Identification of Neuroimaging Biomarkers

Dan Goodwin, Tom Bleymaier, Shipra Bhal
Advisor: Dr. Amit Etkin M.D./PhD, Stanford Psychiatry Department

Abstract

We present a supervised learning approach to identify biomarkers for four distinct psychological disorders. We discuss our usage of multi-class classification using SVMs, and experimentation with using a 1-class SVM as a feature reduction technique. Although the RFE method ultimately had limited ability in prediction, it gave good results in identifying neuro-anatomical features of interest.

1. Introduction

Psychiatric disorders are currently diagnosed through behavioral analysis, an outdated mechanism considering the resolution and availability of modern neuroimaging technology. This paper presents a supervised learning approach to studying psychiatric disorders based on functional MRI (fMRI) data, with the intent of improving their diagnosis and treatment.

The intent of this project is to support Dr. Etkin's research on understanding the underlying neuro-anatomical basis of psychiatric disorders. He and his research team have developed a novel mechanism of imparting emotional stimuli into patients and measuring the corresponding brain activation on an fMRI. At the onset of our involvement, Dr. Etkin was able to successfully apply binary classifications, but was in need to a multiclass approach to help take his research in the next step towards clinical practicality. In this paper we present pattern matching data analytic approaches to understand the underlying neural bases of these disorders and develop neurobiological biomarkers which address the similarities and differences between them.

2. The Data

Dr. Etkin provided us with a high quality data set of 81 patients which he had clinically diagnosed himself. The dataset is ideal because the measurements were taken on the same fMRI and the participants had stopped taking their respective medication prior to the scan, giving the clearest signal possible. The training data set had the following components:

- 18 GAD (Generalized Anxiety Disorder)
- 18 Comorbid (combination of GAD and Major Depressive Disorder)
- 16 chronic pain
- 29 control subjects

The fMRI scans were taken while patients performed a task designed to measure emotional reactivity and regulation. Specifically, subjects were shown a sequence of images similar to those in Figure 1, and were asked to identify whether the facial expression was happy or fearful while ignoring the overlying word [5].

Figure 1: Examples of the emotional stimuli imparted on the study participants

Measuring reaction times over varying sequences of congruent and incongruent images leads to a quantitative measure of a subject's implicit emotional regulation, which is the unconscious regulation that occurs during emotional stimuli[3]. Anxiety and depressive disorders are marked by abnormalities in negative emotion processing, and by examining the neural activation during this test we can find look for the different (as well as similar) activation patterns across the given subject groups.

The 3D fMRI brain images were converted to matrix form using a program called "Statistical Parametric Mapping 5." SPM-5 normalizes the voxels into standard space, and smoothes the 3D brain images to account for unavoidable motion of the patient. As a result, for each patient we obtain a vector of 169,301 voxels. These voxels represent the areas of the brain where cell bodies are present (gray matter), shown in red in the images below.

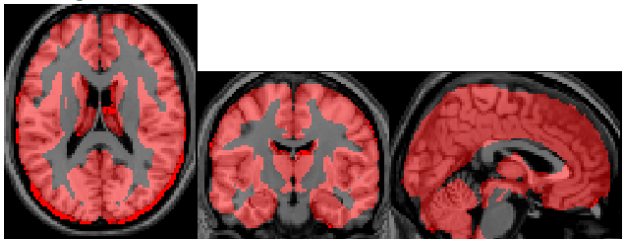
Figure 2: 3D model of the 160k voxels initially of interest

The resulting data set is an 81 x 169,301 matrix where rows represent patients and columns represent feature voxels. To account for scaling differences among patients and groups we normalized the data within each row and then across all columns. The overall dataflow in our system is mapped below in in Figure 3.

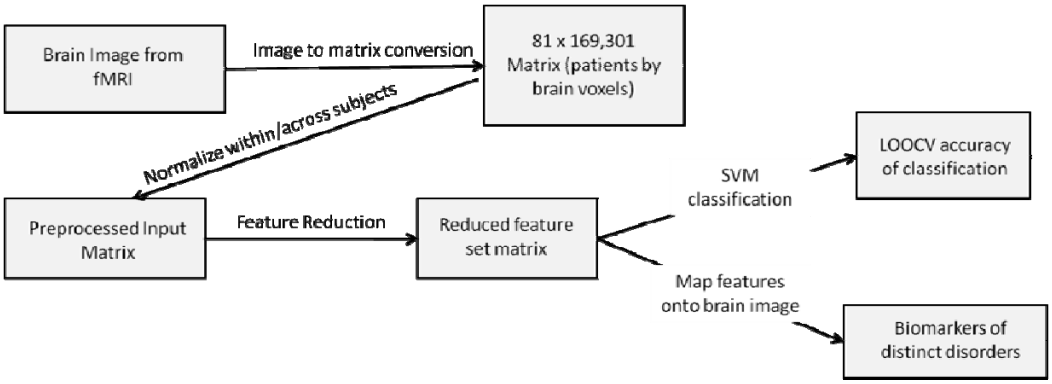

Figure 3: Flowchart for data processing scheme

3. Recursive Feature Elimination (RFE) based on one class SVM

As the data has significantly more features than training samples, reducing the features set is an essential step to producing a robust classifier. In binary classification, a standard methodology for reducing feature size is to train a support vector machine on the given data and recursively remove the features with the smallest weights of the primary Lagrangian. However, our literature survey showed that there is not a clear optimal method to feature reduction in multiclass scenarios.

Most multiclass RFE solutions are a direct extension of the binary problem. Concerned by poor results in various binary classifications, we instead experimented with a less common method of a single-class SVM. The single-class SVM is used to find information about a distribution, and gives a similar set of weights to the binary SVM. However, instead of determining a hyperplane that separates two datasets, the results can be used to differentiate signal voxels from noise. Simulations with synthetic data demonstrated a robust behavior in low signal-to-noise ratio environment (100% identification of markers in a 1.1 SNR) , indicating potential for a method of independently refining the feature set for each category of patient.

Once the largest weighted voxels are identified for each class, they are then recombined into a total region of interest. Specifically, the regions of the brain which are activated by two or more psychiatric conditions are most valuable - not only on anatomical level, but also to identify the minimum number of voxels that can identify all four classes. We therefore experimented with three combination schemes: a union of each class's feature set, an intersection of each class's feature set, and the voxels that have at least two classes in common. To verify these results, these new feature sets were re-formed back into 3-D brain space in order to see if the activated areas of the brain were as expected.

4. Multi-class Classification

Using the libsvm Matlab library, we implemented the One Vs. One approach for 4-class SVM classification [1]. This method finds an SVM classifier for each possible pair of the four groups which results in 6 classifiers. Then, we use the max-wins strategy to classify a test patient into one of the four categories. For each of the 6 decision functions, we count how many times test point x is classified into each class, and assign x to the class with the maximum number of votes.

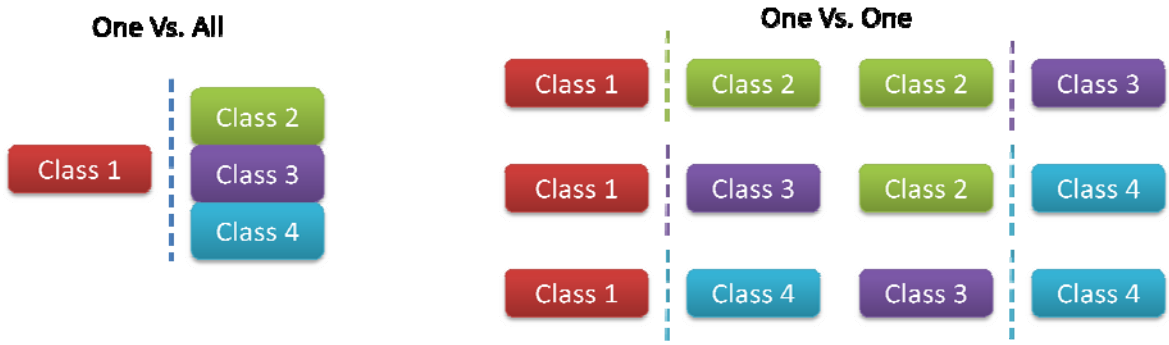

Figure 4: Brief comparison of the two multi-class SVM approaches. Note that, in this example, the One Vs All approach will require three more SVM comparisons

To train the multi-class classifier we randomly selected 16 patients from each of the four groups. This was done to avoid using unbalanced training data and bias the classification algorithm. The smallest

group was pain, at 16 patients, which led to the selection of 16 from each group. Initial studies included the test-sample for the feature selection. Leave One Out Cross Validation (LOOCV) with the linear kernel on the reduced feature set yielded the best results at 98.4% accuracy, while all the other kernels returned a steady 0% accuracy. However, when we removed the test-data point from the RFE calculation, accuracy of the system plummeted below 40%.

It was an initial surprise to see that one out of 81 patients could have such an impact on the SVM training data. To analyze this result, we explored the difference in the feature sets of the GAD category. We first pared the initial 169,301 voxels down to a set of 3000 using all eighteen samples ("the golden set"). Then, we ran identical RFEs on the seventeen of the GAD patients, leaving out a different patient each time. We calculated the percentage of overlap between the golden set and the set with i^{th} element removed. In Figure 5, we see that the change of one patient changes the entire feature set by 50%, which indicates that the variance inherent in the single-class SVM will prevent it from being a stand-alone solution for feature reduction in an online predictive system.

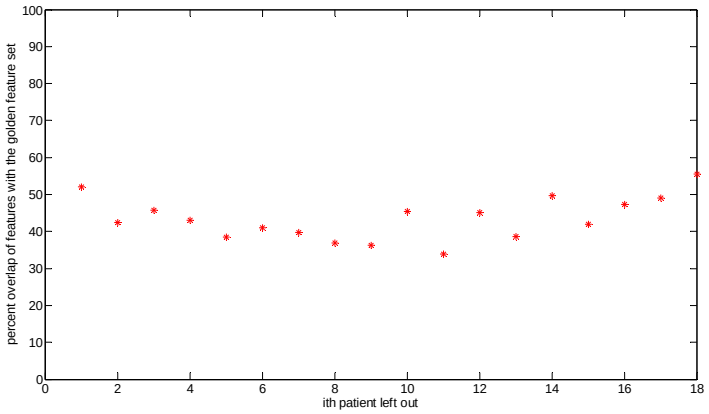

Figure 5: Impact of one missing training sample in the calcuation of the RFE

As a final test to confirm the over-fitting behavior, we randomly assigned our 81 patients a class, then ran the "bogus set" through the 1-Class RFE(using the test point as well) and then tested the accuracy of the system through a One Vs. One SVM. Rather unsurprisingly, we obtained 98.4% accuracy in categorizing the patients to their "bogus class," despite the absence of any actual underlying Given these results, it would not be appropriate to claim high accuracy for the multi-class classification. However what came out of the 1-class SVM, are the biomarkers for the various disorders.

5. Identification of BioMarkers

While the variance in our RFE methodology prevented our algorithm from being a successful predictor, the system showed excellent results in identifying regions of the brain that are activated during the emotional stimuli. Such information has value in designing future feature reduction methods, as the general areas are now known. Figure 6 shows two example of feature reduction done on each class, then plotted together to show comparative areas of activation.

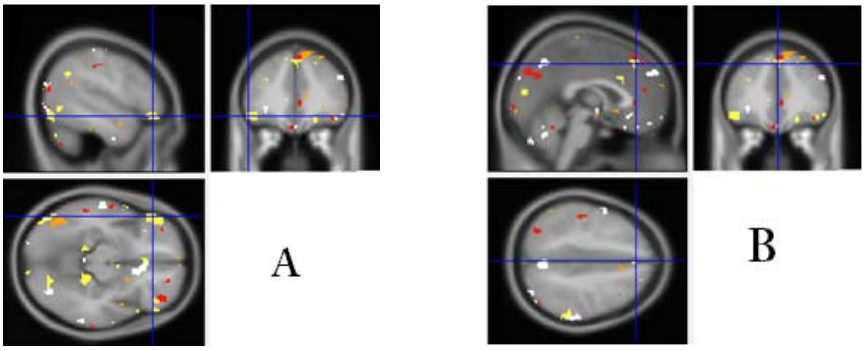

Figure 6: Cluster of activation in the Inferior Frontal Gyrus in the Control Group (b) Area of common activation in the Medial Frontal Gyrus

6. Summary

Ultimately, our system identified the following neuroanatomical features that were activated in patients of different psychiatric conditions.

Area	GAD	CMBD	Ctrl	Pain	Function
Par. Sub L		X			Somatic Perception
Sup. F. Gyrus		X			Self Awareness
Med. F Gyrus	X				High level executive function
Cingulate Gyrus	X		X	X	Emotion formation & processing
Sub Gyral	X				Responds to visual sadness stimuli
PH Gyrus		X			Memory encoding, retrieval
Inf. Parietal		X	X	X	Visual processing
P.C. Gyrus	X		X	X	Touch sensation

Acknowledgements

We would like to acknowledge and thank Dr. Amit Etkin and Dr. Fumiko Hoeft of the Department of Psychiatry at Stanford. They provided us with the data set of 81 patients and shared with us the knowledge and computational progress they had previously made in this field of study.

References

1. C. C. Chang and C. J. Lin (2009) LIBSVM: a Library for Support Vector Machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm/
2. F. De Martino, et al (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage*, v. 43, pp. 44-58, www.elsevier.com/locate/ynimg.
3. A. Etkin, et al (2006) Resolving emotional conflict: A role for the Rostral Anterior Cingulate Cortex in Modulating Activity in the Amygdala. *Neuron*, v. 51, pp. 871-882.
4. A. Etkin, et al (2009) Failure of anterior cingulated activation and connectivity with the amygdala during implicit regulation of emotional processing in generalized anxiety disorder. *Human Brain Mapping annual conference, San Francisco, CA, 6/19/09.*
5. A. Schatzberg, A. Etkin, A. Reiss, M. Greicius (2009) Identification of neuroimaging biomarkers during emotional processing and regulation in anxiety and depressive disorders. *Unpublished Challenge Grant*.
6. H. Zou and T. Hastie (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, v. 67, pp. 301-320.