Discriminative Analysis for Image-Based Studies

Polina Golland[1], Bruce Fischl[2], Mona Spiridon[3],
Nancy Kanwisher[3], Randy L. Buckner[4], Martha E. Shenton[5],
Ron Kikinis[6], Anders Dale[2], and W. Eric L. Grimson[1]

[1] Artificial Intelligence Laboratory,
Massachusetts Institute of Technology, Cambridge, MA.
[2] Athinoula A. Martinos Center for Biomedical Imaging,
Massachusetts General Hospital, Harvard Medical School, Boston, MA.
[3] Department of Brain and Cognitive Sciences,
Massachusetts Institute of Technology, Cambridge, MA.
[4] Departments of Psychology, Anatomy and Neurobiology, and Radiology,
Washington University, and Howard Hughes Medical Institute, St. Louis, MO.
[5] Laboratory of Neuroscience, Clinical Neuroscience Division, Department of
Psychiatry, VAMC-Brockton, Harvard Medical School, Brockton, MA.
[6] Surgical Planning Laboratory, Brigham and Women's Hospital,
Harvard Medical School, Boston, MA.

Abstract. In this paper, we present a methodology for performing statistical analysis for image-based studies of differences between populations and describe our experience applying the technique in several different population comparison experiments. Unlike traditional analysis tools, we consider all features simultaneously, thus accounting for potential correlations between the features. The result of the analysis is a classifier function that can be used for labeling new examples and a map over the original features indicating the degree to which each feature participates in estimating the label for any given example. Our experiments include shape analysis of subcortical structures in schizophrenia, cortical thinning in healthy aging and Alzheimer's disease and comparisons of fMRI activations in response to different visual stimuli.

1 Introduction

Statistical studies of neuroanatomy in different populations are important in understanding anatomical and neurophysiological effects of diseases when comparing patients *vs.* normal controls, or biological processes, for example, comparing different age groups. High dimensionality of the feature space and limited number of examples present a significant challenge for statistical analysis tools. Two different approaches are typically used to overcome this difficulty. The first is to simplify the feature space by using global measurements, such as volume of a structure or average thickness of an anatomical area of the cortex [6, 9, 10, 14]. However, this does not provide detailed information on the type of differences and their location. Another commonly used solution is to analyze each feature separately [2, 11, 12, 18], thus ignoring possible correlations in the feature values.

Using more powerful analysis techniques can potentially improve our understanding of detected differences between populations, as well as identify possible dependencies in the features. In this paper, we demonstrate experimental results of applying discriminative analysis to several different statistical studies of neuroanatomy and function. In each study, the image-based features were chosen based on the question of interest. While selecting representations for statistical analysis is an important problem, it is outside the scope of this paper. Here, we focus on the analysis techniques applicable once the features are extracted from the images.

2 Discriminative Analysis

This section provides a brief overview of discriminative modeling for population comparison (for details, see [8]). We start by training a classifier for labeling new, unseen, inputs into one of two example groups. We then extract an explicit representation for the differences between the two groups captured by the classifier function. This approach is based on the premise that in order to automatically detect statistical differences between two populations, one should try to build the best possible classifier for labeling new examples. In the optimal case, the classifier function will exactly represent the important differences between the two classes, while ignoring the intra-class variability.

Training We use the Support Vector Machines (SVMs) learning algorithm [17] to arrive at a classification function. In additional to the sound theoretical foundations of SVMs, they have been demonstrated empirically to be quite robust and seemingly free of the over-fitting problems to which other learning algorithms, such as neural networks, are subject.

Given a training set of l pairs $\{(\mathbf{x}_k, y_k)\}_1^l$, where $\mathbf{x}_k \in \mathbb{R}^n$ are observations and $y_k \in \{-1, 1\}$ are corresponding labels, and a kernel function $K : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$, the SVM learning algorithm produces a classification function

$$f_K(\mathbf{x}) = \sum_{k=1}^{l} \alpha_k y_k K(\mathbf{x}, \mathbf{x}_k) + b,$$

where the coefficients α_k's and b are chosen to maximize the margin between the two example classes. Training vectors with non-zero α's are called support vectors, as they define, or "support", the separating boundary. In the simplest case of the linear kernel, $K(\mathbf{u}, \mathbf{v}) = \langle \mathbf{u} \cdot \mathbf{v} \rangle$, the separating boundary is a hyperplane whose normal is $\mathbf{w} = \sum_k \alpha_k y_k \mathbf{x}_k$. For non-linear classification, we employ the commonly used Gaussian Radial Basis Function (RBF) kernel family $K(\mathbf{u}, \mathbf{v}) = e^{-\|\mathbf{u}-\mathbf{v}\|^2/\gamma}$ (parameter γ determines the width of the kernel). One of the important properties of this family of classifiers is its locality: moving a support vector slightly affects the separation boundary close to the vector, but does not change it in regions distant from the vector.

We use leave-one-out cross-validation to estimate the expected accuracy of the resulting classifier and, in the non-linear case, to select the optimal set of parameters (e.g., the width γ for the Gaussian RBF kernels).

Classifier Interpretation We use the previously introduced *discriminative direction* [7, 8] to explicitly represent the spatial pattern of differences between populations implicitly captured by the classifier function. Intuitively, the discriminative direction defines the optimal direction of change that makes an input vector look to the classifier more like the examples from the other class while introducing as little irrelevant change as possible. It is easy to see that for the linear classifier, the discriminative direction corresponds to the normal to the separating hyperplane $\mathbf{w}$, which is also the gradient of the classifier function. It can be shown that moving along the gradient of the classifier function minimizes irrelevant changes for the RBF kernels as well. More precisely, the classifier gradient $\nabla f_K(\mathbf{x})$ defines the discriminative direction for the class with label -1, while the discriminative direction for the class with label 1 is $-\nabla f_K(\mathbf{x})$. The gradient of the linear classifier function is the same at every point in the input space, but it varies spatially for non-linear classifiers. This suggests that the most appropriate points for evaluation of the discriminative direction for the non-linear case are the points close to the separating boundary, i.e., the support vectors.

To summarize the steps of the analysis, given a training set of feature vectors and their class labels, we train a classifier (linear and/or RBF), estimate its cross-validation accuracy and compute its gradient, which can be directly visualized in the original input space.

3 Experimental Studies

In this section, we present the results of the analysis for several statistical studies. For each experiment, we describe the original data and the feature extraction procedure, report statistical analysis results and show the discriminative direction identified by the technique. The cross-validation accuracy is reported with the corresponding 95% confidence interval. The online version of the paper (in www.ai.mit.edu/people/polina) contains color images.

Hippocampus-Amygdala Complex in Schizophrenia. This study contains MRI scans of 15 schizophrenia patients and 15 matched controls [14]. The hippocampus-amygdala complex was manually segmented. We used volumetric signed distance transforms as feature vectors defining the shape of the structure, aligning the shapes by bringing them into a "canonical" pose defined by the first and the second order moments. Since the volumetric discriminative direction is difficult to interpret, we project it onto the space of possible deformations of the shape boundary and visualize the resulting surface deformation instead [8].

The linear classifier performance was very close to the 50% baseline, while the best RBF classifier achieved 76.7%($\pm$15.1%) and 70.0%($\pm$16.3%) accuracy for the right and the left hippocampus respectively. Thus, we expect the non-linear classifier to capture relevant structure in the data that allows it to achieve above the baseline performance. In order to visualize the differences represented by the classifier, we compute the discriminative direction for the support vectors.
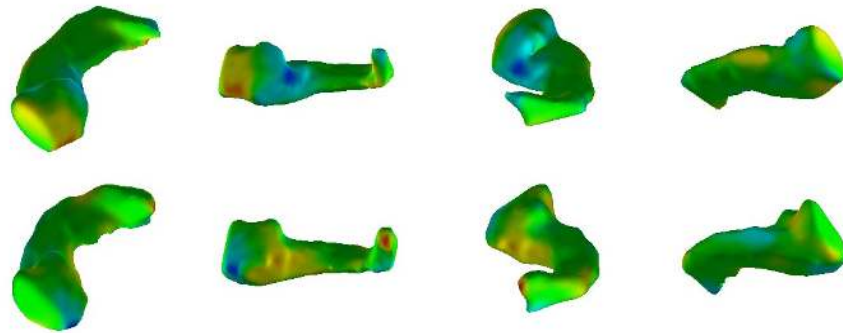
(a) Right hippocampus, the first two support vectors.
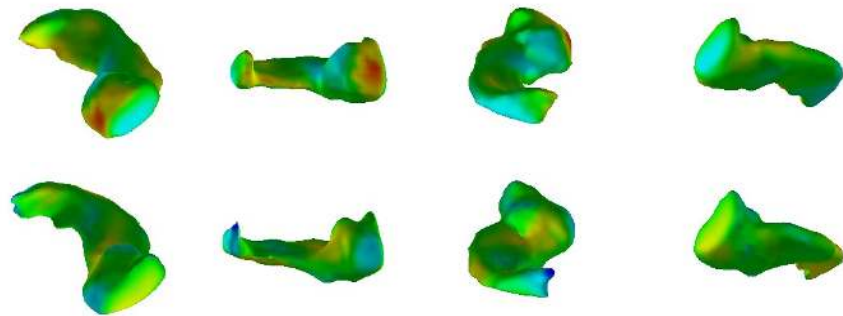
(b) Left hippocampus, the first two support vectors.
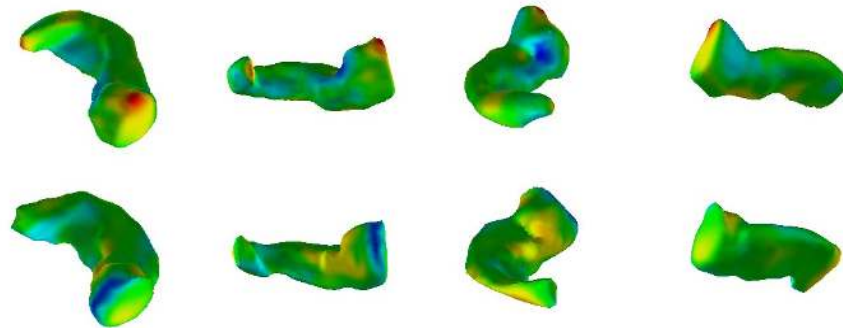
(c) Left hippocampus, the next two support vectors.

Fig. 1. Example support vectors for the hippocampus-amygdala study, shown in pairs of a normal control (top) and a schizophrenia patient (bottom). Four views of each shape are shown (one row per shape). The color indicates the amount of deformation, from blue (moving inwards) to red (moving outwards).

Fig. 1a shows the top support vector from each group with the discriminative direction detected by the algorithm for the right hippocampus. In many experiments, we find that even for the non-linear classifiers, different support vectors often have (visually) similar discriminative direction. In this study, the first four support vectors in each group correspond to essentially identical discriminative direction. Moreover, the support vectors from the two different classes often define deformations of very similar nature, but of opposite signs. Fig. 1(b,c) show the top two support vectors from each group with the discriminative direction for the left hippocampus. Note the deformation in the anterior part of the structure, of a similar nature but localized in different parts of the bulbous head of the amygdala in the two pairs of examples shown. Besides the obvious explanation that the location of this deformation is not fixed in the population, it could also be caused by our method's sensitivity to alignment. Misalignments of the structures in the feature extraction step can cause such size differences to be detected in different areas of the anatomical structure. More powerful alignment technique can potentially help resolve this problem [16].

This example demonstrates the amount of detail in the description of the shape differences between the two populations that can be detected by our technique. This information can help guide the exploration of the relationship between the changes in the subcortical structures and the symptomatic information on the disease (e.g., memory functions and their degradation in the hippocampus study). If a better shape representation is suggested, it can be directly used in conjunction with the presented statistical analysis framework.

Cortical Thinning in Healthy Aging and Alzheimer Type Dementia.
In this study, we compared the thickness of the cortex in 31 young controls, 38 old controls and 37 patients diagnosed with dementia of the Alzheimer type (DAT) [13]. The gray/white matter interface and the pial surface were automatically segmented from each MRI scan [1, 5], followed by a registration step that brought the surfaces into correspondence by mapping them onto a unit sphere while minimizing distortions and then non-rigidly aligning the cortical folding patterns [3, 4]. The cortical thickness was densely sampled at the corresponding locations for all subjects.

The performance of the linear classifier was virtually identical to that of the RBF classifier in both comparisons. Consequently, we only show the gradient of the linear classifier (i.e., the discriminative direction for the second class). The cross-validation accuracy was $98.4\%(\pm 3.1\%)$ for the aging study (young *vs.* old controls) and $77.3\%(\pm 9.5\%)$ for the dementia study (old controls *vs.* DAT patients). Fig. 2 shows the discriminative direction for the two studies. The images show both hemispheres inflated so that the entire cortical surface is visible in the rendering. Grayscale is used to display the sulcal pattern, while color is used to show the differences in the cortical thickness. We can see that the two patterns are significantly different, suggesting that the effects of the Alzheimer's disease on the brain are distinct from those of healthy aging. While the aging pattern is aligned with major sulci, the differences in dementia patients are more localized and confined to a few areas on the cortex. Such analysis could
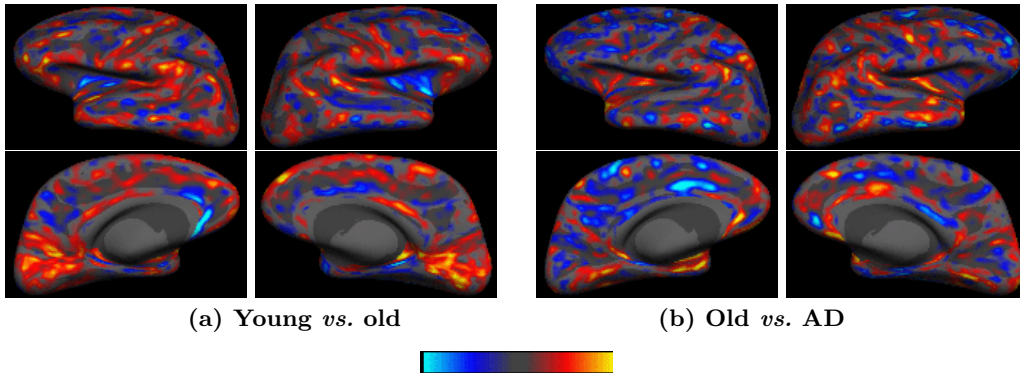
(a) Young *vs.* old　　　　　**(b) Old *vs.* AD**

Fig. 2. Discriminative direction maps for cortical thickness studies. Two views are shown for each hemisphere: lateral (top) and medial (bottom). The color is used to indicate the weight of each voxel, from light blue (negative) to yellow (positive).

potentially be useful for investigating various hypotheses on the development of the disease and its affects on brain structure.

Categorical Differences in fMRI Activation Patterns. This experiment is significantly different from the others in this paper, as it compares the patterns of fMRI activations in response to different visual stimuli in a *single* subject. We present the results of comparing activations in response to face images to those induced by house images, as these categories are believed to have special representation in the cortex [2, 11]. The comparison used 15 example activations for each category (for details on data acquisition, see [15]). The fMRI scans were aligned to the structural MRI of the same subject using rigid registration. For each scan, the cortical surface was extracted using the same techniques as in the study of cortical thickness, and the average activation values were sampled along the cortical surface. A surface-based representation models the connectivity of the cortex and is therefore well suited for fMRI activation studies.

Linear classification achieves $96.3\%(\pm6.8\%)$ cross-validation accuracy for this experiment. Fig. 3a shows the discriminative direction for the linear classifier for the face class. The fMRI images did not cover the entire cortical surface, leaving out the frontal area. We can see several localized areas that were detected as highly predictive of the stimulus category, in general confirming previous findings based on a point-wise t-test. Investigating the differences between the pattern in Fig. 3a and the t-test map will help us understand the spatial correlations between different regions of the cortex.

An additional advantage of using the discriminative approach is that it becomes possible to evaluate how predictive of the category certain sub-regions of the brain are without sacrificing the detailed representation. Fig. 3b shows the resulting discriminative direction if only the "visually active" region of the cortex is considered. The mask for the visually active voxels was obtained using a separate visual task. Note that in general, the method produces a map that is
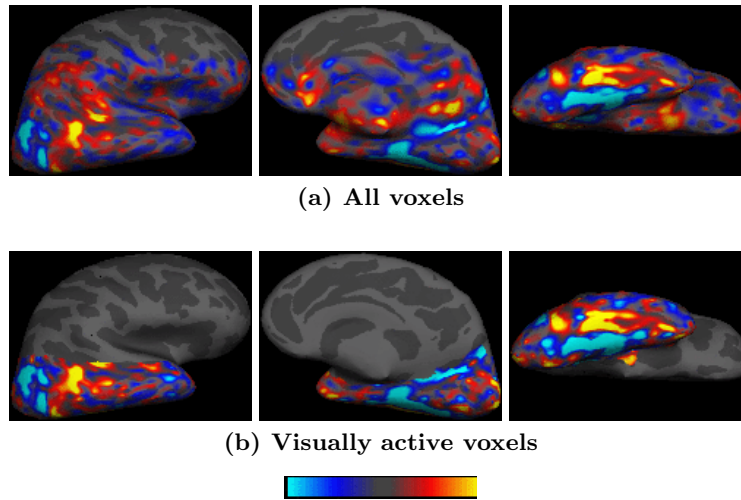
(a) All voxels

(b) Visually active voxels

Fig. 3. Discriminative direction map for the face class in comparison with the house class. Three views of the right hemisphere are shown: lateral (left), medial (center) and inferior (right). The color is used to indicate the weight of each voxel, from light blue (negative) to yellow (positive).

very similar to the corresponding subset of the one in Fig. 3a, indicating robustness of the estimation. Interestingly, the cross-validation accuracy increased to 100% (i.e. the classes are completely separable) in this case, suggesting that the rest of the surface had a lot of noisy voxels making the learning task harder.

4   Conclusions

We presented experimental results for several substantially different image-based studies using the same statistical analysis framework based on discriminative modeling, i.e., training a classifier to label new examples based on the structure in the training data set. The analysis considers all features simultaneously, thus accounting for possible dependencies among the features. By estimating the expected error of the classifier we can effectively assess how separable the populations are with respect to the chosen set of features. Furthermore, visualizing the gradient of the classifier function provides us with detailed information on how predictive individual features and groups of features are of the class label.

Our experience with the technique has suggested several interesting questions to be explored next. Understanding the effects of variability within each population on the detected differences will allow us to provide a better interpretation of the estimated classifier function in terms of the true differences between the populations. Another important problem is assessing significance of the obtained results, especially in light of high dimensionality of the data and small number of samples. Answering these questions will lead towards improving the accuracy and the reliability of the resulting hypotheses of differences between populations.

References

1. A. M. Dale, *et al*. Cortical Surface-Based Analysis I: Segmentation and Surface Reconstruction. *NeuroImage*, 9:179-194, 1999.
2. R. Epstein and N. Kanwisher. A cortical representation of the local visual environment, *Nature* 392:598-601, 1998.
3. B. Fischl, *et al*. Cortical Surface-Based Analysis II: Inflation, Flattening, a Surface-Based Coordinate System. *NeuroImage*, 9:195-207, 1999.
4. B. Fischl, *et al*. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8:272-84, 1999.
5. B. Fischl, *et al*. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *In PNAS*, 26:11050-5, 2000.
6. G. Gerig, *et al*. Shape versus Size: Improved Understanding of the Morphology of Brain Structures. *In Proc. MICCAI'01*, LNCS 2208, 24-32, 2001.
7. P. Golland, *et al*. Deformation Analysis for Shaped Based Classification. *In Proc. IPMI'01*, LNCS 2082, 517-530, 2001.
8. P. Golland. Statistical Shape Analysis of Anatomical Structures. PhD Thesis, MIT, August 2001.
9. C. Good, *et al*. A voxel-based morphometric study of aging in 465 normal adult human brains. *Neuroimage*, 14:21-36, 2001.
10. T. L. Jernigan, *et al*. Effects of age on tissues and regions of the cerebrum and cerebellum. *Neurobiol. Aging*, 22:581-94, 2001.
11. N. Kanwisher, *et al*. The Fusiform Face Area: a module in human extrastriate cortex specialized for face perception. *J. Neurociences* 17:4302-4311, 1997.
12. D. J. Pettey and J. C. Gee. Using a Linear Diagnostic Function and Non-rigid Registration to Search for Morphological Differences Between Populations: An Example Involving the Male and Female Corpus Callosum. *In Proc. IPMI'01*, LNCS 2082, 372-379, 2001.
13. D. H. Salat, *et al*. Early and widespread thinning of the cerebral cortex with normal aging. *Submitted*, 2002.
14. M. E. Shenton, *et al*. Abnormalities in the left temporal lobe and thought disorder in schizophrenia: A quantitative magnetic resonance imaging study. *New England J. Medicine*, 327:604-612, 1992.
15. M. Spiridon and N. Kanwisher. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Submitted*, 2002.
16. S. J. Timoner, *et al*. Performance Issues in Shape Classification, *To Appear in Proc. MICCAI 2002.*, September 2002.
17. V. N. Vapnik. Statistical Learning Theory. *John Wiley & Sons*, 1998.
18. P. Yushkevich, *et al*. Intuitive, Localized Analysis of Shape Variability. *In Proc. IPMI'01*, LNCS 2082, 402-408, 2001.