Data-driven classification of human brain function using Neurosynth

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. While most agree with this goal, the specific functional specialization of most areas in the brain is hotly debated. The advent of neuroimaging techniques has proved useful in achieving this goal because such techniques, especially functional magnetic resonance imaging (fMRI), can record brain activity while the brain performs various functions with reasonable spatial resolution. However, the field has become increasingly aware of the limitations of fMRI and while much has been learned about the functional specialization of the brain, the added knowledge that single studies provide is increasingly limited. Here, we applied machine learning classification algorithms to a large-scale database of neuroimaging studies, Neurosynth, to develop a data-driven approach to distinguishing the functional specialization and diversity of the brain.

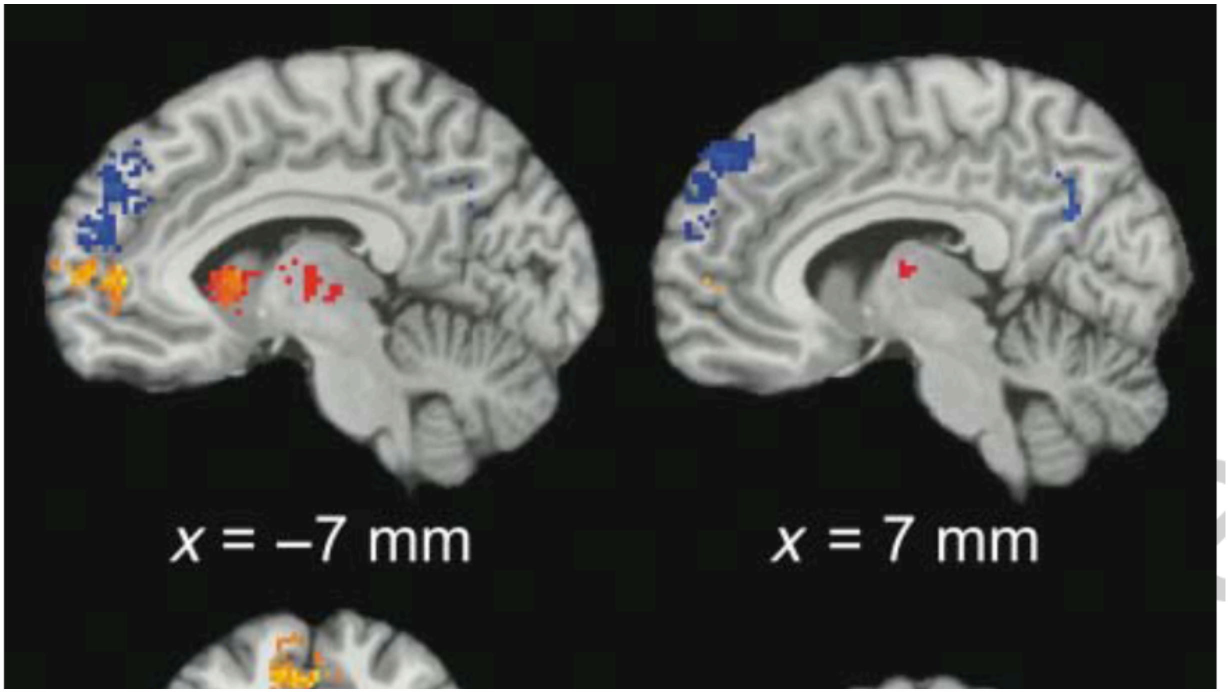
Probing functional specialization in the brain

Commonly used methods

The primary method by which cognitive neuroscientists apply fMRI to probe the functional specialization of the brain is theory-driven studies in which the mental state of various tasks are assumed to be known. That is, subjects are given tasks which are thought to be functionally distinct while their brain activity is recorded, and the resulting activity is statistically compared. Such studies have yielded an enormous literature of findings which have been incorporated into theory about functional specialization. However, it has become increasingly apparent that these studies are often severely underpowered, thereby limiting the added knowledge that can be gained from a single study. Underpowered studies result in unreliable spatial estimates of activity as well as false negatives and positives. While there is certainly valuable signal in such studies, the amount of fine-grained interpretation that can occur for a single study should be limited.

In addition, many studies fail to take into account the base-rate of activity for the region they are interested when interpreting the results of their study. While it is tempting to interpret all of regions active for your processes of interest as indicative of that process, some regions are likely to also be activate for many other processes and only some regions are unique in function to the process of interest. Thus, it is difficult to know from such studies which functions *differentiate* brain regions because some regions are be activated by common functions.

One approach that has been employed to mitigate some of the limitations of single-studies is meta-analysis. Meta-analyses combine a number of studies into a large statistical analysis in order to obtain more reliable estimates of function. Typically, if one is interested in gaining insights into the function of an area of the brain, one would gather many studies which are purported to involve this area but are hypothesized to differ slightly in their spatial organization within it. Thus, for example, if you hypothesized that mPFC is involved in mentalizing but would like to know if there is any spatial specialization for mentalizing about the self versus others, you would collect as many fMRI studies as you could find that involve both processes for a meta-analysis. Crucially, this processes would involve deciding which activations within the studies are representative of which process in order to label them for the analyses. The activations from each function class are then entered into an analysis which essentially results in an average activation map that is statistically thresholded. This analysis in fact has been done previously and reveals an overlap in a central area of mPFC with functional specialization in other subregions (Figure 1).



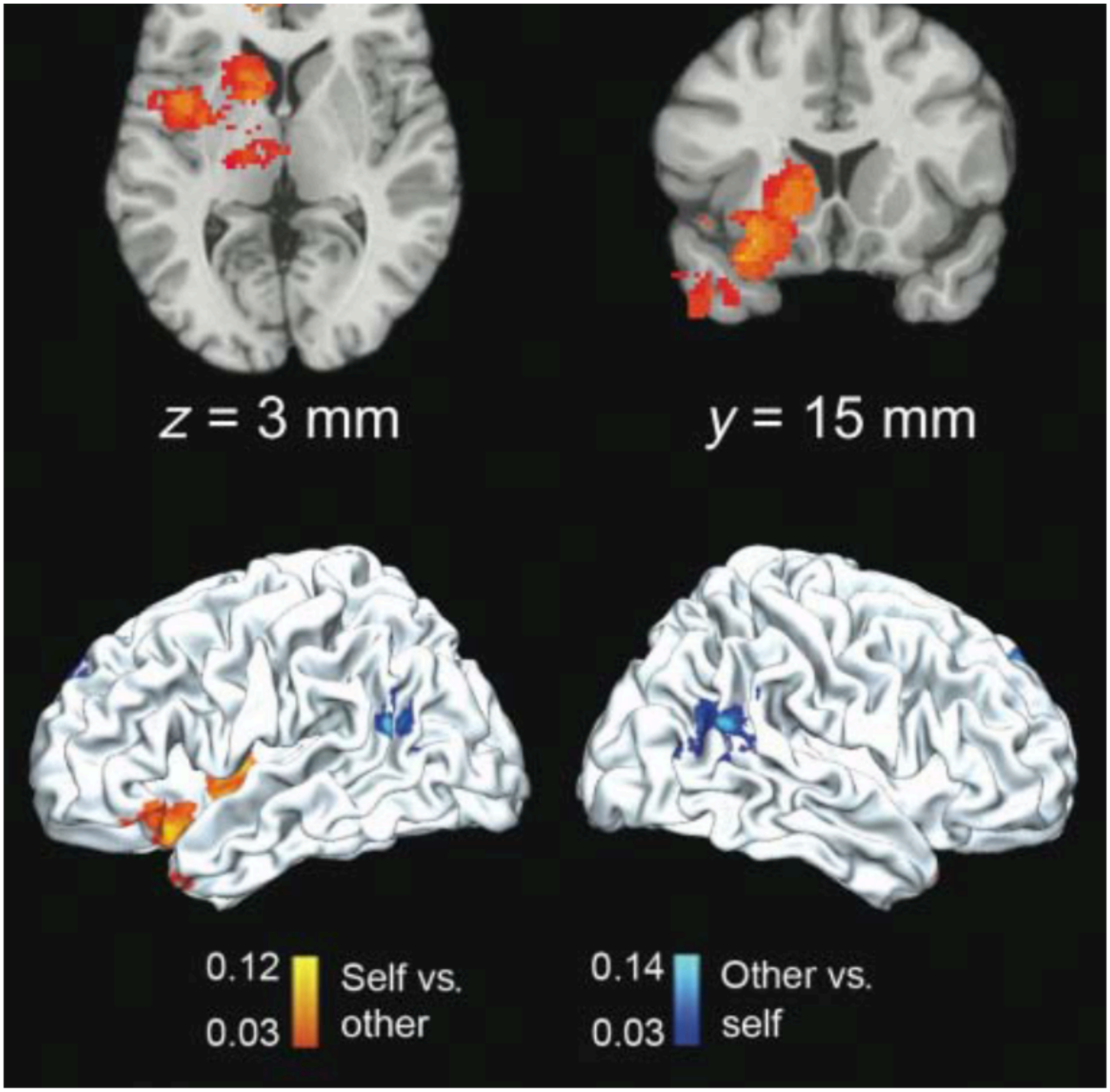


Figure 1. A meta-analysis comparing self (orange) and other (blue) mentalizing processes.

Meta-analyses are a great step forward from exclusively using single studies to glean insights about brain function because they yield statistically reliable results and are not hampered by the low power of fMRI. Due to this increase in power, one can interpret the spatial distinctions with more certainty.

However, meta-analyses are not a panacea for determining the functional specialization of the brain. A series of oft-overlooked steps have potential bias. First, to begin a meta-analysis, one must already have a hypothesis of the functions of an area of the brain in order to select relevant studies. This potentially creates bias because studies that may tap into this region's function but were not within one's scope of plausible function for that region may be overlooked. Second, once studies are selected, the relevant dimensions by which these studies differ must be selected. That is, if one knows a certain general function is involved with a region, one must decide how to subdivide this function into a finer grain in order to assess which regions specialize in these sub functions. This step introduces confirmation bias into the analysis as existing theories of subdivision of function are more likely to be tested, but novel and potentially powerful distinctions in function may go unnoticed.

Additionally, meta-analyses are not necessarily designed to *differentiate* brain regions on a large scale. Typically, one must hand code the coordinates that go into a meta-analysis, necessarily limiting the scope. In addition, while a meta-analysis may reveal specialization and overlap in function, it is not specifically designed to discover which dimensions differentiate function. For example, from the previously mentioned meta-analysis it is clear which regions specialize in self and other mentalization. However, while it is also clear that another region is commonly activated by both processes, it is not clear which functions would additionally discriminate it from neighboring cortex. Simply put, meta-analyses are not designed to yield information about which functions distinguish brain function across all possible cognitive processes. One primary reason is that a very large number of activations and their accompanying functions would be necessary for such an analysis.

Large-scale neuroimaging informatics

Fortunately, due to the realization of the need for more data, large databases of neuroimaging studies have recently been created. The availability of such a great amount of neuroimaging data associated with cognitive function allows the possibility of using cutting edge analytics to determine the functional specialization of the brain in a data-driven way that reduces the introduction of bias and allows for novel distinctions in specialization. Large amounts of data allows for greater statistical power and reliability, two aspects necessary for data mining. In addition, given the large variety of tasks that are contained in these databases, they allow for novel discoveries and connections to be made that would be difficult using a smaller variety.

Resting-state analyses

One approach that has readily taken advantage of large neuroimaging databases is the analysis of resting state functional MRI (rs-fMRI) data. During rs-fMRI subjects lie in the scanner at rest while their brains are scanned for 5–15 minutes. It is thought that spontaneous activity between brain areas reflects the intrinsic connectivity structure of the brain. Sophisticated graph-theoretic methods have been applied to such data to reveal potential models of how different brain regions are organized. These analyses have revealed that some areas of the brain, such at the dorsal attention network, behave as a community, or small-world network, which is likely to be performing a type of process. On the other hand, other areas such as the posterior cingulate cortex seem to community to many areas, likely serving to integrate multi-modal information across domains. Certainly, there is much to be learned from analyzing the connectivity structure of brain regions; however, it is important to note that the connectivity pattern of certain regions is being used to infer its likely to function. This assumption, while reasonable, still requires an inference to be made and is not a direct measure of a brain region's functional specialization.

BrainMap

Databases which pair functional data with ontological labels of cognitive function, such as BrainMap or Neurosynth, allow you to gain more direct insights as to the functional specialization of brain areas. BrainMap is a database in which authors upload their statistical maps from their fMRI studies and manually tag each contrast using an regulated ontology that describes the type of cognitive process that was targeted in the study. The advantage of this approach is that the data is curated with care, resulting in accurate ontological labels. As of this writing, BrainMap features 2336 studies with 45188 subjects - clearly a great leap forward from small-scale meta-analytic approaches.

Data mining has been applied to BrainMap in various cases and have led to interesting discoveries. Data decomposition methods (ICA) have been applied, resulting in brain networks that converge with those produced by rs-fMRI mining. In addition, large-scale meta-analyses have been performed to generate a functional atlas based on the different domains of cognitive processes defined in BrainMap's ontology (Figure 2). BrainMap has also been used to characterize the diversity of function across regions by determining the entropy of function within areas (Figure 3).

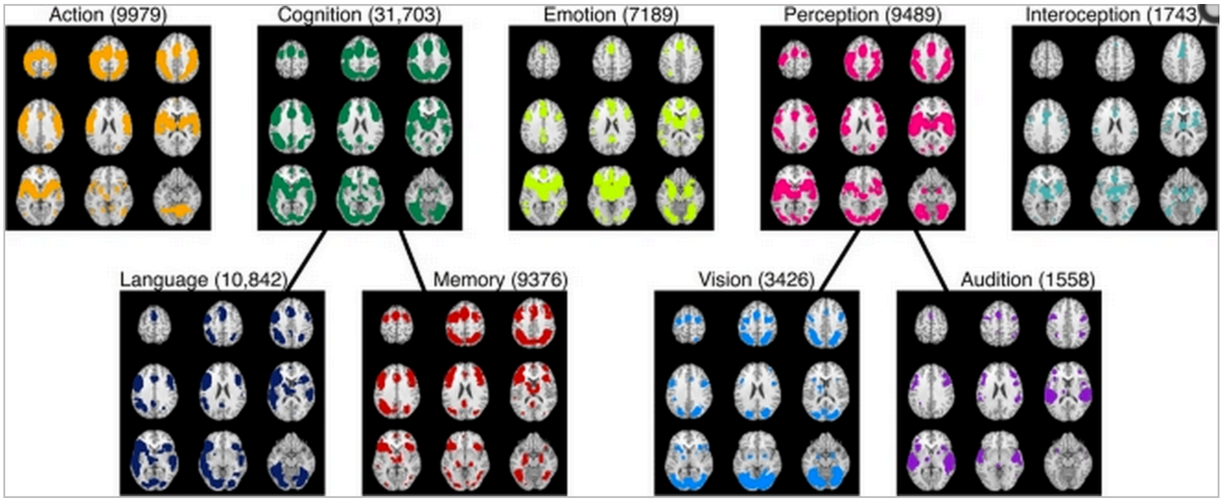


Figure 2. Atlas of the brain as generated from the BrainMap database

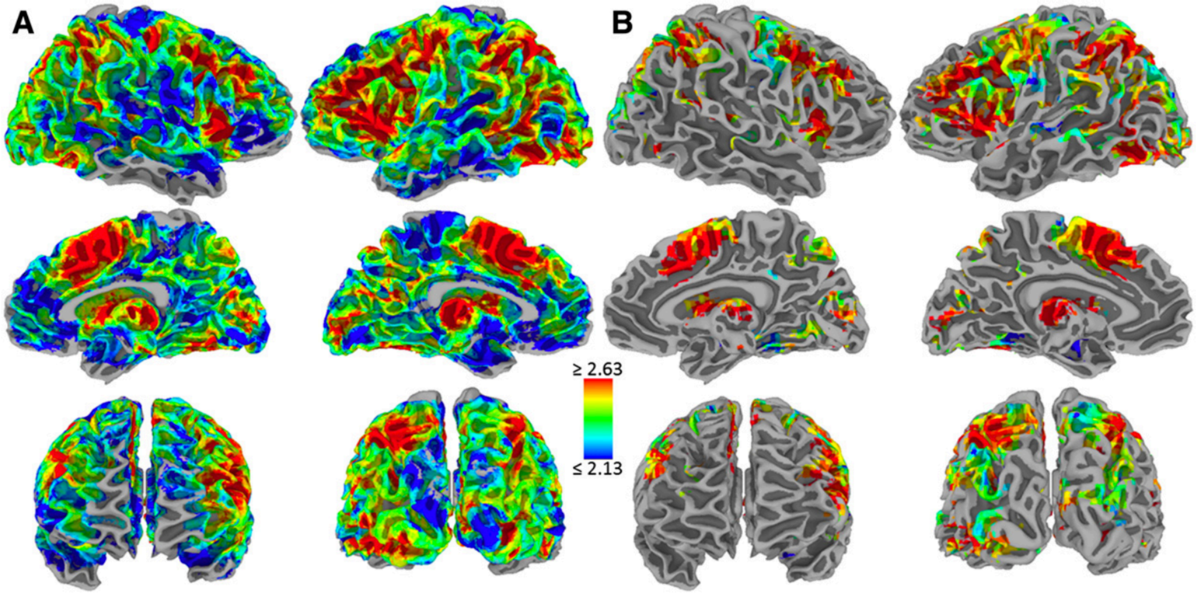


Figure 3. Functional diversity map generated from BrainMap. (A) Areas of higher diversity are shown in warm colors and areas of lower diversity are shown in cool colors (color bar represents Shannon entropy values). (B) Same as in part A, but masked so as to show only the locations where diversity was estimated with greater confidence.

Such approaches have been extremely fruitful due their ability to generate statistically robust, data-driven characterizations of brain function and diversity. However, several draw backs of BrainMap, in general, and the specific approaches exists. While BrainMap has been a fantastic leader in the generation of large-scale fMRI databases, it features a very limited and potentially artificial ontology. When submitting articles, researchers themselves must select the behavioral domain of their experimenter; for example, a brain researcher may need to decide between “Cognition.Memory.Working” and “Cognition.Memory.Implicit”.

Clearly, the problem of designing an appropriate ontology is extremely difficult; more-so, it may be inherently problematic because the ontology of the database defines the structure of the studies within it. In other words, the ontology that is used by the database may lead to some bias in the results that result from it. While this problem is be impossible to fully solve unless we understand fully understood cognition, more data-driven approaches that bypass a formal ontology may provide novel insights. As an aside, BrainMaps obtuse interface and restrictive data-sharing policy further inhibits its usability.

Neurosynth

The Neurosynth database sidesteps a formal ontology and manual entering of data by automatically mining the internet for fMRI studies. Neurosynth scrapes fMRI journals, such as NeuroImage, for coordinate tables that display the peak activations of the study. To obtain a measure of the cognitive function studied in a paper, Neurosynth calculates the frequency of the words in the text of the paper, after disposing of semantically useless words such as “the”. Surprisingly, this approach is successful and manages to produce sensible looking maps of cognitive terms; for example the term “value” shows a clear activation in the nucleus accumbens and the media prefrontal cortex, regions hypothesized to be crucial for valuation. In addition, while the quality of the data in Neurosynth is suspect, it is outweighed by the ease by which data can added. As of writing, there are near 5,000 studies in the database, with the number of studies set to double this year. In addition, the open nature of the Neurosynth interface makes it a perfect target for data-mining.

Focus on differentiation of function

An additional limitation of previous studies that has been not been addressed is the lack of focus on *differentiation* of function. As mentioned previously when discussing meta-analyses, many analyses of brain specialization focus on a specific dimension that is hypothesized to differentiate function in a specific area of cortex. For example, previously a meta-analysis attempted to differentiate the activation of mentalizing studies by whether the focus of the mentalization was on the self or on others. As already mentioned, this first requires an *a priori* differentiation of function and the analyses reveals if that dimension differentiates cortex spatially. Second, the analysis reveals which areas are common in activation, and which differ. However, for the areas that are common in activation, due to the scope of the analysis and the algorithms employed, it is impossible to tell what functions differentiation it from other regions.

Current Study

In the present analysis, we sought to apply machine-learning algorithms that specifically focus on classifying different areas in the brain, to reveal which cognitive functions differentiate amongst them. The algorithms employed here specifically allow one to reveal which features *differentiate* cortical regions spatially. We applied these methods both on a whole-brain basis, in order to determine the best functional identifiers of all brain regions, in addition to applying the method to targeted comparisons between regions that overlap greatly in function and their differentiation is debated.

Methods

Binary classification problem

The goal of this analysis was to classify two spatially distinct brain regions based on Neurosynth terms. In other words, we wanted to know which functional labels *distinguish* two brains regions. We opted to use a binary classification instead of a large multi class classification because there were not observations to support it.

In Neurosynth, there are two sources of data: a table that matches studies with the coordinates presented within it and a table that matches studies with the frequency of the words within a paper. Thus, the basic unit of classification for our analysis is a study in Neurosynth.

First, we selected studies that activated Studies were selected using a threshold of activation; if a study activated greater than 6% of the voxels within a mask, it was labeled as containing activation for that region. This was done for both regions, resulting in a list of studies that activated for each. Studies that activated for both regions were omitted from analysis (Figure 4).

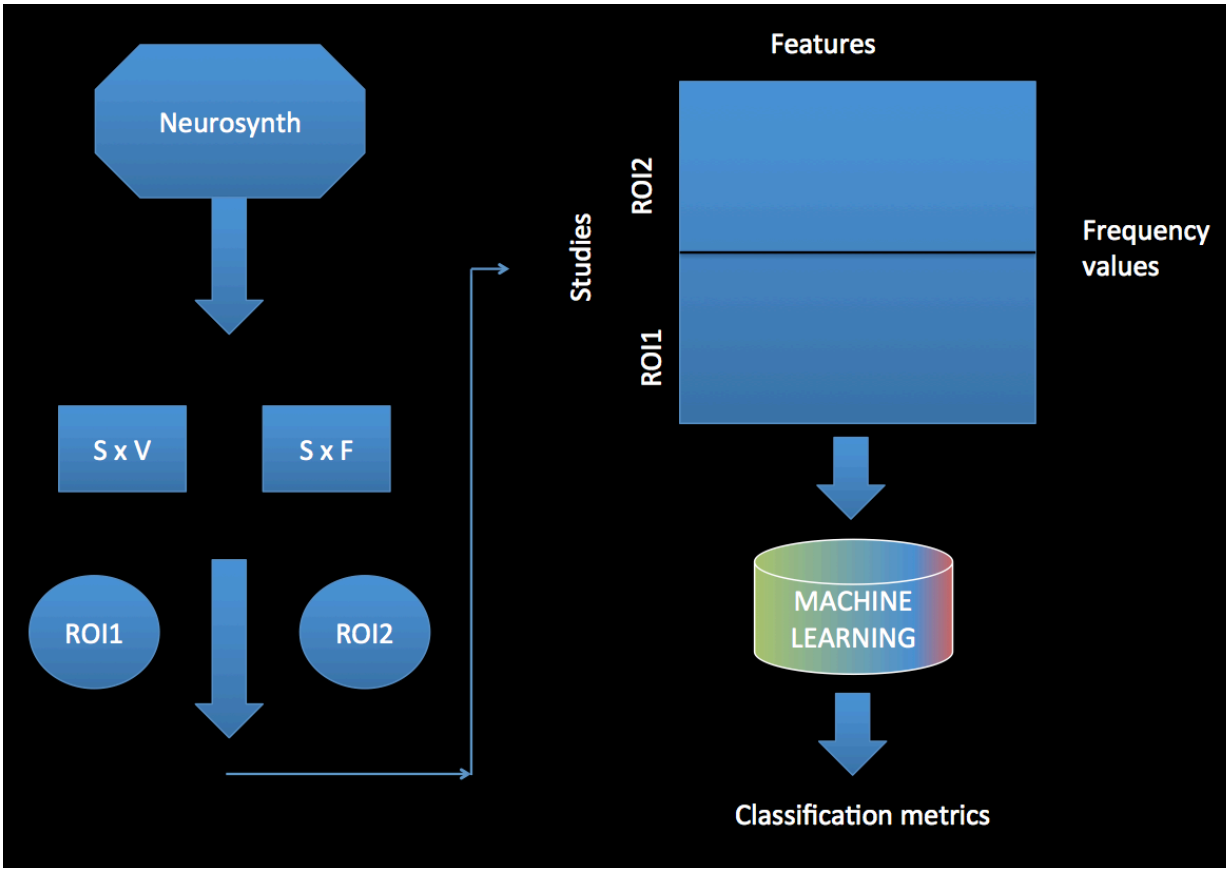


Figure 4. From Neurosynth two matrices, studies by voxels and studies by features, are extracted. Using two region of interest (ROI) masks, the relevant studies are filtered and labeled as activated by ROI1 or ROI2. Machine learning algorithm learns to classify studies and is tested using 4-fold cross validation.

Having obtained a list of studies that activated the two regions, we used a machine-learning classification algorithm to classify the studies based on their NeuroSynth term frequencies. The algorithm attempted to “learn” which features could be used to classify whether a study activated one region versus another. 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 likely to activate the visual cortex. The classification was tested using stratified 4-fold cross validation; classifiers were trained on 3/4ths of the data and tested on the remaining 1/4th, ensuring that an even distribution of classes was found in each fold.

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 classifier.

Because decision trees exhibit high-variance but low-bias, it is possible to aggregate their predictions using a vote and result in a low-variance low-bias prediction.

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.

Scoring metric

Accuracy was used as the scoring metric. Because the number of studies that may go into any binary classification may shift, the chance rate will also shift. Thus, we calculated the chance classification rate and subtracted it from the actual accuracy, resulting in the percentage above chance classification rate.

Application to whole-brain and resulting metrics

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. The brain was divided using Craddock’s 30 region parcellation of the brain based on rs-fMRI parcellation analyses. The results of the analyses were then aggregated resulting in various metrics for each region.

Overall classifiability

The rate at which a specific region was able to be classified against all other regions. The mean of that region’s binary classifications against all other regions.

Top features for each region

Terms that were most important in classifying a given region versus all other regions. Because terms varied in their overall importance across the entire brain, we also demeaned each region’s feature importances with respect to the entire brain to focus on the features that were important for a specific region in particular.

Diversity of functional specialization

We sought to asses the diversity of functional specialization across the brain. Regions with greater diversity of function are hypothesized to be involved in a greater number of cognitive or behavioral functions. We can approximate this by assessing the diversity of the feature importances for each region. In other words, if our algorithm exclusively uses a set of features to classify a region against all others, it is likely that that region's function is quite specific to those terms. For example, if studies the terms "vision" and "shape" are consistently used to predict studies that activate visual cortex versus other areas, it follows that visual cortex's function is very specific to vision. However, regions with a broader range of function may be classified by a wide range of terms, and these terms are likely to shift depending on which region is being used as comparison. For example, mPFC, a region that behaves as a hub and is likely to be involved in many functions, may be differentiated from memory areas with terms such as "emotion" while it may be differentiated from emotion areas with terms such as "memory.

We calculated Shannon's diversity to the feature importances output by the classification algorithm. Shannon's diversity is a measure of entropy which can be calculated for any vector with the following equation:

$$H' = -\sum_{i=1}^{R} p_i \ln p_i$$

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 varied in importance depending within a region depending on which region it was being classified against. In the mPFC example, the importance of the term "emotion" was high against one region but low in another, resulting in high diversity. In the visual cortex example, vision was high again all other regions, resulting in low diversity. We calculated Shannon's Diversity of each term within a region across all other regions and took the mean diversity across all terms.

Sparsity is the extent to which feature importances showed sparsity within a specific comparison between two regions. That is, if within a comparison between two regions the algorithm identified many terms as important for classification, sparsity would be low. If only a few terms were identified as important while others were near zero, sparsity would be high.

Neurosynth terms as features

Thus far, we have described our initial analysis in which we used neurosynth term frequency within papers as our measure of the cognitive or behavioral function that a study was about. Due to the variation in language across papers and general instability of nature language, these terms can sometimes be noisy or reflect variability across papers that is not cognitively important. This can be problematic because the algorithms used herein are greedy and will use any signal, irregardless of its theoretical cognitive importance, for classification.

Recently, Poldrack et al., (201x), applied topic modeling techniques to reduces the space of neurosynth features into 40 topics. These 40 topics reflect much of the variability found in Neurosynth but reduce the features into 40 coherent “topics” to which specific terms load onto. For example, a topic reflecting memory related terms has high loading by terms such as “memory”, “retrieval”, “encoding”, and “recognition”. The advantage of using these terms instead of raw term frequencies as the features for our classification algorithm is that they are more interpretable as a group of semantically sensible terms that individually. In addition, “junk” terms that do not reflect theoretically interesting signal group together (e.g. “age”, “adults”, “children”) and are easier to identify. Given that our goal is to understand the nature of what differentiates brain regions, this approach may yield theoretically more sensible results.

Results

In class I will present the results

Discussion