COS 424: Interacting with Data

Lecturer: Ken Norman | Lecture #23
Scribe: Andrew Ferguson | May 1, 2007

Today's guest lecture was presented by Professor Ken Norman of the Psychology Department. Professor Norman is a neuroscientist and applies pattern classifiers, such as the one's learned about in this class, to functional magnetic resonance imaging (fMRI) data to try and decode the brain. His lab is known as the Princeton Computational Memory Lab and their website is located at: http://compmem.princeton.edu/.

1 Introduction

Figure 1: **fMRI machine in the basement of Green Hall**

The standard approach for brain studies is to place subjects in an MRI machine (see Figure 1), have them perform a cognitive task, and then examine the regions of the brain that are "lit up" on the MRI results. For example, you might perform a study in which subjects memorize names of celebrities and locations and then perform scans while asking them to retrieve those memorize. Examining which regions of the brain become active while retrieving memories can provide clues to identify the neural signatures of particular thoughts and memories. One interesting area of study is to track those signatures over time.

Neuroscientists work to answer the question: How does the brain process information? We can further ask how information is represented and what information *is* represented. Another big question is how information is transformed at different stages of neural processing.

The practical side of Professor Norman's talk today was to expose us to the problems of neuroscience. Today's techniques for analyzing fMRI data are not as good as the state-of-the-art techniques in machine learning. Professor Norman and others want to bring brain imaging "up to speed."

2 fMRI Basics

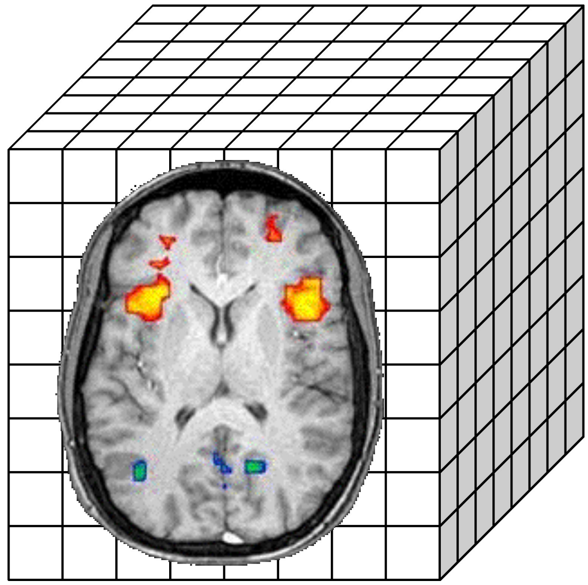

Figure 2: **Visualization of fMRI data separated into voxels.**

Regions in the brain that are more active use more oxygen than those regions that are not. The small difference between oxygenated and deoxygenated regions is detected by fMRI. The images are then analyzed by imagining the brain in a sliced 3-D cube. Each slice of the 3-D cube is known as a volumetric pixel or "voxel" (see Figure 2). Each voxel is 3mm on each side.

The accuracy of fMRI data suffers for several reasons. It requires two seconds to take a picture of the whole brain (one fMRI scan). Secondly, the size of each voxel is BIG relative to neurons, which are on the scale of micrometers. Furthermore, we are not directly measuring the electrical activity in the brain - only a level that is correlated with electrical activity.

One final note about fMRI is that the blood flow response is smeared out in time. Indeed, the peak occurs about six seconds after the neuron fires. Researchers have to adjust the timeline of their images to account for this fact.

3 Applications of Machine Learning

3.1 An Example Brain Study

A key insight is that cognitive states (such as recalling a specific memory) correspond to distributed patterns of brain activity. If we can identify those patterns in the fMRI data, then we can associate them with different cognitive states. This is the idea behind the study "Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex" (Haxby et. al, 2001).

In the study, subjects were shown images of faces, cats, five man-made objects, and nonsense pictures. The researchers recorded the subjects' brains' responses to the images and then applied a machine learning algorithm to them to predict the type of image from the brain scan. The algorithm calculated the average response to one-half of the data for each category, and then performed nearest-neighbor approximation on the second half of the data using the correlation across voxels as the distance metric. The algorithm was able to predict the type of image with an overall accuracy of 96%. The subjects were in the scanner for a total of thirty minutes each.

Prof. Norman's group extended the Haxby study by tracking two different cognitive states (looking at pictures of shoes or pictures of bottles) using single images. They converted the data into vectors that reflected pattern of activity across voxels at a point in time. They then trained a neural network classifier (logistic regression) in a supervised setting on the vector data and tested on new data. Professor Norman noted that fancier machine learning algorithms generally don't do better than basic ones on this data.

3.2 Free Recall and Mental Time Travel

The study "Category-Specific Cortical Activity Precedes Retrieval During Memory Search" (Polyn et al, 2005) examined how we retrieve memories selectively. If we think intuitively about the process of retrieving memories, one might start by remembering where we were, or who was around. These general thoughts trigger memories and prod recall. We get ourselves into the frame of mind that we were in at that time. Effectively, we are trying to "rollback" our brain state. This is described as "mental time travel." This study's goal was to image this rollback process.

The subjects were presented with three types of stimuli: pictures of famous faces, places, and objects and were asked to rate them. Next, they were instructed to recall what images they saw, in any order. A pattern classification algorithm was then trained on the fMRI data from the showing phase, and then used to predict the probability that the subject is remembering each type of picture during the recall phase. The success of the classifier would indicate how similar one's brain state is during the recall phase to how it was during the showing phase.

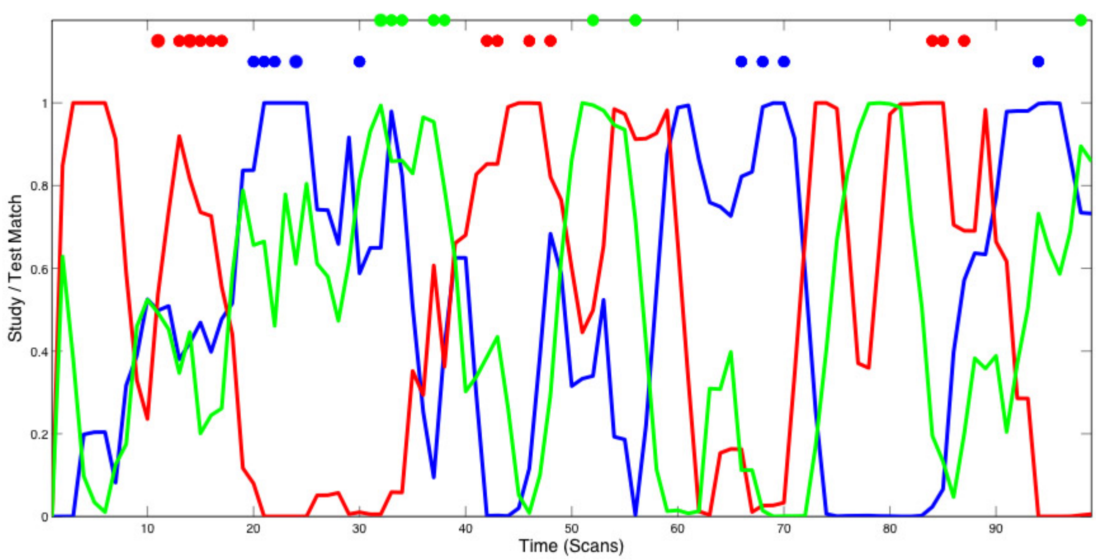

Figure 3: **Classifier predictions for Subject 9 - faces are blue, places are red, and objects are green. The dots indicate when pictures of each type were recalled.**

An example result from this study is shown in Figure 3. It is clear that the classifier has successfully assigned a high probability of remember a given type, just as the subject is indeed doing so. Furthermore, the pattern of probability for each type *rises* before the actual event of recalling pictures of that type.

¿From studies such as this one, neuroscientists have learned that some areas of the brain respond most strongly to certain types of pictures. For example, there is a "face area" that responds strongly to faces. The location of this face area tends to be consistent across people. However, it remains important to study all areas, looking at the distributed pattern of responses, since there is information contained throughout the fMRI scan.

3.3 Identifying Patterns in the Images

In the study "Decoding the visual and subjective contents of the human brain" (Kamitani and Tong, 2005), subjects were shown striped patterns of various orientations. The researchers used a machine learning algorithm to try and predict the orientation of the pattern being shown from the brain data. Their resulting predictions were accurate to within twenty degrees.

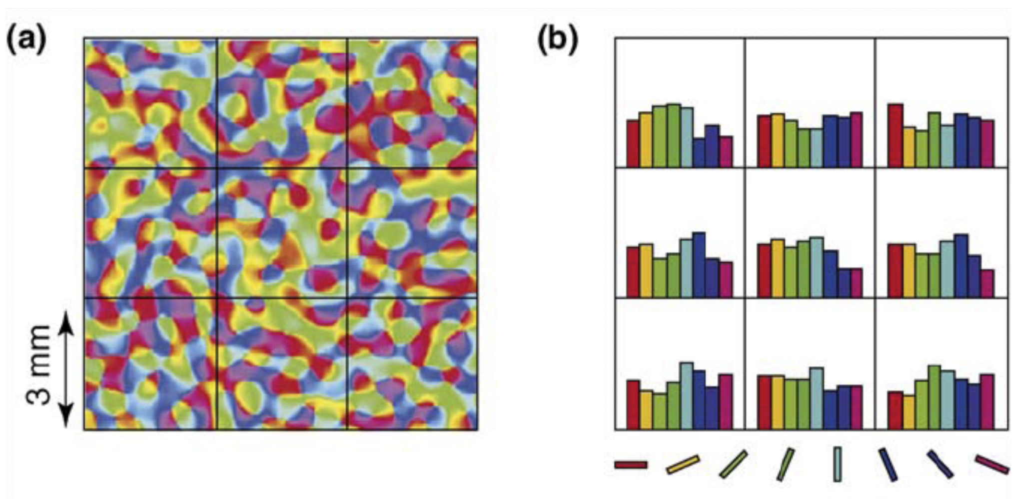

Figure 4: **Example of the association of neurons to orientations inside voxels.**

Different neurons in the brain are attuned to different orientations (Figure 4a). Scientists learned this by studying the physical properties of animal brains. Inside each of voxel are many neurons attuned to different orientations. However, the precise distribution over the orientations varies between voxels (Figure 4b). Using these small asymmetries to recover strong information is exactly the strength of linear classifiers.

3.4 The Pittsburgh Competition

Professor Norman and his lab group compete in the Pittsburgh Brain Activity Interpretation Competition (www.braincompetition.org). In the 2006 edition of the event, subjects were scanned while watching three episodes of Home Improvement and simultaneously making time-varying ratings of features such as amusement, food, tools, faces, etc. Competitors received the brain scans for all three episodes and the ratings from the first two. Using this data, they were asked to predict the ratings for the third episode. The Princeton team of Denis Chigirev, Greg Stephens and collaborators won second place.

This year's competition involves brain scans from subjects who wandered around in a virtual reality videogame while looking for things such as fruit and weapons. Competitors will have to predict what was going on in the game from the fMRI data. Professor Norman is looking for students to join this effort.

4 Technical Challenges

Interpreting brain scans is a particularly difficult problem for machine learning algorithms (Mitchell et al., 2004, *Machine Learning*). The data is high dimensional (20,000 x 20,000 matrices `[check]`) and very noisy. We are looking for big patterns, of which there are relatively few. Furthermore, most of the variance in the data is from non-interesting brain data (such as breathing) and experimental procedure. Of course, the challenges that make this problem hard also make it interesting.

How can we improve the situation? For starters, Professor Norman's lab group has tried every classifier they can implement in `MATLAB`. The choice of classifier doesn't seem to matter (much) however, which is consistent with the data being noisy and ill-posed. They did find that regularization helps a lot. For example, ridge regression outperforms standard linear regression.

4.1 Feature Selection

Feature selection has proven to be a useful means of improving the accuracy of the predictions. The data is preprocessed to remove noisy voxels. The standard approach to doing this is to run analysis of variance (ANOVA) on each voxel to see if it is able to make discriminations above a certain threshold. Voxels that by themselves make poor predictions generally make little positive contribution to the overall prediction. Of course, throwing away even weak voxels is theoretically disadventageous since the benefit of linear classifiers is that they can produce a strong output signal from many weak inputs.

To reach beyond this problem, we need to evaluate voxels on a multivariate basis. Ultimately what we are looking for is the smallest set of voxels that carries the most information. However, naively searching for this set is combinatorially explosive.

A promising heuristic, called "searchlights," for approximately finding this set is outlined in the paper "Information-based functional brain mapping" (Kriegeskorte et al., 2006). The paper proposes sliding a sphere of voxels around inside the brain image, and computing in a multivariate manner how well the voxels in the searchlight will perform. Good searchlights are kept track of, so they can be added to the final set.

4.2 Dimensionality Reduction

Another promising technique for reducing the complexity of the problem is dimensionality reduction. Recoding the data is successful because there is extensive redundancy across voxels (especially spatially proximate voxels). This is due to biological reasons such as the evolution of the brain and the anatomy of the circulatory system. Clearly, the voxels are not independent. Thus, searching for a more efficient way to represent the data seems fruitful.

An example technique used for dimensionality reduction of fMRI data is manifold learning. Manifold learning tries to find a lower dimensional space in which useful information lies embedded in the high dimensional space of all the information. Other techniques include spatial wavelet decomposition, ICA, and generative models for brain states. This work, by David Weiss and David Blei models brain states as being a linear combination of "neural topics," which are specific patterns of voxel activity across the brain.

5 Future Directions

An obvious direction for future research is to try and increase the amount of information we are feeding the classifiers. Currently, we know a lot more about the brain, fMRI response, and cognition than we give the classifier to work with. Similarly, in current models, each time slice is treated as distinct. In actuality, we know there is strong temporal autocorrelation due to the sluggishness of the blood flow and because our brains usually shift slowly between topics.

There is also spatial correlation to contend with in the data. Researchers have tried to use spatial smoothing, but the constraint is that we don't want to smooth over useful information, which is already coarse-grained as compared to the scale of neurons. We can try to measure the correlation between voxels and give that information the classifier.

Current analyses are focused on a single subject at a time. Can we leverage data from other subjects to improve our classifier? For example, in the Haxby study (Section 3.1), can we use other subjects' activity when viewing shoes to intelligently set priors for the next subject? Can we use data from other fMRI studies? We already know that there are areas of the brain that tend to respond to faces or places. But how can we organize all of this data, controlling for variations in experimental protocol?

Additional information that the classifiers are not considering is the hierarchical nature often inherent in cognitive states. For example, if we are looking at pictures of animals – such as bears – and classifying them according to dangerousness, our brain scans will probably look different than if we are classifying them according to whether they are land or sea animals. This is connected to how people control their thoughts. Our prefrontal cortex biases which brain regions are active based on which tasks are performed. If we can detect that bias in the fMRI data, then we can look for correct kind of "bear" pattern.

Graphical models have the flexibility to teach classifiers about these different constraints (temporal autocorrelation, spatial correlation, data from other subjects, data from other experiments, and hierarchical structure).

6 Information from the Q&A Session

When a researcher is showing a subject pictures of houses, we never actually know that they are thinking about "houses," but we trust that there is a higher probability of that, so our results are meaningful on average. The concept of "task switching" when the brain changes which task it is performing, such as from concentrating on the experiment to dreaming about a vacation. In psychology studies, we can see "cognitive lapses" in the brain data where the subject wasn't doing what the researcher asked. Behavioral cognitive studies are forced to assume that the subject is actually in the desired cognitive state, whereas in fMRI studies, researchers have the advantage of actually "looking" at the brain.

As always, it is perilous to over-interpret the results of the classifier. They can predict that there is information in a given region of the brain, but not necessarily what it means. For example, we could theoretically apply classifiers to the early visual cortex, learn the coding of images there, and then predict the type image (for example, “a person”) based on the pixel-like response of the nerves in that region. This does not, however, mean that there are neurons in the early visual cortex that detect people. Classifiers can provide hypotheses for further studies that might examine subjects who have brain damage (or temporary malfunction) in the region of interest.

Classifiers are not useful for assessing information voxel by voxel. They make reliable predictions from an ensemble of voxels.