Clustering fMRI data using a spatiotemporal geodesic k-means algorithm

R. A. Masterton[1,2], D. F. Abbott[1,2], and G. D. Jackson[1,2]

[1]Brain Research Institute, Melbourne, Victoria, Australia, [2]The University of Melbourne, Melbourne, Victoria, Australia

Introduction: For some problems, it is desirable or even necessary to reduce fMRI data into clusters of voxels rather than having to deal with tens-of-thousands of individual voxel time courses. While sacrificing spatial resolution, reducing the data into a smaller number of data elements makes it possible to complete, in a much shorter time, computationally expensive operations such as pair-wise comparisons between all data points.

Simple sub-sampling, using a uniform neighbourhood averaging kernel, suffers from partial volume effects, as the shape of the averaging kernel (e.g. sphere, box etc) is usually not matched to the underlying structure of the data. Better approaches cluster voxels together based upon measures of 'similarity', and summarise with a statistic (such as the mean) of the member voxels. The similarity measure may be based on structure (e.g. using *a priori* region definitions) or functionality (e.g. correlated time-courses). However, approaches using spatial or temporal information in isolation only take into account one aspect of the underlying nature of the data.

Here we present a data-driven clustering algorithm that uses *both* spatial and temporal information to group voxels into spatially connected and temporally homogeneous clusters. We present the results of applying this method to both simulated and real fMRI data.

Methods: We model the spatiotemporal topology of the fMRI using a weighted graph. Each voxel in the brain is represented as a vertex in the graph with edges connecting adjacent voxels. We assign each edge a weight, based upon the correlation between the associated voxel time courses - highly correlated neighbouring voxels are assigned a low weighing and *vice versa*. The geodesic distance between two vertices in this graph (i.e. the sum of weights along the edges of the shortest path connecting the two vertices) hence provides a measure of the spatiotemporal relationship between the corresponding voxels – a short distance implies the two voxels are close in both spatial location and temporal correlation.

We performed clustering on the graph representation of the data using a geodesic k-means algorithm which partitions the data into k separate clusters. Each cluster has an associated centre – the vertex with minimum eccentricity within the cluster – and the k-means algorithm minimises the sum of geodesic distances between each vertex and its cluster centre. We implemented this method using C and Matlab (The Mathworks, Natick, MA, USA) as an SPM2 toolbox (www.fil.ion.ucl.ac.uk/spm/) which allows for quick clustering of whole brain fMRI data sets (e.g. <1 minute for 200 volumes each 128x128x25 voxels using a 2.8GHz Intel Pentium processor).

For testing, we created a simulated fMRI time-series with 200 pre-defined regions. Voxels belonging to the same region were assigned a common time-course, and Gaussian noise was added to produce a signal-to-noise ratio of 0.5. We also created a "noise" data-set with the voxel time-courses not grouped into regions. We processed both sets of simulated data using our clustering algorithm. We rated performance by matching the clusters produced by our algorithm with the pre-defined regions and measuring the degree of overlap (i.e. the percentage of mutual voxels within the cluster-region pairs). To ensure the cluster-region pairs provided maximum overlap, an optimal matching was chosen using the Kuhn-Munkres algorithm. The algorithm's performance when clustering the noise data-set provided a null-distribution for comparison.

We also tested our method on real data acquired from 8 healthy subjects undergoing 10 minutes of fMRI scanning in an eyes-closed resting state. In this case, where there was no ground truth to compare the clusters against, we rated performance upon the temporal homogeneity of the clusters - as measured by the mean, within each cluster, of the correlation between individual voxel time-courses and the mean time-course of the cluster. Again we compared this to a null result created by running our clustering on a noise data set.

Results: In fifty trials of simulations, our method was consistently better than the null distribution – with the geodesic k-means clusters always providing a better match to the template regions (Figure 1). On average, at least half of the clusters produced by our method shared 70% or more voxels with a region defined in the template.

For the real data, each subject was clustered into 1000 clusters. Figure 2 shows the distribution of the cluster homogeneity for the pooled 8000 clusters of all subjects. Compared to the null condition there was an increase in the mean homogeneity of the clusters (0.81 versus 0.73), but the more important result was the reduced variability. The 5^{th} and 95^{th} percentiles of the cluster-homogeneity distribution were 0.71 and 0.91 respectively whereas in the null condition they were 0.51 and 0.90.

Discussion: We have shown that a geodesic k-mean clustering algorithm can decompose fMRI data into clusters of relatively homogeneous, spatially-connected voxels. This decomposition provides a reduced representation of the data whilst preserving its topology and retaining much of its temporal structure. One possible application for this method is the investigation of distributed brain networks by examining coherence between cluster time-courses.

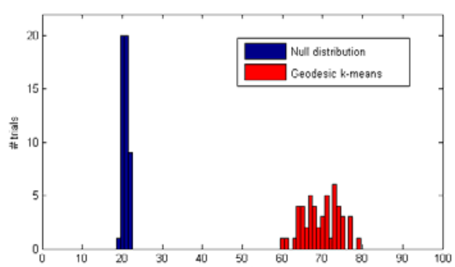
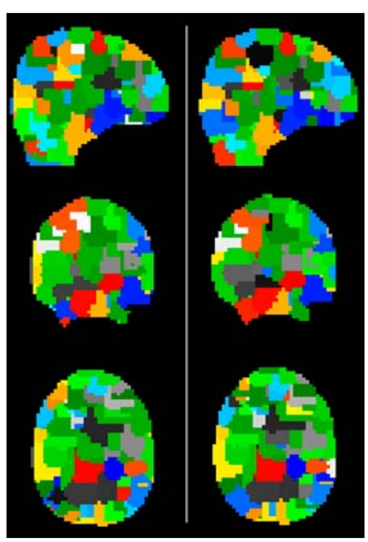


Figure 1 – Simulation results:

Above) Histogram summarising the results of running fifty trials of clustering a simulated data set. The clustering performance is summarised with the median overlap (in terms of the percentage of mutual voxels) between the matched cluster-region pairs.

Right) An example of the clusters produced by our method (left), compared with the regions in the simulated data (right). A significant overlap between the two is evident.



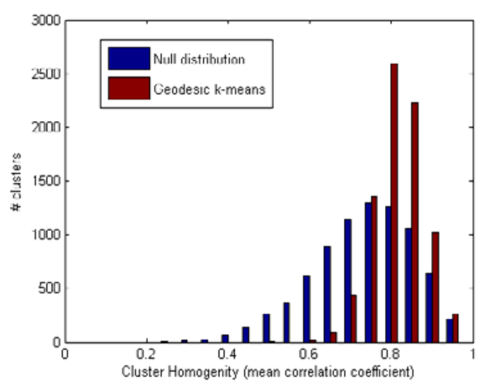


Figure 2 – Resting-state fMRI results:

Histogram of the cluster-homogeneity for all eight subjects (8000 clusters). The cluster-homogeneity is scaled from 0-1 and represents the homogeneity of voxel time-course within a cluster.