Using neural network algorithms to investigate distributed patterns of brain activity in fMRI.

Sean M Polyn, Leigh E Nystrom, Kenneth A Norman, James V Haxby, M Ida Gobbini & Jonathan D Cohen

Center for the Study of Brain, Mind & Behavior; Dept. of Psychology, Princeton University, Princeton NJ, USA

Poster presented at OHBM Conference, Budapest, Hungary 2004

Introduction

Many recent papers have focused on the analysis and classification of distributed patterns in neuroimaging data (Haxby et al., 2001; Cox & Savoy, 2003; Carlson et al., 2003; Hanson et al., in press).

Backpropagation networks are a powerful nonlinear classification algorithm. Here, we show that they can be used to classify subtle perceptual categories with high accuracy. However, in our dataset the advantage of these networks over a simple linear classifier is small.

We also explore techniques for reading out which voxels are important for a given category. We show that some techniques that have been proposed (e.g. by Hanson et al.) may yield false positives.

Respresentative pictures from each of the seven categories.

Experimental design

Subjects performed a simple task in which they had to detect one-back stimulus repeats in a stream of pictures all drawn from the same category; repeats were presented in a different orientation from the original stimulus. The design closely matches that of Haxby et al. (2001), but with more subtle categories.

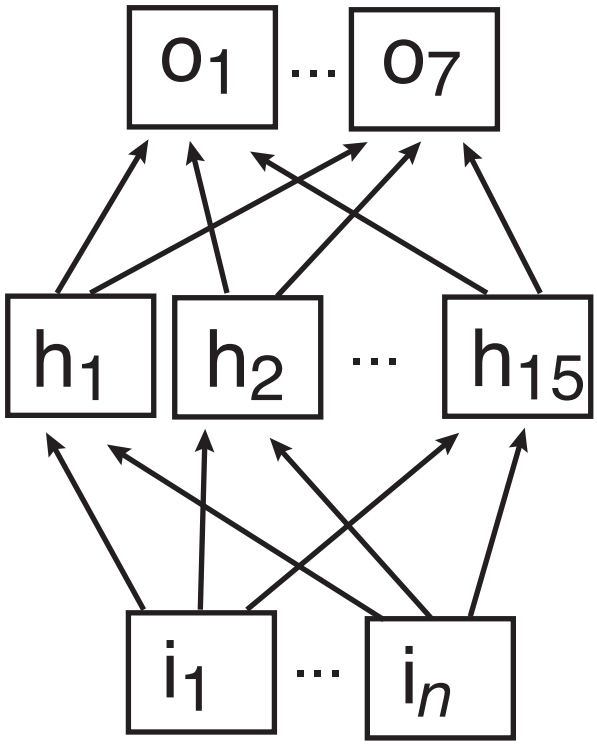
Imaging methods

Data was acquired from a Siemens Allegra 3T: TR = 2.5 s; TE = 34 ms; flip angle = 80; 2.2mm x 2.2mm inplane resolution; 32 axial slices, 2mm thick with 0.2 mm gap; 153 volumes per run, 8 runs.

Analysis techniques

Backpropagation

Backprop networks: Each input unit corresponds to a voxel in the brain volume. Training patterns are presented and weights in the network are altered to reduce classification errors.



The networks are tested for generalization on patterns that were not presented during training.

Backprop specs: 2 & 3 layer networks. Cross-entropy error function. Trained with conjugate gradient backprop with Powell-Beale restarts. 3 layer nets had 15 hidden units (hyperbolic tangent transfer function). Output layer used sigmoidal transfer function. Matlab neural network toolbox was used for all simulations.

Analysis techniques

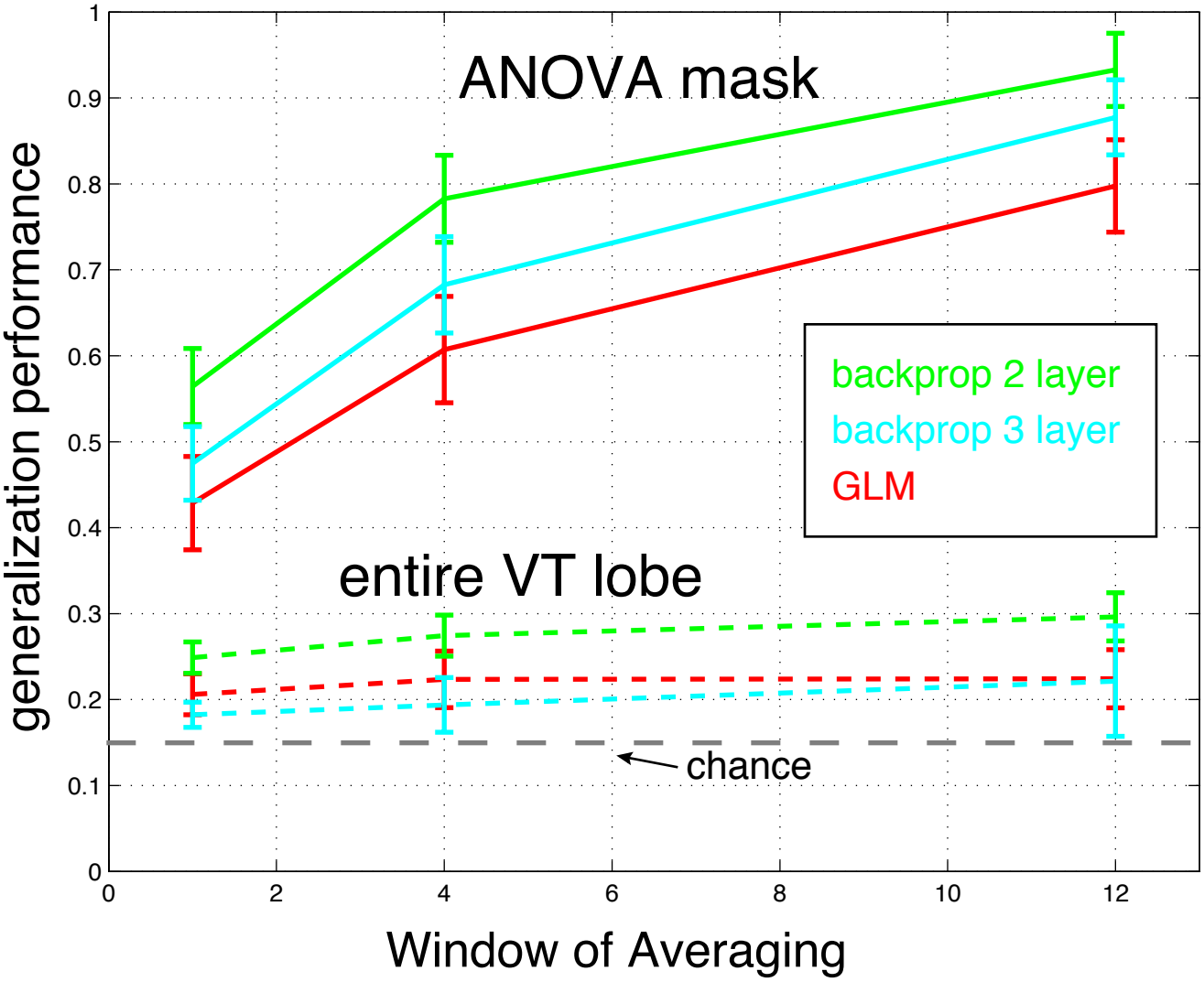
General Linear Model (GLM)

A linear regression is run on the training patterns; for each category we get a map of beta weights (one beta per voxel).

We perform a dot product of each test pattern with each of the beta maps. The category producing the largest dot product is our guess at the category membership.

This method is similar in spirit to the correlation analysis carried out in Haxby et al. (2001).

Comparison of classification algorithms



Generalization performance across subjects. Performance is helped in all cases by the use of an ANOVA: Voxels that do not vary across categories are removed. Two layer backprop is best, but all perform well.

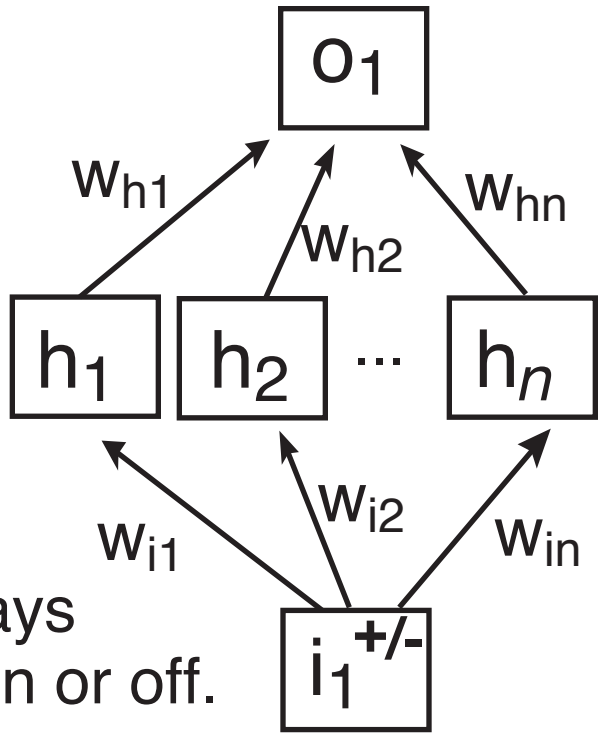
Analysis of classifiers: Which voxels are important for representing a given category?

Weight-based analysis

Take the dot product of the w_i's with the w_h's, and multiply that by the average input value for the category.

This effectively gives the 'path strength' from voxel i_1 to o_1.

The sign of the path strength says whether voxel i turns output o on or off.



Noise-based analysis

Described in Hanson et al. (in press):

Replace signal in a voxel with large noise source.

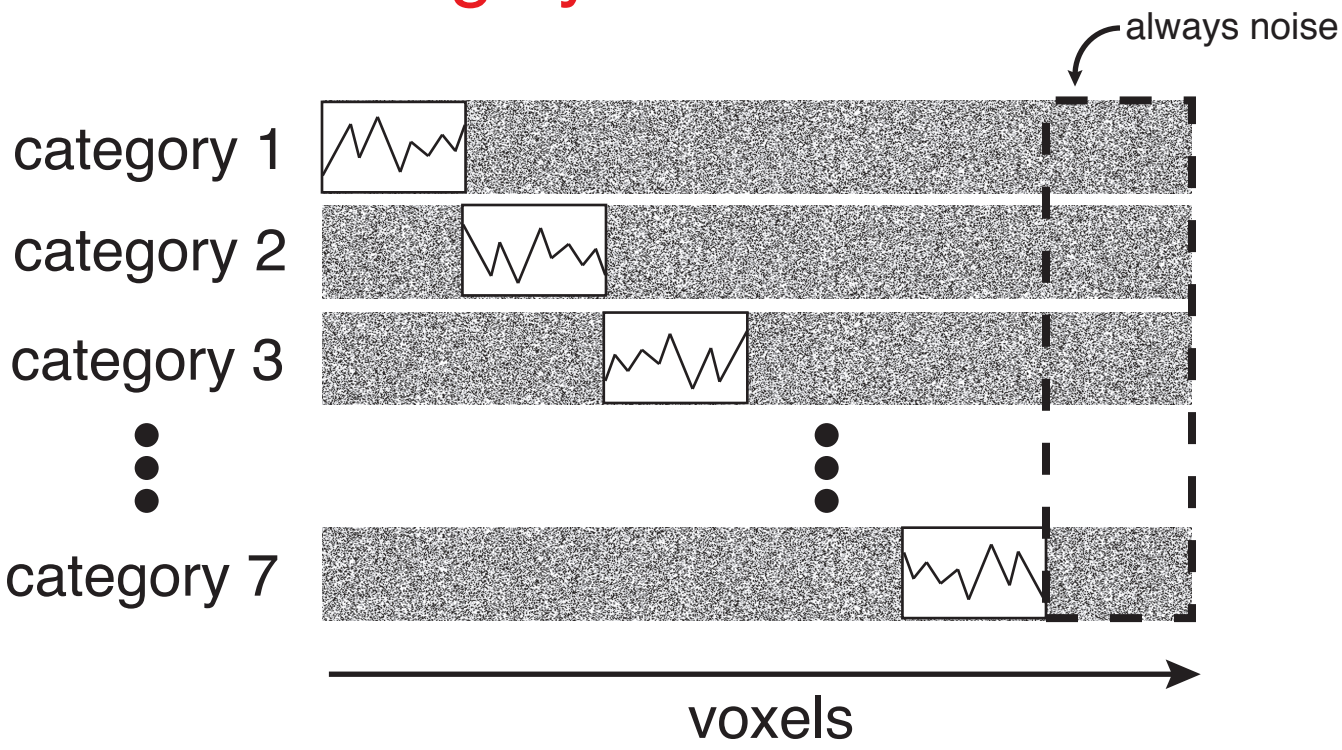
Test network on these noised patterns. If errors for a given category exceed some threshold, the voxel is sensitive to that category.

Conclusion: "substantially all of the same VT lobe voxels contribute to the classification of *all* object categories [...]".

GLM-based analysis

Use the absolute value of the beta weights to index the importance of each voxel.

Creating synthetic data



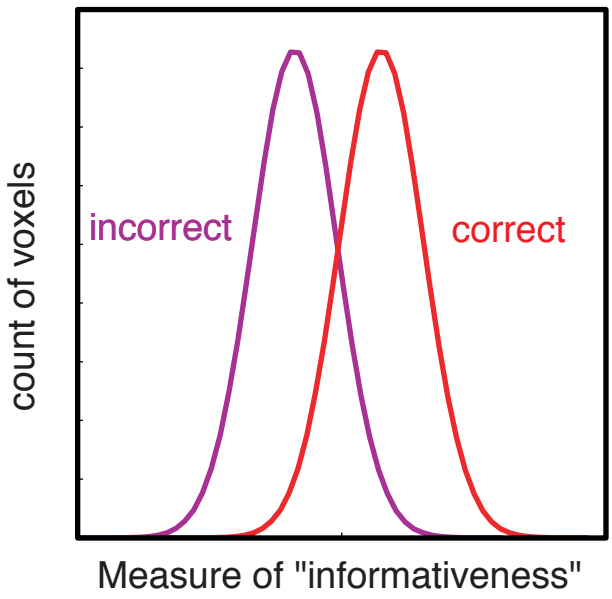
To test the analysis methods we create a synthetic dataset in which the category representations are entirely non-overlapping.

Each analysis method assigns a set of "informativeness" scores to each voxel, indicating how sensitive the voxel is to each category. Since we know which voxels contain meaningful signal, we can quantify the reliability of each method for this dataset.

Measuring sensitivity

There will be some distribution of scores for the units that actually code for a category, as well as a distribution of scores for the units that do not.

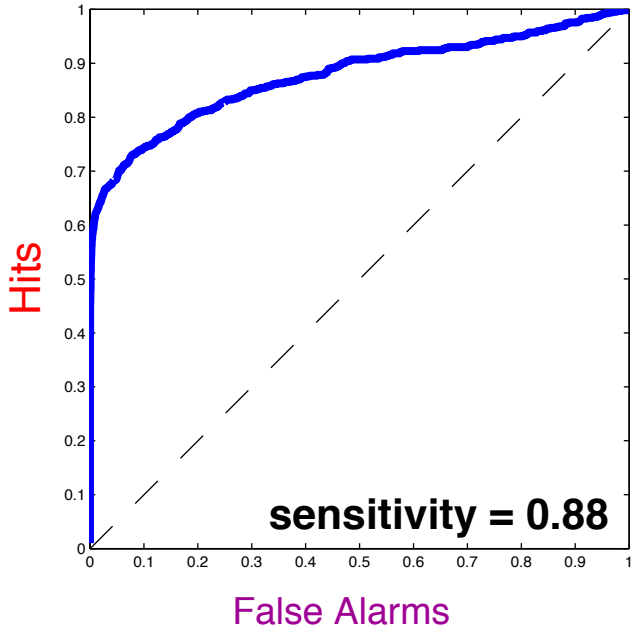
hypothetical distributions



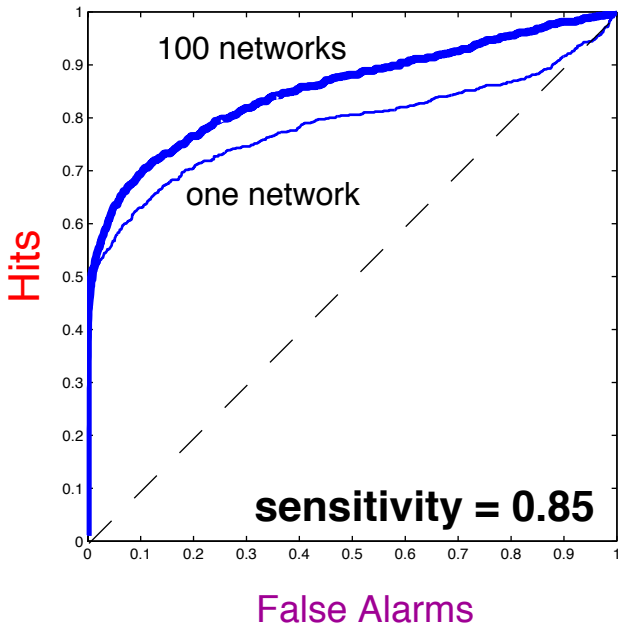
We can set a threshold value on the "informativeness" score and determine the number of hits as well as false alarms that the analysis method makes. By sweeping the threshold across the range of scores, we can create an ROC curve; this gives us a quantitative measure of the sensitivity of the method.

Comparing the analysis methods

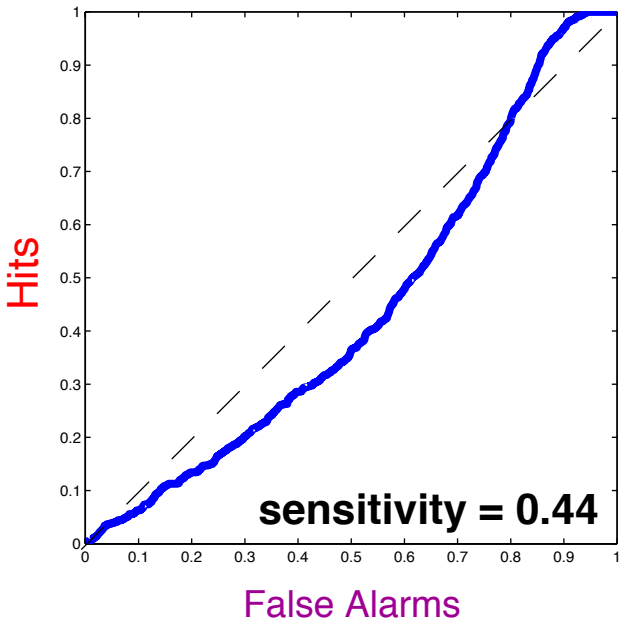
GLM-based analysis



Weight-based analysis



Noise-based analysis



The ROC curves tell us how well each analysis method captures the category membership of each voxel. A larger area under the curve means greater sensitivity to category membership.

Interpretation of synthetic data analysis

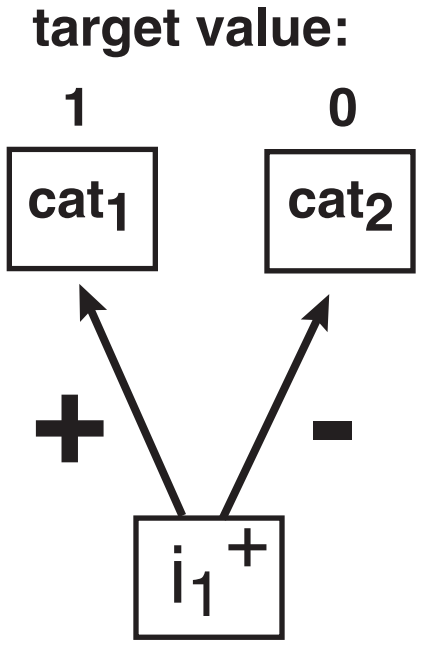
GLM and weight-based methods are sensitive to category membership. Both methods miss units that have characteristic values close to the mean of the noise distribution.

The noise-based analysis misinforms about the sensitivity of particular voxels. In a synthetic environment containing completely non-overlapping patterns, the noise-based method suggests that voxels sensitive to single categories are sensitive to *all* categories. It correctly classifies voxels that are always noise.

A caveat: It is possible that using a competitive activation function on the output layer (Hanson et. al, in press) would change the dynamics; this will be investigated using the synthetic data method.

Why does the noise-based analysis overestimate category sensitivity?

Backprop alters weights to drive the category units to their target values.



If input 1 has a positive value for category 1, the weight to cat_1 will be increased and the weight to cat_2 will be decreased.

When a noise source perturbs input 1, both cat_1 and cat_2 will be affected.

Weight-based is ok because is assigns a strong positive path strength to cat_1 and a strong negative path strength to cat_2.

Conclusions

It is possible to determine, with near ceiling accuracy, whether a subject is viewing a male face, female face, monkey face or dog face, based on just a few seconds worth of brain signal.

Several methods give similar results in this domain.

Interpreting what backprop has learned is complex: Synthetic data analysis shows that a weight-based analysis finds most, but not all relevant voxels.

For the type of backprop we used, a noise-based analysis suggests that representations are highly overlapping when in fact they are completely non-overlapping.

References

Carlson TA, Schrater P & He S (2003) Patterns of activity in the categorical representations of objects. Journal of Cognitive Neuroscience 15(5):704–717.

Cox DD & Savoy RL (2003) Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19:261-270.

Hanson SJ, Matsuka T & Haxby JV (in press) Combinatorial codes in ventral temporal lobe for object recognition: Is there a "face" area? NeuroImage

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425-2430.

An electronic version of this poster is available online:
http://compmem.princeton.edu/publications.html