed in normal meta-analyses. This increases our ability to find novel associations in these data that may have been otherwise excluded.

These databases have already been used to generate statistically robust, data-driven characterizations of brain function and diversity. Primarily, they have been used to generate automated meta-analyses of brain function. For example, Neurosynth can be used to generate maps that show which regions are likely to be activated during a certain types of cognitive functions (e.g. reward, Figure 1). While these approaches are very useful for what brain regions are activated by specific cognitive processes, they are not designed to infer what processes differentiate regions of the brain. For example, if a meta-analysis reveals that episodic memory activates the medial

prefrontal cortex (mPFC) while semantic memory does not, one may conclude that the functional specialization of mPFC is episodic memory. However, this analysis has failed to take into account all other processes that also activate mPFC, such as self-referential processing, that may lead to different conclusions about it's functional specialization.

Thus, to determine spatial functional organization, it may be more useful to query neuroimaging databases to determine which functions, across all domains, are most associated with a given region. However, a limitation of typical correlation approaches to this question is that while certain cognitive process may be highly associated with a given region, these processes may not be unique to this region and thus do not differentiate it from others. For example, while mPFC is very high associated with reward processing, the is true of the nucleus accumbens. Thus, a different cognitive function, such as context integration, may be required to differentiate these two regions. Importantly, the function that differentiates regions may not be the most associated with any given region, making it difficult to find. This is particularly important because a key aspect of functional specialization is not only what a region is commonly involved in, but what processes it is *uniquely* involved in that allow one to differentiate it from others.

Here we applied a data-driven approach to the Neurosynth data in order to determine which cognitive functions are not only associated with different brain regions, but which functions allow one to reliably differentiate them from one another. First we clustered the brain into spatially distinct regions by applying supervised clustering algorithms to the Neurosynth database. Next, we trained classification algorithms to reliably classify different brain regions using Neurosynth terms and determined which cognitive terms were most important for classification. The importance of a feature for performing was used a measure of region's functional specialization. Finally, we used Shannon's entropy measures to quantify the relative diversity of a brain region's function. Presumably, regions with a greater range of cognitive functions would require a greater number of terms to classify against others, while fairly specific areas would require less.

Methods

Neurosynth database

The Neurosynth database (neurosynth.org), is repository of neuroimaging studies that

is populated by scraping online neuroimaging journals (such as *Neuroimage*). The scraping algorithm finds the peak activation coordinates within the paper and stores that along with the text of the paper. The text is first filtered to remove useless words (e.g. “the”) and stored as a table of word frequencies.

The resulting database is noisier than manually crafted repositories but has the advantage of including nearly 10,000 studied due to the automated nature of the algorithm. Meta-analyses performed on Neurosynth data have corroborated more careful manually performed analyses with less noisy but much smaller datasets.

Binary classification problem

The basic unit of analyses was to attempt to classify two brain regions, (e.g. visual and auditory cortex). Since the unit of analysis in the Neurosynth database is studies, we selected sets of papers that activate our regions of interest above a threshold (0.05% of voxels in ROI activated in paper). Importantly, studies that activated both regions were removed.

Once having obtained two sets of studies that activate each region, we used machine-learning to attempt to classify the studies into the regions they activate only using the word frequencies for each studies. Using 4-fold cross validation, the algorithms were given 3/4th of the data for training and were tested on the remaining 1/4 of the data. This was done four times in order to train and test on all of the available data. If the algorithm was able to generalize from the training data and classify studies using the words frequencies from the studies’ papers, it follows that the activity of the two regions can be differentiated based on the words used in the papers. For example, if one was attempt to classify visual cortex versus motor cortex, the algorithm may learn that if a study mentions the word “vision” often it is a study that activated the visual cortex and not the motor cortex.

Classification algorithms

The specific classification algorithm that was used was an ensemble Gradient Boosted Random Forest. Random forests of trees were generated and added to the ensemble classifier based on a gradient function in order to only add useful classifiers. This classifier was chosen because it results in state-of-the-art classification algorithms and describes which features were important for classification for any given pair.

Feature importances

An important aspect of this analysis is not only to classify regions, but to determine *which* features are important for classification. The classification algorithms also produced a vector which describes the importance of each term in Neurosynth for the paired classification. Again, when classifying visual cortex against motor cortex, the terms “vision” and “motor” would be expected to have high feature importances as articles that activate these regions are likely to talk about those terms often.

Neurosynth topics

Because single word frequencies can be very noisy and include cognitively irrelevant signal, we instead classified on a set of topics derived using topic modeling (Poldrack et al., 2013). The topics are a reduction on the space of Neurosynth words and group together words that are semantically related. For example, an example topic that reflects reward processing may show high loadings for the words “outcome”, “anticipation”, “loss”, and “gain”.

The grouping of words made it easier to filter cognitively irrelevant features such as a topic reflecting methodological details of papers (e.g. “study”, “analysis”, “revealed”, “compared”). Out of the original 40 topics, we removed 15 cognitively irrelevant topics, resulting in 25 topics used in classification.

Application to whole-brain

The above classification problem depicts how we classify any two brain regions against each other using Neurosynth features. We applied this approach on a whole-brain basis by dividing the brain into many small regions and applying binary classification between all the possible permutations.

For example, if 30 brain regions were entered into the analysis, each region would be classified against the other 29 regions individually. We would then know how accuracy we were able to classify any given region against any other, and which words were important for the classification. Averaging these metrics within a region would reveal which words were important for that region overall, and which features were important for classification.

Brain parcellation

In order to ensure that the regions that we entered into the analysis would be sensible

given the Neurosynth data, we parcellated the brain into 11, 20 and 60 regions using the Neurosynth data. We applied k-means clustering, a form of supervised learning, to the Neurosynth database.

Classification Metrics

The following metrics resulted from our classification analysis.

Above chance classification (ACC) accuracy

The classification accuracy that our algorithms were able to perform binary classification between any two regions, minus the classification algorithm that a “dummy” classifier would produce. A “dummy” classifier simply used frequency as the classification metric (e.g. the most frequent region was always chosen). By subtracting the dummy classification accuracy we were able to control for unbalanced classification problems.

ACC was calculated for all pair-wise binary comparisons between regions, averaged within regions to determine how well a specific region was able to be classified against all others, and averaged across all regions to determine how well regions were able to be classified across the brain.

Top features for each region

Features (e.g. terms or topics) that were most important in classifying a given region versus all other regions. This was determined by averaging the feature importances for all the classification pairs within a region. Because features varied in their overall importance across the entire brain, we also normalized each region’s feature importances with respect to the entire brain.

Diversity of functional specialization

In addition to determining the functional specialization of brain regions, we wanted to assess the extent to which a region has a high level of specialization or exhibits a broad range of cognitive function. In order to do so, we calculated the diversity of the feature importances for each region using Shannon’s Diversity. A high level of diversity reflects that many Neurosynth features were required to classify a region against others, possibly reflecting that region is involved in various types of cognitive processes. Alternatively a low level of diversity indicates that a region is involved only with a small set of cognitive functions and has a high level functional specialization.

We applied this metric across two dimensions of the resulting feature importances for each region resulting in two metrics:

Consistency is the extent to which features were used for classification consistently within a region across all other regions. For example, if the term “vision” was always important for classifying the visual cortex against all regions, that would result in a high consistency rating. Alternatively, if one region required a set of terms (e.g. “emotion”) when classifying against one region but an entirely different set of terms against another (e.g. “memory”) , that would yield a low consistency metric.

Sparsity is the extent to which feature importances were concentrated on a small set of features within a comparison. If classifying two regions requires many features to be performed, that would result in low sparsity. However, if only a few features were required for a comparison, that would result in high sparsity.