

One-Class Pattern Recognition in Brain Activity via Neural Networks

Omer Boehm, David Hardoon and Larry Manevitz

University College, London
University of Haifa

Abstract:

In this paper, we show how one-class recognition of cognitive brain functions across multiple subjects can be performed at a good (above 85%) level of accuracy via appropriate choices of features. This betters the initial work of Hardoon and Manevitz (where such classification was first shown to be possible in principle, with 60% range on the same data) and work of various groups around the world (e.g. Mourao-Miranda (UCL) and Mitchell et al (CMU) which have concentrated on two-class classification.) The methodologies for one class classification investigated in this paper are the compression neural network (originally due to Cottrell) and developed by many, including Manevitz and Yousef and a version of one-class SVM due to Scholkopf et al. and Yousef and Manevitz. A genetic algorithm is used to help perform the feature selection.

Introduction:

In recent years, the possibility of obtaining fMRI studies as data and seeing if one can identify cognitive activity directly from the brain scans has become a real possibility [Mi], [MM1] [HM]. This correspondence between physiological information and specific cognition lies at the very heart of the goals of brain science.

Note that this work, is, in a sense, opposite to the problem of identifying which areas of the brain are associated with various cognitive activity, another area of central concern for brain science. However, in some ways there is a strong synergy between these two activities, for, while it might, in principle be possible to identify the cognitive activity from full brain data, most researchers in this area, starting with Mitchell [Mi] have realized that the strong noise to signal in brain scans requires aggressive feature selection.

This noise to signal ratio has several origins;

- One – the inherent noise in the technological scan,
- Two – the variability within a single subject,
- Three – the fact that a brain is actually performing many tasks simultaneously and one can not control for all of them,
- Four - brains are physically distinct across individuals and the mappings between them are only approximate [TT].

Five – MRI technology has limited resolution, so in a sense the original data is always "smeared".

In addition, considering the dimensionality of the data, one always has very few data points. A typical scan has about thirty thousand voxels in real values; while the expense and difficulty in data taking in fMRI of an individual is in the order of a hundred samples.

So, the problem being tackled has small data size, huge dimensionality, and a huge noise-to-signal ratio; *a priori* it would seem an unlikely endeavor. Nonetheless, the results reported (beginning with Mitchell [Mi], and others) show that this is possible. In these works methods to aggressively reduce non-relevant (noise) features were applied.

Note that if one manages to reduce the number of features, one is essentially finding the areas of the brain that are associated with the cognitive problem; i.e. the complementary problem.

In this work, we investigate using a "wrapper approach" [KJ], using the machine learning tool itself to find the features.

One-Class versus Two-Class

When one is looking for a two class (or n-class with $n \geq 2$) the assumption is that one has representative data for each of the classes and uses them to discover separating manifolds between the classes. While the most developed machine learning techniques address this case, this is actually a very unusual situation.

First, while one may have invested in obtaining reasonably representative data addressing one class; it is unusual to have a representative sample of its complement (in 2 class learning). The same is true for the n-class case.

Moreover, each filter must be re-created as data for each class is obtained and divisions between sub-classes must all be trained separately. Furthermore, one can essentially never have data to distinguish between class A and "anything else". Thus while one may initially have data representing class A, B and C and then using two-class methods to find a filter distinguishing between class A and B, class A and C, and class B and C; or alternatively finding a filter between class A and class (B or C) and class B between (A or C); etc.

The assumption in using these filters will be that the data comes from one of these classes. Should one wish to add class D then many additional filters must be trained.

It is very unlikely that one has the data to say Class A versus "everything else". It is more likely to imagine a scenario where data is gathered for a particular kind of cognitive task and then, when data for another task is gathered, a different filter is made for the new class. Thus one can incrementally build up a library or "battery" of classification filters; and then test a new data point by this battery. Of course, it would then be possible for a data point to pass several such filters; but since the brain works in parallel, this can be considered an asset to the method rather than a deficit.

However, as expected, the results for two class classification were superior to those of one-class. Hardoon and Manevitz [HM] (IJCAI 2005) showed that while one class can be done in principle, for this fMRI task, their classification results (about 60%) were not sufficient.

In this work, we have remedied this deficit, by showing that one can obtain, automatically, filters with accuracy close to their two-class cousins. The main methodology was finding the appropriate features. This was a reasonable hope given the large dimension of features given by the fMRI map (which were all used in Hardoon and Manevitz [HM] .

To do this we proceeded with three main tools:

1. A choice of a one-class classifier approach. The two that we considered were
 - a. The “compression neural network”.
 - b. Different versions of “one-class” SVM.
2. The use of the “wrapper approach” to judge the quality of features.
3. A manual binary search proceeding by a “binary dissection” approach to the brain (at each stage using the one-class wrapper as an evaluator.)
4. More generally, the use of a genetic algorithm to isolate the best features.

The one class learning method was used to perform the evaluation function in the genetic algorithm.

Each of these steps has its own complications and choices.

For example:

Step 1a) requires choosing an appropriate compression ratio, for the one-class NN; and of course the training method.

Step 1b) has many variants; we did not exhaust all of them; but we found the results too sensitive to the choices and so in the end used a version of 1a almost exclusively.

Step 3, being manual took too long; we used its results to help decide on the initial conditions of the genetic algorithm.

Feature Reduction:

To recapitulate, this reduces the question of finding the features to a search for subsets in the space of features. In this work, we have examined both one class SVM and compression neural networks as the machine learning tool. These were investigated in Hardoon and Manevitz (IJCAI 2005), where it was found that, perhaps surprisingly to some, the NN approach worked somewhat better. This was less surprising when considering the work of Manevitz and Yousef, where it was shown, in a comparative study in a textual classification task, that while both seem to have similar capabilities; the SVM was much more sensitive to the choice of parameters.

The main emphasis of this work is the feature selection, using the wrapper approach. We followed two paths – initially, we worked by hand and did a primitive greedy search on the subsets as follows:

First, start with a "reasonable" area of the data scan; i.e. all background dead area thrown out; the most external levels of the brain discarded. That is, the raw scan had about 120,000 real valued voxels; after reduction we had about 70,000 voxels.

Second, *divide* (various options to do that) the brain into overlapping two or three geometrically contiguous boxes (by selecting along one dimension) - run the classifier and discard the lowest returns; Continue with the best box as long as it classifies better than the previous loop.

When all boxes do worse; consider either (i) perform a different division of the boxes along the same dimension as before, but now of different sizes that overlaps the previous chosen boxes or (ii) select the boxes by slicing in a different dimension. (i.e. if the search was on boxes defined by the row indices, now use the best row indices found and try to create boxes with different column indices).

Cease when no improvement is found.

Figure 1. Manual binary search via the one-class bottleneck NN

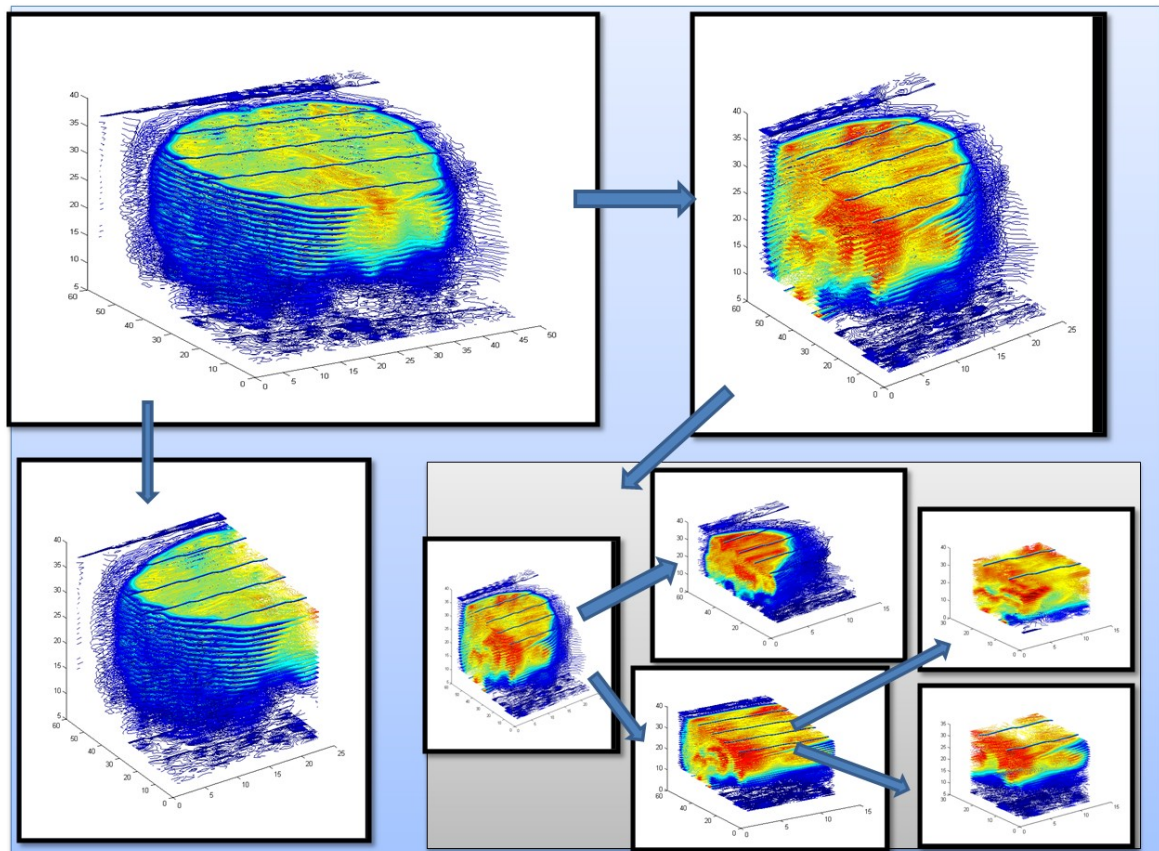


Figure 1 illustrates the a sample process of the manual greedy binary search. The assumption was that the task 'relevant' features reside in a relatively small contiguous chunk in the brain.

Following this work, we were able to produce the following table of results: (obtained on one of the possible search paths)

Each data point (3D matrix) which originally contained about 120,000 features was reduced as explained above into about 70,000 features . ([58 x 46 x 46] => [48 x 39 x 38]).

Iteration	[rows, columns, height]	# features	Houses	Objects	Patterns	Blank	Avg
1	[1-17,1-39,1-38]	25194	58%	56%	55%	60%	57%
	[15-33,1-39,1-38] *	28158	62%	55%	64%	65%	62%
	[30-48,1-39,1-38]	28158	55%	52%	50%	60%	54%
2	[15-33,1-39,1-15]	11115	61%	63%	55%	60%	60%
	[15-33,1-39,13-30] *	13338	69%	68%	72%	70%	70%
	[15-33,1-39,27-38]	8892	58%	57%	60%	60%	59%
3	[15-23,1-39,13-30]	6318	63%	69%	68%	62%	66%
	[20-26,1-39,13-30] *	4914	70%	67%	76%	79%	73%
	[25-33,1-39,13-30]	6318	60%	67%	70%	75%	68%
4	[20-23,1-39,13-30] *	2808	74%	70%	71%	73%	72%
	[22-25,1-39,13-30]	2808	65% ⁰	73%	60%	80%	71%
	[24-26,1-39,13-30]	2106	70%	69%	69%	68%	69%
5	[20-21,1-39,13-30]	1404	67%	65%	74%	63%	67%
	[21-22,1-39,13-30]	1404	60%	63%	70%	64%	64%
	[22-23,1-39,13-30]	1404	65%	63%	72%	68%	67%
back							
6	[20-23,1-18,13-30]	1296	67%	66%	70%	72%	69%
	[20-23,19-39,13-30]	1512	67%	70%	72%	78%	72%

Table 1 – Manual binary search via the one-class bottleneck NN for 'faces' data.

* Chosen part. (If no part is chosen, the current path is terminated, and a different division step is performed – see text).

The table represents the results for the 'faces' (fMRI data acquired for subjects while viewing 'Faces'). We used a bottleneck NN (with compression rate of 60%) which was trained (5 folds) solely for the 'faces' data and then tested against the rest of the categories. The decision how to continue was according to the average over all categories. As can be seen, this method brought us up to 80% accuracy on blank data, and 72% on average.

For a control, we also considered random selection of about the same proportion of features; and the results were not much above random.

The Genetic Algorithm

Of course such a work is very tedious and there are many intuitive choices. In an attempt to automate it, we decided, to apply a genetic algorithm approach to this problem, although the computational requirements became almost overwhelming.

This means, that each member of the genetic algorithm population is a binary vector of dimension 70,000 (corresponding to the “trimmed” brain) where a “1” indicates that a feature is chosen. In the initial generation, we typically chose about 10% of the features selected randomly with a population size of 20..

We have implemented and tested several representations for the genomes and a corresponding crossover and mutation strategies.

We produced the following results, on the same data of the [HM] Haroon-Manevitz study.

Category Filter	Faces	Houses	Objects	Patterns	Blank
Faces	-	-	-	-	-
Houses	84%	-	-	-	-
Objects	83%	91%	-	-	-
Patterns	92%	85%	92%	-	-
Blank	91%	92%	92%	93%	-

Table 2 – Genetic algorithm results. The genetic algorithm was able to find a filter for each class with a success rate almost similar to the ones produced for the two class classifiers.

Location of Areas of Brain Associated with Cognitive Tasks.

Having discovered features appropriate for classification; it is interesting to enquire whether or not these features are local (i.e. presented in a particular area of the brain) or distributed. Of course, this can only be asked up to the resolution of the fMRI data themselves.

To get a feel for this, we produced a primitive visualization tool of the data; where we give a three dimensional location of the genetic algorithm chosen features.

Although we have not yet quantitatively analyzed these results yet, a visual analysis does indicate a locality of the features and one can see that certain areas of the brain are important for the classification of each task. Interestingly, they are not identical; i.e. it is not just the values of these features that are important for each task but their choice.

Summary and Future Work

1. We have shown that classifying visual cognitive tasks can be done by one-class training techniques to a high level of generalization.
2. We have shown that genetic algorithms; together with the one class NN compression network can be used to find appropriate features that, on the one hand, increase the accuracy of the classification to close to that obtainable from two-class methods.
3. Preliminary results show that this method can indicate areas of the brain critically associated with the cognitive task

This work needs to be extended to other (non-visual) cognitive tasks; and it needs to be seen to what resolution the work can be carried out. Can specific styles or specific faces of people be identified from these kind of mechanisms? Is there a theoretical limit on either the accuracy or the resolution?

References

- [CMZ] Cottrell, G. W., Munro, P. & Zipser, D. (1988). Image compression by back propagation: an example of extensional programming. *Advances in Cognitive Science*, 3.
- [HM] David R. Hardoon, Larry M. Manevitz (2005) .fMRI Analysis via One-class Machine Learning Techniques. *IJCAI 2005*, 1604-1605
- [KJ] R.Kohavi, G.H. John (1997). Wrappers for Feature Subset Selection. *Artificial Intelligence 1997* .
- [JMG] Japkowicz, N., Myers, C. & Gluck, M. A. (1995). A novelty detection approach to classification. *International Joint Conference on Artificial Intelligence*(pp. 518–523).
- [MY1] Manevitz, L. & Yousef, M. (2001). One-class svms for document classification. *Journal of Machine Learning Research* 2 (pp. 139–154).
- [MY2] Manevitz, L. & Yousef, M. (2007). Document classification via neural networks trained exclusively with positive examples. *Neurocomputing*, 70, 1466 –1481.

- [Mi]** Mitchell, T. M. , Hutchison, R. Niculescu, R.S. , Pereira, F. , Wang, X, Just, M. and Newman, S. (2004) Learning to decode cognitive states from brain images., *Machine Learning*, 57:145-175, 2004
- [MM1]** Mourao-Miranda, J., Reynaud, E., McGlone, F., Calvert, G. & Brammer, M. (2006). The impact of temporal compression and space selection on svm analysis of single-subject and multi-subject fmri data. *NeuroImage* 2006, doi:10.1016/j.neuroimage.2006.08.016..
- [SP]** Scholkopf, B., Platt, J., Shawe-Taylor, J., Smola, A. & Williamson, R. (1999). Estimating the support of a high-dimensional distribution. Technical Report MSR-TR-99-87, Microsoft Research.
- [TT]** Talarich, J. and Tournoux, P. Coplanar Stereotaxic Atlas of the Human Brain, Thieme Medical, 1988, p. 122.
- [Y]** Yousef, M. (2000). Document Classification Using Positive Examples Only. PhD thesis, University of Haifa.