Machine Learning Classification of fMRI Data in Semantic and Syntactic Tasks

Jay Pujara

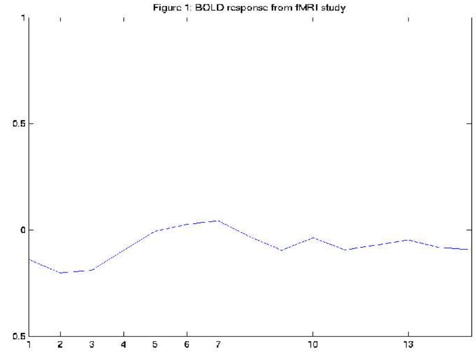
Abstract

The mental processes underlying the understanding of words and sentences are still poorly understood. Computational analysis of data from Functional Magnetic Resonance Imaging (fMRI) experiments has been an invaluable tool for understanding the patterns of neural activation associated with mental processes. Analysis of data can be very difficult due to interference between stimuli as well as noise in fMRI measurements. This thesis focuses on discovering if there are differences in brain activity between classifying words and classifying sentences by using a naïve Bayes classifier. Further, the differences in neural activity are quantified at various levels of detail to determine where such processing takes place. To achieve this goal, proper methodology and procedure for analysis of fMRI data was developed and refined to ensure legitimate conclusions.

1. Introduction

The emergence of functional Magnetic Resonance Imaging (fMRI) in the last decade has created many new opportunities for neuroscientists and computer scientists. fMRI allows the collection of data correlating to neural activity at fine temporal and spatial granularity throughout the entire brain. Specifically, neural activity can be sampled at a temporal resolution of half a second and a spatial resolution on the scale of millimeters. The metric of neural activity in fMRI studies is the blood oxygen level dependent (BOLD) response. The BOLD response (Figure 1) measures the ratio of oxygenated hemoglobin to deoxygenated hemoglobin, and reaches its peak five seconds after stimulus onset. By measuring the BOLD response of a 3x3x5 mm^3 portion of the brain, referred to as a voxel, observed neural activity can be assigned to functionally or anatomically similar regions of the brain, referred to as regions of interest (ROIs). However, the high resolution of fMRI data yields on the order of 10,000 observations for each second of a study and millions of observations during the course of an entire experiment. The vast quantity of data gathered through fMRI experiments makes human analysis of observations impractical. Instead, the quantity, sparseness and noisiness of this data make machine learning methods particularly appropriate for analysis of fMRI studies. The goal of this thesis is to find the answers to some compelling questions about the brain with the aid of machine learning classifier.

Figure 1: BOLD response from fMRI study



The fundamental question this thesis answers is "What are the differences between the processing of words and sentences?" More specifically this paper investigates the differences between syntactic processing, where a subject classifies the part of speech of a given word, and semantic processing, where a subject evaluates the meaning of simple sentences. Similarly, this work may lend insight to current research on shallow processing, where superficial qualities of a word are considered, and deep processing, where the meaning and attributes of the word become important. While previous approaches to answering similar questions have used average activity in specific regions of the brain, the motivation behind this study was to use a classifier to differentiate between semantic and syntactic tasks. By creating a classifier with a high success rate, the systematic differences between syntactic and semantic tasks can be isolated. Based on comparable research, classification accuracy greater than 90% would be considered a successful classifier.

Results of this research can be used to understand how humans process language, aid those with linguistic impairments, and establish a protocol for analyzing fMRI data from such studies. Based on prior work and the sparse and noisy nature of the data, the naïve Bayes classifier is best suited for the analysis of data. Prior machine learning studies for similar analysis of fMRI data have found good accuracy (greater than 80%) in classifying brain activity in such tasks as sentence ambiguity and word categories. Naïve Bayes classifiers are conventionally used for large data sets with sparse data. Additionally, research contrasting different classifiers on fMRI data sets has shown that the naïve Bayes classifier is particularly suited for fMRI analysis. (Mitchell, 4/2003), (Specific references, results).Prior studies of fMRI studies on sentence comprehension have discovered significant activity in the left temporal and inferior frontal cortex regions of the brain, among others. (Description of studies and results).

3. Methodology

3.1 Experimental Setup

The Brain Lexicon experiments were engineered with three phases, the first focusing on slow syntactic categorization, the second requiring fast syntactic categorization, and the third requiring slow semantic categorization. In phases one and two words are classified as nouns or verbs, while in phase three noun-verb sentences are classified as true or false. In phase one, words were presented for one second during which the word was categorized as a noun or verb, followed by fourteen seconds of fixation, yielding a total trial length of fifteen seconds. In phase two (Figure 2), words were presented for one second during which the same noun-verb categorization occurred, followed by two seconds in fixation, yielding a total trial length of three seconds. Finally in phase 3 (Figure 3), the subject was given surprise instructions requesting they categorize simple noun-verb sentences as true or false. The noun was presented for one second followed by two seconds of fixation. Next the verb was presented for one second during which the true-false categorization occurred, followed by fourteen seconds of fixation, yielding a total trial length of eighteen seconds. Phase one contained 16 trials, phase two contained 120 trials, and phase 3 contained 18 trials. Although phase 2 provided the most data with the most intense activity, it provided some particular challenges. Specifically, the BOLD response requires five seconds to reach its peak value, and can take up to twenty seconds to fully subside. As trials lasted three seconds, the noise and intensity of the observations presented a challenge to analysis. In the initial attempt to create a classifier, the data used for analysis was taken from a single subject.

3.2 Feature selection

The desired comparison was between Phase 2 and Phase 3 due to the quantity of data available in Phase 2. As the length of Phase 2 trials was three seconds, three seconds of the fifteen second long Phase 3 trials had to be selected to make a reasonable comparison. Three approaches were evaluated to select the appropriate window to compare against Phase 2: the response to the noun of the sentence, the response to the verb of the sentence, and the average of these two responses. One issue behind each of these selection criteria was the onset of semantic processing. If semantic thought began at the presentation of the subject of the sentence, the use of the noun would be most appropriate. However, if semantic processing occurred only after the predicate was presented, the verb presentation would be the appropriate stimulus for data. If processing was occurring continually, the average presentation might be a suitable feature to present to the classifier. Each of these methodologies for feature selection were tested in limited scope and the most successful approach seemed to be the presentation of the verb data. The features finally selected for classification were three seconds of fMRI data concatenated into a single data vector.

3.3 Approach

3.4 Testing procedure

Testing and evaluating the data was divided into several phases. First, statistical tests were used to evaluate the semantic and syntactic data against a fixation condition. Next, neural activity was classified into semantic or syntactic through the uses of a Bayes classifier. Finally, the statistical tests were paired with classification accuracy to distinguish precisely what portions of the brain were responsible for semantic processing, and which portions were used specifically for syntactic processing. Validating the data statistically involved determining the mean response of semantic and syntactic stimuli and using t-tests to measure the significance of neural activity with respect to fixation. Classification tests included evaluating accuracy of classifying data as semantic or syntactic activity in a variety of sampling configurations, such as using the entire brain for classification, using specific regions of the brain, using the most active voxels, and using voxels individually. In each scenario, the classifier was trained using leave-one-out training. Pinpointing the neural activity specific to semantic or syntactic processing involved selecting the individual voxels that had the highest accuracy classifying between the two conditions. Next, if the mean activity of the voxel was consistently greater throughout the semantic or syntactic processing task, these voxels were thought to be important specifically to semantic or syntactic processing.

4. Results

4.1 Classification

4.1.1 Classifier Accuracy

The fundamental question considered in this thesis is whether there are significant differences in syntactic and semantic categorization tasks. Results of classifier training (Figure 4) show that there is a significant difference as evidenced by the accuracy of a naïve Bayes classifier. The classification accuracy was over 90%, and achieved nearly 95% accuracy in select regions of the brain. Additionally, the classification results show that feature selection of verbs represent the most reasonable approach in data analysis. To elaborate, the presentation of an average between the noun and verb resulted in artificially high accuracies. It seems reasonable to assume that by averaging two different portions of the task, some artificial data state was created. Presentation of only the noun feature of the sentence presentation yielded lower accuracy, possibly signifying that syntactic processing was still occurring before the predicate of the sentence was presented.

Sampled Region	*Classification Accuracy*
Full Brain (10,342 voxels)	93.48%
300 most active voxels	94.93%
30 most active voxels	96.38%

Sampled Region	*Classification Accuracy*
Left Dorsolateral Prefrontal Cortex	94.20%
Left Temporal	88.40%

Figure 4

4.1.2 Salient differences in syntactic and semantic processing

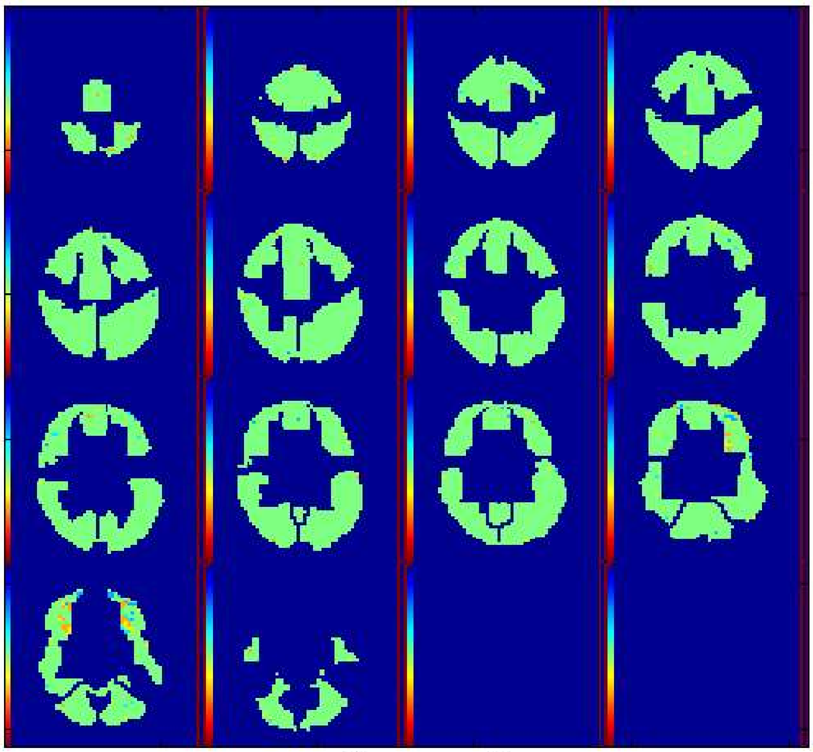


figure 5

The regions demonstrating the best accuracy in classification include left and right dorsolateral prefrontal cortext, right temporal, left and right extrastriate cortex. Examining the data at further granularity shows that voxels in left and right extrastriate, and left and right dorsolateral prefrontal cortex are specific to syntactic processing and voxels in left dorsolateral prefrontal cortex are associated with semantic processing. In sum, separate portions of left dorsolateral prefrontal cortex play the largest distinguishable role in semantic and syntactic processing. The figure below (Figure 5) quantifies this claim, with orange sections representing semantic processing and blue regions representing syntactic processing.

4.2 Characterization of data

4.2.1 Mean response characterization

The mean response shows strong BOLD responses beginning at the second and fifth seconds of Phase 3 trials while no BOLD response seems apparent in Phase

2. Categorization of mean response by region shows more significant responses. (*Images* of mean responses and per-region response).

i. **Tests for validity**

1. T-tests
   Discussion of t-tests, showing data significant from noise in specific regions. (*Image*).
2. Linear-regression
   Discussion of linear regression results. Show where activity conforms to typical response (*Image)*.

II. Conclusion

a. Interpretation of results

Classification of syntactic and semantic processing is successful, with 96% accuracy. Furthermore, specific areas found to be critical to syntactic processing include (list areas). Specific areas found to be critical to semantic processing include (list areas). (Contrast to other studies).

b. Evaluation of Methodology/Future directions

Experimental procedure made analysis of the data difficult. (Evaluation of strengths and weaknesses of methodology). Future experiments would be improved by taking measures to separate trials by at least five seconds and taking significantly more measurements for semantic categorization. (Discussion of experimental revisions).

III. References