456

456

On clustering of fMRI time series

P. Toft, L.K. Hansen, F.Å. Nielsen, C. Goutte — Department of Mathematical Modelling, Technical University of Denmark
S.C. Strother — PET Imaging Service, Minneapolis VA Medical Center, Minnesota, USA
N. Lange — Brain Imaging Center, McLean Hospital & Harvard University, Boston, USA
N. Mørch, C. Svarer, O.B. Paulson — Neurobiology Research Unit, Copenhagen University Hospital, Denmark
R. Savoy, B. Rosen — Massachusetts General Hospital, Boston, USA
E. Rostrup, P. Born — Danish Research Centre for Magnetic Resonance, Hvidovre Hospital, Denmark

Abstract

- We use *k-means* to cluster fMRI time series.
- Instead of clustering on the raw time series, we use the cross-correlation function with the *paradigm* (the activation stimulus).
- The performance of k-means is very dependent on the initial setting of class centers.

fMRI time series

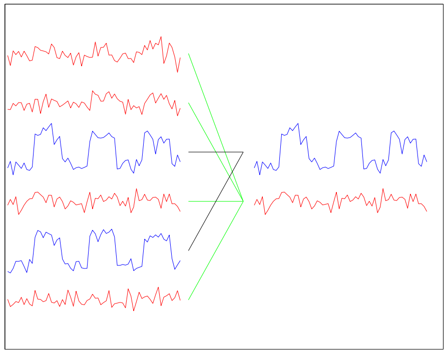
The fMRI signal measures the hemodynamic response to focal neural activity.
It is a combination of several interacting components:

- actual response,
- neuronal activations,
- global components (breathing, etc.).

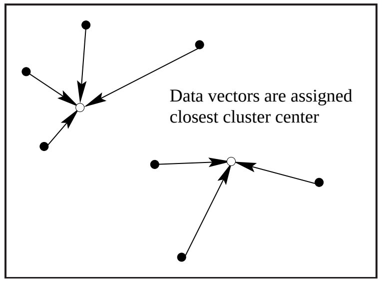
Signal correlated in *time* and *space*.
Both *short-term* and *long-term* correlations.

Clustering

The data vectors are labelled according to the nearest cluster center.



In a high dimensional space the data vectors are assigned to the closest center:



In this figure, only two of the dimensions are shown. Black circles represent the data, empty circles the cluster centers.

k-means Clustering

Clustering is an exploratory data analysis technique.
The k-means algorithm is a simple non-parametric clustering method.

1. Initialise K clusters $k = 1 \ldots K$, with centers $C_k^{(0)}$. Iteration $i = 0$.
2. Assign each data vector x_j to the cluster k with the nearest center $C_k^{(i)}$.
3. Set new cluster centers $C_k^{(i+1)}$ to the center of gravity of the cluster members:
   $C_k^{(i+1)} = E\{x_j\}_{x_j \in k}$
   This formula can also be modified to use the median and/or include a smoothing term.
4. Goto step 2 until convergence.

Convergence is usually very fast: only few iterations are needed.
However, the result is very dependent on two factors:

- The number of clusters used,
- the initial setting of the centers.

The Metric

The k-means algorithm relies on calculating the distance from data vectors x_j to cluster centers $C_k^{(i)}$.
This requires the use of a proper *metric*.
We will use a standard Euclidean distance: $d(x, y) = \|x - y\|_2^2$.

The trivial clustering would use the raw data vectors as input to the k-means algorithm. The results do not allow to identify functionally meaningful regions.

Using the Cross-correlation

In this work, we cluster on the cross-correlation function between the measured fMRI time series and the paradigm, rather than on the raw response.

- This emphasizes the similarity between the data vectors and the paradigm.
- Noise influence is reduced, due to the low frequency dominance in the paradigm.
- The dimensionality of the data is reduced, especially the high frequency part.

Experiments

The data has been collected in the following situation:

- **subjects :** single subject.
- **task :** left-handed finger-to-thumb opposition.
- **modality :** whole-brain echo planar fMRI images.
- **paradigm :** multiple runs of 72 images. 24 baseline, 24 activation, 24 baseline
- **time step :** 2.5 second.
- **location :** axial slice through primary motor cortex and SMA.
- **slice :** 39x49 voxels.
- **voxel size :** 3.1x3.1x8mm.

We present two kinds of experiments:

1. k-means clustering on the raw response,
2. k-means clustering using the cross-correlation.

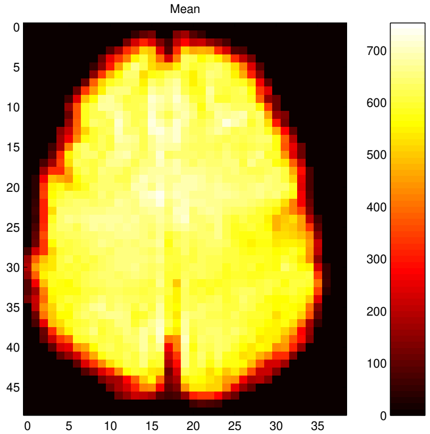
The number of clusters has been set to 10 in both experiments, and we use 10 iterations of the k-means algorithm—enough to stabilize the position of the cluster centers. Each cluster center is the "average cross-correlation" of the cluster members.

In each voxel, the actual cross-correlation between the voxel response and the paradigm is replaced by the cross-correlation between the corresponding cluster center and the paradigm.

Results are presented on the following figures.

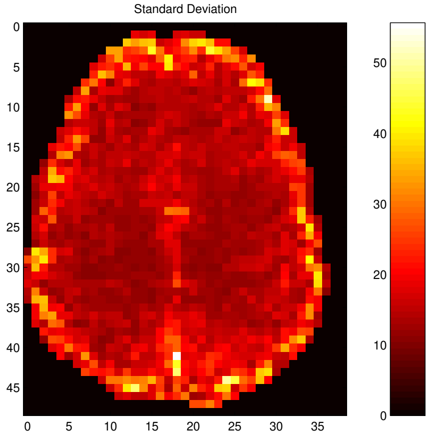
Data description

The mean of the fMRI time series in each voxel:



The mean carries virtually no information about the activated areas.

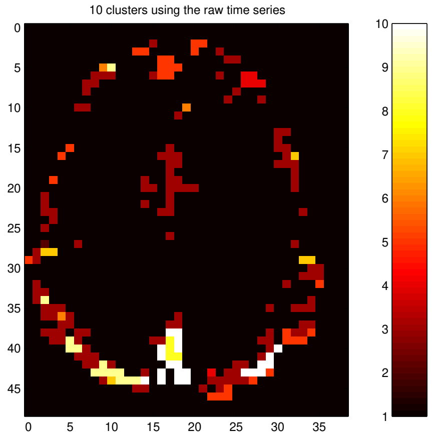
The standard deviation of the fMRI time series in each voxel:



The standard deviation does not seem to correlate with the activation paradigm.

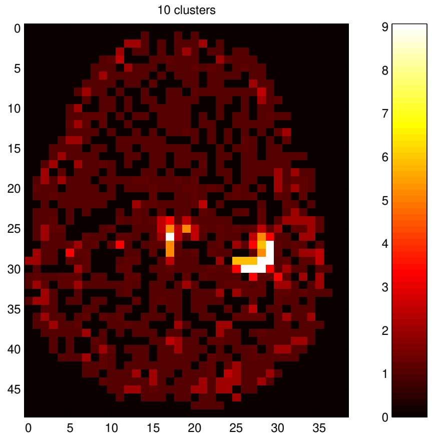
Results

On the raw hemodynamic response: max cross-correlation for the cluster center for each voxel. There are 10 centers, thus 10 colors.



No information about activated areas.

On the cross-correlation: max cross-correlation for the cluster center for each voxel. There are 10 centers, thus 10 colors.



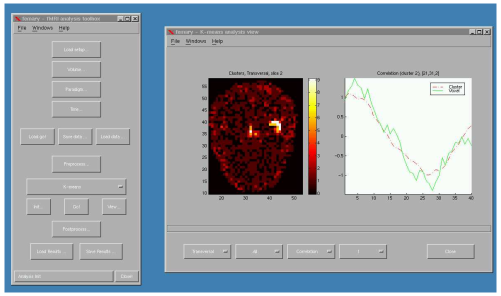
Primary motor area and SMA are picked up as two spatially separated voxel groups with similar correlation with paradigm.

Discussion

Clustering on the raw fMRI time series does not seem to provide much insight into the actual activation induced by the paradigm.
On the contrary, using the cross-correlation between the data and the paradigm, we easily identify the relevant activated areas.
It should be noted that the results are highly dependent on the initialization of the cluster centers.
A possible solution would be to pre-process the data using k-means clustering and a large number of clusters (e.g. 50), then use a deterministic, hierarchical clustering method. This is currently under investigation.

Implementation

The algorithm is part of a Matlab toolbox currently under development at DTU.



References

G. Scarth et al., *Human Brain Mapping*, 3, S89, 1996.
B. Ripley, *Pattern Recognition and Neural Networks*, Cambridge, 1996.

The project is funded by The Danish Research Councils and the Human Brain Project P20 MH57180 "Spatial and Temporal Patterns in Functional Neuroimaging".