Predicting mental functions from brain activations

Madhura Parikh   Subhashini Venugopalan

1   Introduction

Advances in neuroimaging techniques and rapid growth of scientific literature in neuroscience have made large improvements in our understanding of human brain functions. Over the past few decades neuroscientists have studied brain images from EEG/MEG, fMRI and other sources to identify associations between psychological tasks and activity in brain regions [6]. Although these studies have led to large amounts of literature and several discoveries of bodily functions associated with certain brain regions (or networks) the mapping between functions to brain regions and vice-versa still remains largely unclear. In particular, different studies have associated the same regions to a wide range of functions [6] and scientists believe that there could exist unique underlying functions that brain regions specialize in.

2   Problem Definition

For the purposes of this project, we look at enhancing a new automated framework NeuroSynth [9] that combines text-mining and machine learning techniques to generate probabilistic mappings between cognitive and neural states. The framework addresses the following problems:

Forward Inference: In simple terms, given a psychological function or concept, the forward inference(a.k.a encoding) answers what regions of the brain are activated during that function. e.g. pain, vision, conflict, etc

Reverse Inference: The reverse inference(a.k.a. decoding) looks at the reverse mapping. i.e. Given a signature of neural activity, reverse inference identifies the cognitive state(s) and functions that the activations correspond to.

Both forward and reverse inference problems are quite challenging. However, forward inference has been studied well in medical literature and appears to not be of as much interest to the medical community since, neuroimaging techniques provide an easy way to verify the associations. On the other hand, reverse inference, which seems to be of greater interest to the community is also a difficult problem since multiple cognitive states could have very similar neural signatures [9]. In this project we therefore focus on the latter problem, and propose to use *transfer-learning* as a part of our solution. We also describe an forward inference approach for out-of-vocabulary concepts, as a possible future extension.

2.1   Related Work

Forward and reverse inference problems have been addressed by several contemporary works [7, 6, 3, 9]. However most of these approaches have been limited to manual efforts by individual

research groups, thereby failing to scale up with the increasing availability of fMRI data. Recent works like [9, 3] have therefore attempted to build automated frameworks for these tasks using machine learning and statistical techniques. In [9], the authors use a very simple model for both forward and reverse inference. Using some 3000 odd fMRI studies, they characterize each study by 2 feature vectors - **Activation vector (A_i)** and **Term Vector (T_i)**. A_i is a binary vector with $\approx 10^6$ features, a particular feature is set to 1 if the study reports that voxel to be activated and 0 otherwise. Additional pre-processing is carried out to ensure that the $3D$ coordinates that appear in the document are mapped smoothly to the voxels. T_i is a vector of length $\approx 10^4$, each feature indicating whether the corresponding term of interest appears above a pre-defined cut-off frequency in the document or not.

The forward inference i.e $\Pr[A|T]$was approximated using a simple frequency count with smoothing (i.e it was set to be the ratio of the number of times a term and activation point appeared together in a document, to the total number of occurrences of the term in the document, assuming that the terms had a uniform prior). For the reverse inference, a Naive Bayes approach was used by assuming that each of the activation points were independent and by applying the Bayes rule to the calculated forward inference probabilities. The Naive Bayes classifier was trained as a binary classifier (predicting which of the two given terms was the likely class label) as well as a multiclass classifier, considering sets of upto 10 terms (since considering all the 10^4 terms for presence/absence would result in an exponential blow-up). While both the feature extraction as well as the models used in this work are very simplistic, the results they obtain are very promising.

There have been other approaches that have further built on NeuroSynth. For instance in [3], the authors use label decomposition techniques in a one-vs-all setting to predict multiclass labels for the reverse inference problem. They use Support Vector Machines (l_2 regularized), logistic regression and ridge classifier for these tasks, comparing the outcomes based on different criteria like precision, recall and Hamming loss, to show that the multiclass approach is effective for reverse inference. They also propose to use other multiclass approaches and regularization techniques and analyze their performance as future work.

In another similar work [7] uses a Generalized Linear Model(GLM) for forward inference. For the reverse inference they use logistic regression with Ward clustering to counter the high dimensionality of the problem.

3   Proposed Work

Our project aims to primarily address the reverse inference task. We propose to more sophisticated algorithms to solve the classification problem currently addressed by the Naive Bayes classifier. Additionally, we plan to use techniques from transfer learning to determine the concept from the brain activations (and images) collectively. In addition, time permitting, we propose "zero-shot" forward inference for unseen concepts. i.e. Given a new concept for which we have not seen brain activation points, we hope to predict the activation network. A rough outline of our ideas to achieve this goal are mentioned below:

1. **Reverse Inference via more sophisticated classification techniques.** Currently NeuroSynth is available as an open source tool for the benefit of the neuroscience community. At its core, though it still uses the simple models that were introduced in [9] such as Naive Bayes classification. While there have been other approaches that have used more powerful

models, they have not been implemented concretely as open source tools. We propose to use a combination of logistic regression with hierarchical clustering and integrate these approaches to the current NeuroSynth framework. While we start with logistic regression as a simple yet reasonable model, we will be flexible in our approach and switch to a different model such as Support Vector Machines, if our experiments indicate it to be more promising. We also propose to use hierarchical clustering as a means of reducing the high dimensionality of the problem. By using hierarchical clustering we hope to reduce the large number($\approx 10^6$) of activation points to a more manageable size while at the same time not losing too much information, that may render our labels too general to be of much interest to neuroscientists.

2. **Reverse inference transfer learning.** A very interesting extension to NeuroSynth would be to make it possible to infer concepts given raw fMRI images. Currently NeuroSynth works only on binarized images (i.e images are thresholded so that voxels of sufficient activation are set to 1 and the rest are set to 0). However there are a vast number of real fMRI images that would be hugely benefitted by the various inference tools in NeuroSynth if they could be analyzed directly. This would require to do transfer learning where the source and destination have different domains but the same target. [4, 8] is a concise survey of existing transfer learning approaches. We propose to use some form of transductive transfer learning that enables us to make reasonable predictions even in the space of real images. We propose to use various co-clustering techniques as a first step towards this. Various such techniques like *Self Taught Clustering(STC)* [1] and *Transfer Spectral Clustering* [2] are some examples.

3. **Zero-shot forward inference.** (Time permitting) Given a new concept we would like to predict the brain activation points for this unseen concept. In order to do this, we plan to leverage contextual meaning of the context from text documents. We plan to represent each context in a distributional vector space constructed based on word co-occurrence statistics in neuroscience literature. These vectors give us a way of measuring similarity between two concepts in literature. ow, given a new concept, we can then identify the concepts/functions that are most similar and report the activation points that lie in their intersection.

4   Data Source

We intend to use OpenfMRI [5] as a source of raw fMRI images at the subject level. Apart from this we will also work with NeuroVault[1] data for fMRI images at the group level. NeuroSynth itself has an excellent repository of 8000 odd scientific papers that will aid us in the text-mining relevant sections of the project. All the data is open source and freely available via a convenient API.

References

[1] Wenyuan Dai, Qiang Yang, Gui-Rong Xue, and Yong Yu. Self-taught clustering. In *Proceedings of the 25th international conference on Machine learning*, pages 200–207. ACM, 2008.

[2] Wenhao Jiang and Fu-lai Chung. Transfer spectral clustering. In *Machine Learning and Knowledge Discovery in Databases*, pages 789–803. Springer, 2012.

[3] Koyejo Oluwasanmi and Russell A Poldrack. Decoding cognitive processes from functional mri.

---

[1] http://neurovault.org

[4] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *Knowledge and Data Engineering, IEEE Transactions on*, 22(10):1345–1359, 2010.

[5] Russell A Poldrack, Deanna M Barch, Jason P Mitchell, Tor D Wager, Anthony D Wagner, Joseph T Devlin, Chad Cumba, Oluwasanmi Koyejo, and Michael P Milham. Toward open sharing of task-based fmri data: the openfmri project. *Frontiers in neuroinformatics*, 7, 2013.

[6] Russell A. Poldrack, Jeanette A. Mumford, Tom Schonberg, Donald Kalar, Bishal Barman, and Tal Yarkoni. Discovering relations between mind, brain, and mental disorders using topic mapping. *PLoS Comput Biol*, 8(10):e1002707, 10 2012.

[7] Yannick Schwartz, Bertrand Thirion, Gael Varoquaux, Parietal Team, and Inria Saclay Ile-de France. Mapping cognitive ontologies to and from the brain. *Advances in Neural Information Processing Systems (NIPS)*, 2013.

[8] Matthew E Taylor and Peter Stone. Transfer learning for reinforcement learning domains: A survey. *The Journal of Machine Learning Research*, 10:1633–1685, 2009.

[9] Tal Yarkoni, Russell A Poldrack, Thomas E Nichols, David C Van Essen, and Tor D Wager. Large-scale automated synthesis of human functional neuroimaging data. *Nature methods*, 8(8):665–670, 2011.