Computer Science 229
December 14, 2013
Primary authors: Paul I. Quigley '16, Jack L. Zhu '16
Comment to piq93@stanford.edu, jackzhu@stanford.edu

Decoding the Human Motor Cortex

Abstract: A human being's motor cortex becomes engaged when he or she performs specific motions. However, little is known about the relationship between the regions in the motor cortex. Neuroscience has not reached consensus over how motions for specific joints relate, or how engagement changes when a motion is performed more forcefully. We obtained fMRI BOLD scans of subjects' motor cortices as they performed various movements. We found that we could distinguish between movements with considerable accuracy. Moreover, we confirmed that these techniques relied solely on data within the motor cortex, not skull-movement or other disruptions.

BACKGROUND: BRANTNER EXPERIMENT

A recent paper authored by members of the Stanford Artificial Intelligence Laboratory, *SVM Based Classification of Motor Tasks on fMRI-BOLD*[1], explored the technique of classifying motor tasks to gain insight into the structure of the motor cortex. We were inspired to first replicate their results, in order to verify our implementation of the machine learning techniques, and also to gain insight into where their model and techniques fell short. In their experiment, a subject performed one of ten tasks while being monitored by an fMRI scanner. During each task they were either holding a light (.05 kg) or heavy (.5 kg) weight. Table 1 lists these tasks.

TABLE I: The list of tasks. For simplicity in later results, each task was given a number 1-10.

Task # (light/heavy)	Motion
1/6	Wrist-Flex
2/7	Wrist Rotate
3/8	Elbow-Flex
4/9	Shoulder-Flex
5/10	Shoulder-Rotate

Each of three subjects performed a total of 116 motions. The pertinent data gathered from the experiment was a measurement of the activation in each voxel, a $2.5 \times 2.5 \times 2.5 mm^3$ region in the brain; the scan area was a rectangular prism of $64 \times 64 \times 24$ voxels, resulting in a scan volume with a total of 98,304 voxels. The fMRI volumes were taken every 1.57 seconds.

The data was preprocessed with GLMdenoise, a software package by Kay et al [2]. GLMdenoise uses statistical methods to derive global noise regressors for use in a General Linear Model (GLM) analysis. GLMdenoise denoises the raw time-series data, and produces a "modelmd" volume for each task type, that represents the peak percent signal change in the BOLD (Blood Oxygen Level Dependant) signal when it is fitted to the Haemodynamic Response Function (HRF). Each of the "modelmd" volumes represent, intuitively, which voxels gave the clearest signals of engagement during a task. Figure 1 and Figure 2 are examples of such "modelmd" volumes, corresponding to the unweighted wrist-flex and unweighted elbow-flex movements, respecitvely. Looking at the figures, it is visually evident that the activation regions are different for different motions. This is a strong indication that distinguishing between motions using fMRI data may be an appropriate machine learning task.

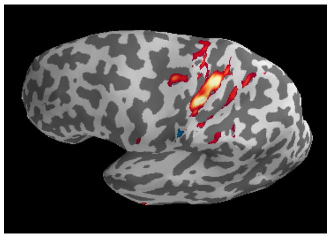

FIG. 1: : Highlighted have voxels "modelmd" scores above a threshold for the unweighted wrist-flex.

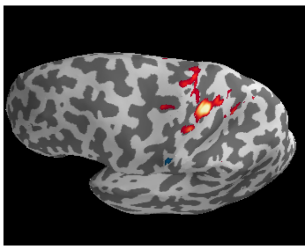

FIG. 2: : A different, though not disjoint, set of voxels exceeds this threshold for the unweighted elbow-flex.

In Brantner, an SVM was trained with the following model: A subject performed tasks P and Q each p and q times respectively. Define a time interval around each task so that a total of T measurements were made during each time interval. Also, choose some set of voxels $V = \{v_1, ..., v_j\}$. Now for each voxel v_k we will measure its activation at some time t within the interval for the motion r. We can denote this $a_r^t(v_j)$. We seek for each of our training examples to contain information about every voxel over the entire response. So we have $p + q$ training examples of the form:

$$x^{(i)} = \begin{bmatrix} a_i^1(v_1) \\ \dots \\ a_i^T(v_1) \\ \dots \\ \dots \\ a_i^1(v_j) \\ \dots \\ a_i^T(v_j) \end{bmatrix}$$

$$y^{(i)} = \begin{cases} 1, & \text{if } i\text{th task is } P \\ -1, & \text{if } i\text{th task is } Q \end{cases}$$

Brantner achieved the highest classification accuracy when they let V be 50 voxels sampled from the top 500 R^2 voxels (see **Voxel Selection** for more detail), and a time window of 12-14 seconds.

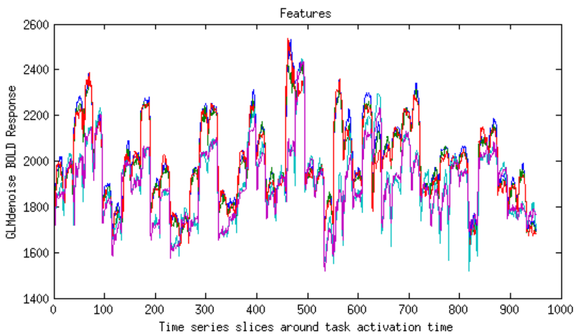

FIG. 3: An example of the time-series features of five high R^2 voxels.

REPLICATING RESULTS

We used this model, and ran an SVM with a linear kernel, using SMO and leave-out-one-cross-validation. The SVM implementation we used was MATLAB's *svmtrain* and *svmclassify*, in the Statistics Toolbox. We obtained results nearly identical to those reported in Brantner, who used the LibSVM implementation. We used non-linear kernels as well, but found much lower classification accuracy. As in Brantner, we see a curious a result when we divide our comparisons into 4 groups. In group 0, we compare light-weight tasks to other light tasks; in group 1, heavy to heavy; in group 2, light to heavy; in group 3, we compare light-to-heavy, but of the same motion.

We see that our accuracy is significantly lower in group 3, barely over 60%, while the other groups have accuracies of about 90%. A central goal of our project was to explore the causes of this lower accuracy in group 3, and devise ways to raise it.

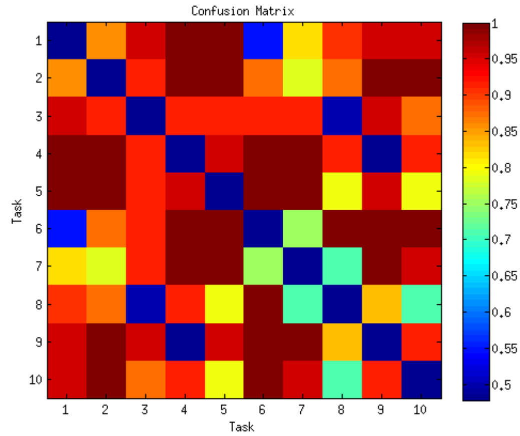

FIG. 4: : Confusion matrix. ijth cell represents the accuracy of classifying task i vs. task j

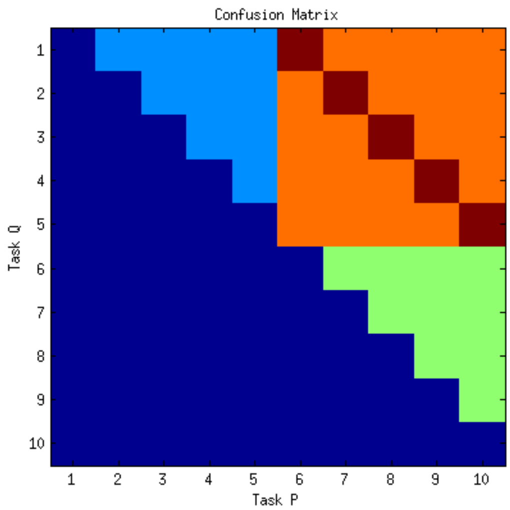

FIG. 5: : Each comparison is colored according to its group.

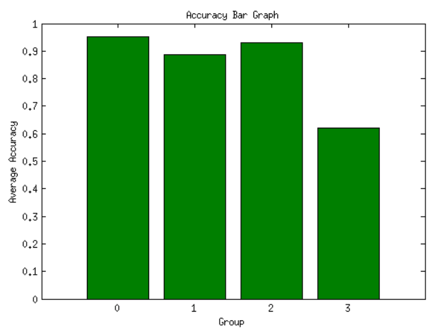

FIG. 6: : The average accuracy for comparison within each group. Group 3 shows notably lower accuracy.

VOXEL SELECTION

In Brantner, voxels were included in V based on their high R^2 values, calculated by GLMDenoise. The R^2 value indicates the variance in the time-series response for that voxel expalained by the onset of the task. Like the "modelmd" scores, these serve as an indicator of which voxels showed the clearest signs of engagement; unlike the "modelmd" scores, these take into account responses over the entire experiment, not just one task. The logic is that the voxels with the highest R^2 are most reliable; however, in Brantner it was not explored where these voxels actually were. We built a data visualization pipeline to plot the location of specific voxels on to an image or 3D model of the brain.

(a.) (b.) (c.) (d.)

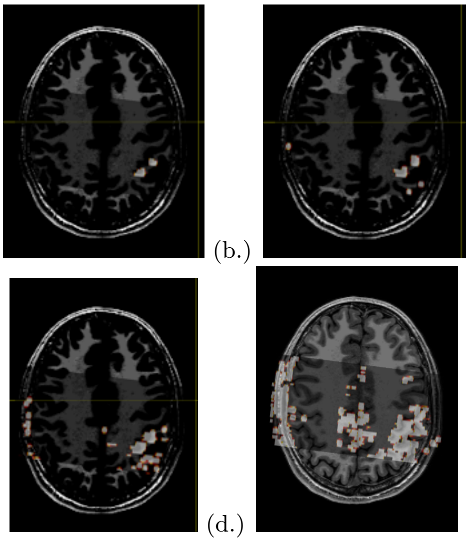

FIG. 7: : The top voxels by R^2: (a.) Top 50 (b.) Top 100 (c.) Top 500 (d.) Top 5000

The top 50 R^2 voxels are clearly located in the brain, and are located near the motor cortex. The top 500 R^2 voxels begin to include some voxels in the skull. With the top 5000 voxels, many skull voxels are included. Brantner found that the classification accuracy began to decrease when including voxels outside of the top 500. See **Conclusions** for discussion of the possible implications of this finding, in light of our visualization.

MAXIMUM MODEL

Based on the findings from our visualization pipeline, (see Figure 6) we designed a new model. Instead of including the entire time series for each voxel, $\{a_i^1(v_k), ..., a_i^T(v_k)\}$, we included only the maximum engagement achieved over that time series. Formally, if we define

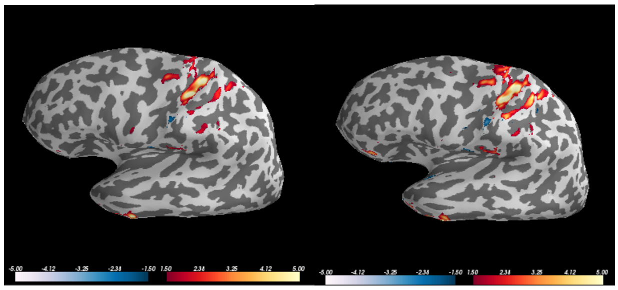

FIG. 8: The top "modelmd" voxels for the unweighted (left) and weighted (right) wrist flex. Both are located in the same region, but have different values. See Figures 1 and 2 for comparison.

$$m_i(v_k) = \max(a_i^1(v_k), ..., a_i^T(v_k))$$

then in this new model

$$x^{(i)} = \begin{bmatrix} m_i(v_1) \\ ... \\ m_i(v_j) \end{bmatrix}$$

Under this model, we were able to achieve higher accuracy for group 3, over 80%. Moreover, we found that we could include a larger number of voxels, and still achieve accurate classification. The other groups showed slight decreases in accuracy under this model

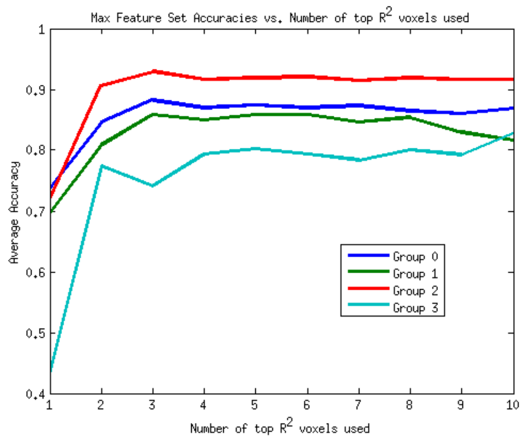

FIG. 9: When we the maximum model, we were able to achieve over 80% accuracy in group 3, even when using large numbers of voxels.

VOXEL SPHERE SEARCH

Our visualization pipeline showed us that the top 50 R^2 voxels and the top 500 R^2 were relatively clustered.

Using voxels from each of these groups gave relatively high accuracy; we considered that we could use every voxel from this cluster. We wrote a script that given a radius (in voxel lengths), and center within the brain, would extract every voxel within that sphere.

Formally, for center c and radius r we let

$$V = \{v : ||v - c||_2 < r\}$$

For our center, we first picked the centroid of our top 50 R^2 voxels, since this appeared to be roughly the center of the motor cortex. We then varied the radius, and found that under this method of voxel selection, we achieved the highest accuracy across all groups with $r = 3$.

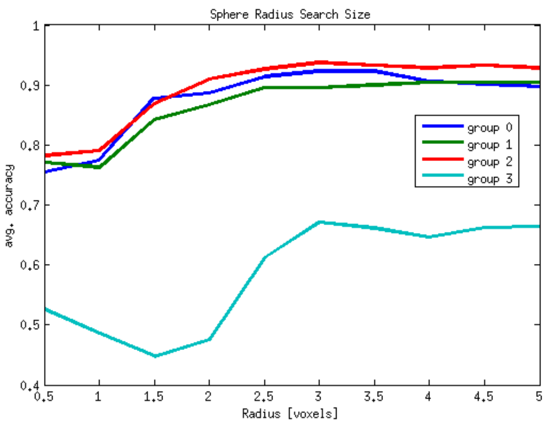

FIG. 10: With our sphere centered at thecentroid of the top 10 R^2 voxelss, we achieved maximum accuracy with a $r = 3$ voxel lengths

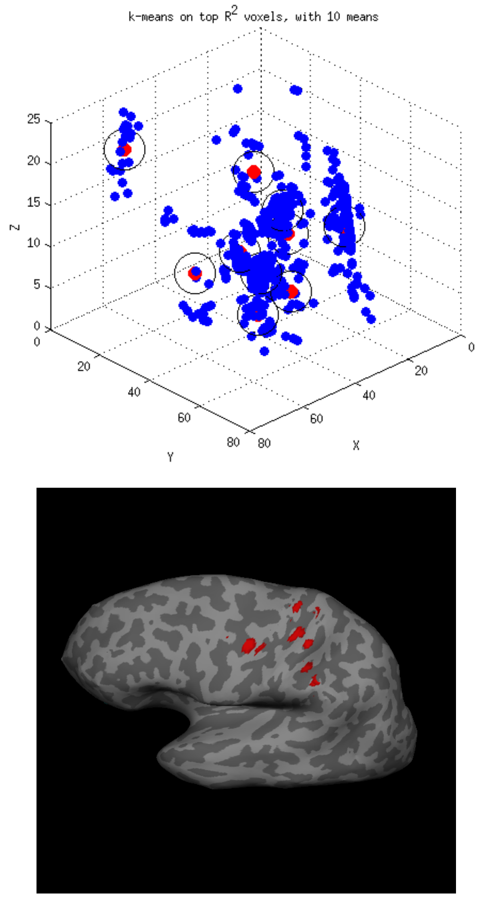

FIG. 11: We created balls around our means within the top 500 R^2 voxels. We used these means as the centroids in our voxel balls. Top: In three-space, Bottom: centroids projected on to the brain

We then explored various methods of maximizing accuracy with respect to the choice of centroid. Our first approach was a grid-search across the voxel space. Due to computational contraints, we were unable to perform an exhaustive grid search with a significantly small step size, so we instead formulated a k-means centroid search on the top voxels sorted by R^2 as a heuristic for good "means" to search in. With this technique, we need only need to train and classify on k spheres. Figure 9 (top) shows the top 500 R^2 voxels in the voxel space, and their respective centroids as estimated by k-means. Figure 9 (bottom) shows which parts of the brain these means correspond to.

For this study, we did not do a rigorous selection for the value of k, but qualitatively verified that the means capture clusters within the motor cortex around $k = 10$. Future analysis may use the elbow-method or some other heuristics for finding the optimal value of k. Overall, the idea of using the K-Means algorithm to find centroids of high R^2 voxels is a good heuristic for optimizing the efforts of the sphere search algorithm because it includes many of the high R^2 voxels, but also includes voxels in their vicinity. If these voxels are part of some cluster in the motor cortex that correlates well to one of these motions, this algorithm may be able to detect and identify that cluster. In this study, however, we were unable to find any correlations between the task group and the cluster that had the highest classification accuracy.

Summary of results

TABLE II: **Time-Series Model:** The classification accuracy for each group using the time-series model by Voxel selection

Group #	Top 50 R^2	Top 500	Top 5000	Ball (r=3)
0	.95	.91	.87	.92
1	.89	.90	.83	.90
2	.93	.95	.84	.95
3	.62	.71	.67	.67

TABLE III: **Max Model:** The classification accuracy for each group using the max model by Voxel selection

Group #	Top 50 R^2	Top 500	Top 5000	Ball (r=3)
0	.82	.89	.74	.83
1	.84	.80	.83	.80
2	.91	.91	.84	.88
3	.74	.79	.67	.60

Conclusions

Perhaps the most significant conclusions from our project is also the simplest; using our visualization pipeline we showed that all of the top 50 - 500 R^2. Moreover, these voxels were around the region classically defined as the motor cortex. Outside of this range, more of the top voxels were located in the skull. Brantner found that using voxels outside of this range decreased classification accuracy; with our visualization, we may be able to attribute this decrease to rising portion of skull-voxels. Brantner's classification accuracy within the 50-500 range is then indicative of the ability to learn the identity of motor tasks from motor cortex data.

Another significant finding was that the max model provided higher classification accuracy for group 3. Including only the maximum BOLD response instead of the entire time series removes information about the shape of the response to a motion and its delay. Since the group 3 comparisons were between tasks of the same motion and and different forces, the fact that we were able to classify more accurately without this information, for group 3 but not the other groups, may imply that it is only the amplitude of the BOLD response, and not the shape or delay that changes with force.

We were unable to draw any strong conclusions from the use of our sphere search method. Our goal was to use this algorithm to find distinct clusters in the brain that provided greater classification accuracy for different tasks and groups. In particular, finding a cluster that gave highest classification accuracy for group 3 than the other three groups could have been strong evidence for the theory that there exists a seperate region in the brain that encodes for force. We were unable to find any evidence of this claim. However, we still believe that this method could be a powerful tool for future researchers, as we discuss in the next section.

Further Research

Several steps could be taken for further, more accurate results, and a greater understanding of the motor cortex. We believe the first step that should be taken in future research is gathering more training data, since had only 116 motions per subject. We assumed that the data would be specific to a subject; in other words, that training an SVM on data from subject a for classification of data from subject b would not be reliable. (Exploring whether or not this is possible to classify accurately with such an SVM may be an interesting topic for future research.) This meant that we were training an SVM on a relatively small amount of data. Scanning the same subjects over multiple sessions would be one way to gather more motions per subject without having to keep subjects in an fMRI scanner for too extensive a period of time.

Had our sphere search been successful, the implications would have been strong. Finding distinct clusters of voxels which classified best for different tasks, could imply that these clusters were control centers for their respective tasks. We believe that our sphere search is a powerful method, because it does not rely on the number of tasks, the type of tasks, or any prior knowledge on the structure of the brain. We expect it will be a useful tool in the future when we use haptic robotics with a much larger task set. Unfortunately, performing the full grid-search over the entire voxel space would have taken an exceedingly long runtime. And even if we could have performed this grid search, the actual regions that encode for motions may be much smaller than our $2.5 \times 2.5 \times 2.5mm^3$ voxels; in fact, a neuron is about $.004mm$ in diameter.

Overall, there is a lot of work to be done to fully understand the motor cortex, but we believe we made significant progress over the course of this project.

Acknowledgements

We would like to thank our advisor Samir Mennon for providing us with data and guidance throughout the quarter.

References

[1] Brantner, G., Menon, S., Schoepp, G., Khatib, O. SVM-based Classification of Motor Tasks on fMRI-Bold Data

[2] Kay, K.N., Rokem, A., Winawer, J., Dougherty, B., Wandell, B.A. GLMdenoise: A fast, automated technique for denoising task-based fMRI data