Hierarchical clustering for network analysis in functional connectivity MRI

G. J. Thompson[1,2], M. Magnuson[1,2], and S. D. Keilholz[1,2]
[1]Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, United States, [2]Biomedical Engineering, Emory University, Atlanta, GA, United States

Purpose of Study

An increasing number of studies of functional connectivity within the brain are being performed using MRI, and such work has shown the potential to both better understand the healthy brain [1] and diagnose diseases of the brain [2]. These techniques are based on the blood oxygenation level dependent (BOLD) signal and areas of the brain known to be connected anatomically have been shown to have voxel time courses which have higher correlation coefficients than others [3].

Seed-based correlation is the method most commonly used to analyze functional connectivity data [3], but requires user input in selecting the seed region, which can introduce bias. Independent component analysis is a data-driven approach that can identify functional networks [4], but the components it produces are unranked and can be difficult to interpret. Due to these limitations, it would be beneficial to develop a data-driven algorithm which does not require user input and produces easily-interpreted results. Hierarchical clustering based on cross-correlation values between individual voxels is a promising candidate. This work examines the use of hierarchical clustering to characterize the functional connectivity of the sensorimotor network in the rodent. Resting state fluctuations similar to those observed in humans have also been detected in rats [5], and the relatively simple structure of the rat cortex facilitates evaluation of the analysis method.

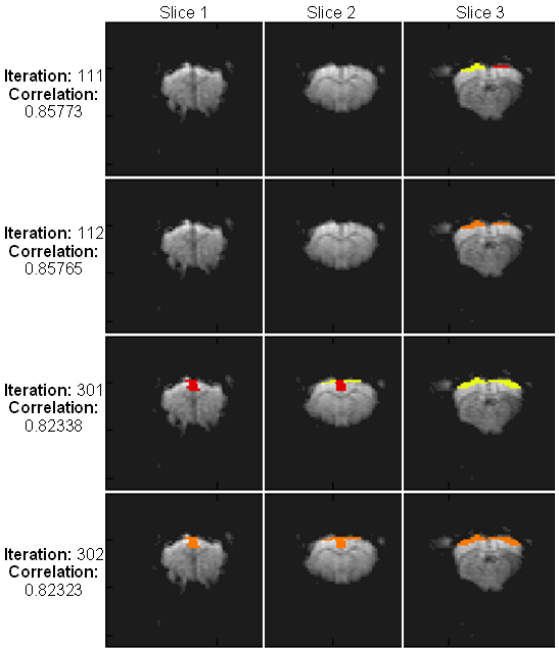

Figure 1: 3 slices from a single rat at different iterations into the algorithm, with corresponding cross-correlation values. From iterations 111-112 the left (yellow) and right (red) SI groups merge (orange), and from iterations 301-302 the SI (yellow) and motor (red) groups merge (orange).

Algorithm

In an image of size N x N, K slices, a cross correlation matrix of size $(K*N^2)^2$ is created where the entry at row x and column y represents the correlation coefficient calculated between the x^{th} and the y^{th} voxels. The first group is formed including only the pair of voxels with the maximum correlation coefficient. Then further groups are formed in an iterative manner as follows; The pair of voxels with the next highest correlation is considered. If one of the pair has already been added to a group, then the other voxel in that pair is also added to that group. If both voxels have already been added to different groups, then those two groups are merged into a single larger group. If neither voxel is in a group, then a new group containing only the pair is created. Repeat.

While this algorithm is capable of forming groups, it has hundreds of iterations due to the hundreds of voxels in the image, and eventually will merge all voxels together into a single group. Because of time constraints, this study stopped forming groups at a correlation value of 0.65, chosen because most of the cortex was classified by that point. Because there were hundreds of merges, a method was needed to determine when a significant merge between two large groups was occurring, to be able to determine the significant large groups at varying levels of correlation. Therefore only merges of two groups, where each group was at least 6 voxels in size and the smaller group was at least 20% of the larger group, were considered. This ensured that small groups due to noise weren't considered, and that insignificant additions to an existing large group weren't considered. This algorithm was implemented in *MATLAB*.

Computational Considerations

To reduce the size of the cross correlation matrix, rather than using all N^2 voxels, only voxels which have amplitude indicating that they contain anatomical data are used. As not all voxels are used, 3 extra columns are added to the cross correlation matrix to indicate each voxel's x, y and slice number position. In addition, as the correlation coefficient operation is commutative, only the upper triangular half of the correlation matrix need be calculated.

The limiting step of the algorithm is determining the maximum within the cross correlation matrix, to determine the next pair. If the maximum is calculated in the normal *MATLAB* manner the algorithm will take on the order of $(K*N^2)^4$ operations to complete. To reduce this slow computation the cross correlation matrix was sorted beforehand using a binary sort, on the order of $(K*N^2)^3\log_2(K*N^2)$ operations, and reducing the order to only $(K*N^2)^2$ operations.

These considerations reduced run time from several days to less than 1 hour for all data sets tested.

Results

The algorithm was run on data collected from 6 rats anaesthetized with medetomidine, TR 500 ms, TE 20 ms, matrix size 64 x 64, field of view 1.92 cm x 1.92 cm, 4-5 2 mm thick slices centered over primary somatosensory cortex (SI), 1200 repetitions, as done in [5].

It was observed that groups formed corresponding to the anatomical features of the left and right lateral primary somatosensory cortex (SI), the left and right lateral secondary somatosensory cortex (SII), and a dorsomedial region likely to correspond to the motor cortex. Based on its criteria for determining significant events, the algorithm found two large groups merging to form a SI group in 3 rats, two large groups merging to form a motor group in 4 rats, left and right SI groups merging in 4 rats, SI and motor groups merging in 4 rats, and two large groups merging to form bilateral SII in 2 rats. Each event and the correlation coefficient at which it occurred is shown in figure 2. Figure 1 shows a typical observation; groups form separately in SI, SII and the motor area. SI and SII merge, then merge with the motor area.

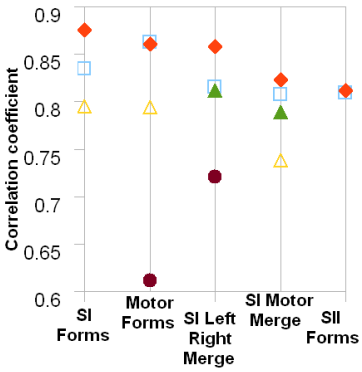

Figure 2: The cross-correlation values at which each group formation event occurred in each rat. Each rat is represented by a different color and shape, events which did not occur in particular rats are not displayed.

Discussion

This study demonstrates that an algorithm can be created which constructs anatomically relevant groups from functional connectivity MRI data in a repeatable manner across subjects. This algorithm fulfills the goal of not requiring anatomical information to be specified, and partially fulfills the goal of not requiring a correlation threshold to be pre-determined, though one was set due to time constraints.

This algorithm could allow detection of networks in human data with minimal or no manual input, however despite efforts discussed above this algorithm is still computationally intensive. Therefore it might be necessary to group voxels of high correlation in human data prior to execution of this algorithm.

References

1. Fox et al., Proc Natl Acad Sci USA, vol.102, pp.9673-8, Jul 5 2005. **2.** Rombouts et al., Hum Brain Mapp, vol. 26, pp. 231-9, Dec 2005. **3.** Biswal et al., Magn Reson Med, vol. 34, pp. 537-41, Oct 1995. **4.** Calhoun et al., Hum Brain Mapp, vol. 13, pp. 43-53, May 2001. **5.** Williams et al., Proc Int Soc Magn Reson Med, p. 2119, 2006.