Novel metrics and application of nearest-neighbor feature selection for creating resting-state fMRI brain atlases

Bryan A. Dawkins[1], Trang T. Le[2], Alejandro A. Hernandez[1], and Brett A. McKinney[1,3,*]
[1]Department of Mathematics, University of Tulsa, Tulsa, OK 74104, USA
[2]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA 19104
[3]Tandy School of Computer Science, University of Tulsa, Tulsa, OK 74104, USA.

Abstract

Resting-state functional connectivity MRI (rs-fMRI) data consists of correlation matrices, where correlations are computed between the time series from brain Regions of Interest (ROIs). There are many different parcellations of the human brain into collections of ROIs. These parcellations, or atlases, can be used in case-control studies in order to understand and accurately classify subject phenotypes. We present new metrics for nearest-neighbor distance-based feature selection at the ROI level. Using our new metrics, we apply a novel nearest-neighbor feature selection algorithm to calculate relative importance of ROIs in two existing brain atlases. We use integer programming to derive a mapping between brain atlases to determine spatially similar ROIs. With ROI importance scores and spatial similarity between atlases, we create a new brain parcellation that combines aspects of both brain atlases.

1 Background

Resting-state fMRI data exists in high dimensions and has many sources of noise, such as physiological or motion related [1]. Feature selection is typically done with the purpose of determining brain regions of interest (ROIs) that accurately discriminate between cases and controls in order to understand a particular phenotype. The data consists of pairwise ROI-ROI correlations, where each ROI is a time series measuring brain activity in a particular region or regions of the brain while a subject is not performing a task. A typical data set consists of m subject-specific correlation matrices of dimension $p \times p$, where the pairwise correlations are computed between p ROIs from a brain atlas. Nearest-neighbor distance-based feature selection in rs-fMRI data has been performed using the private evaporative cooling method, which used pairwise ROI-ROI correlations as predictors of a particular phenotype. However, nearest-neighbor feature selection algorithms have not been applied at the ROI level to assess the relative importance of ROIs for a given phenotype. To address this, we have previously proposed a new distance metric that allows us to compute the importance of individual ROIs using a nearest-neighbor distance-based approach. We use this new distance metric with a novel nearest-neighbor feature selection algorithm called Nearest-neighbor Projected Distance Regression (NPDR) in order to compute ROI importance and the corresponding pseudo P values [2]. Our analysis is done on subject rs-fMRI correlation matrices generated by two well known brain atlases [3, 4].

In order to make spatial comparisons between any pair of brain atlases, we first compute a distance matrix containing all pairwise distances between the different collections of atlas ROIs. Distances are defined based on a set dissimilarity metric that accounts for differences in voxel collections between pairs of ROIs. In a particular

coordinate system, voxels have well defined three-dimensional locations in a given brain atlas. As long as two different atlases are in the same coordinate system, we can compare voxel membership between opposing atlas ROIs. We use an integer program that defines the standard Assignment Problem (AP) to find the one-to-one mapping between the two sets of atlas ROIs [5]. The collection of all mapped ROIs gives the closest spatial analogy between the two atlases, which tells us the closest relationship between the two sets of ROIs from different atlases. The collection of unmapped ROIs gives an indication of spatial uniqueness in the two atlases, respectively. All ROIs can be further mapped to a well defined anatomical region of the brain, which allows us to point out potential targets for better understanding the phenotype of interest.

Our spatial mapping between atlases and relative importance scores for ROIs in each respective atlas provides a way to combine relevant and distinct aspects of each brain atlas into a new parcellation. This new atlas includes important ROIs that are in the optimal one-to-one mapping from the solution to the assignment problem and any important unmapped ROIs from each atlas. Spatial overlap and attribute importance can serve as a useful tool for other researchers to compare, contrast, and combine two atlases. In particular, our results show how one might choose either of the two atlases to study the phenotype of interest we are considering in this work.

2 Methods

In this section, we first describe real rs-fMRI data generated from healthy controls (HC) and subjects with major depressive disorder (MDD), eating disorder (ED), substance abuse (SA), or anxiety disorder (AD). Using integer programming, we then derive a one-to-one mapping between the ROIs in two brain atlases used to generate the real data mentioned previously. Finally, we use our new distance metric for rs-fMRI data, along with NPDR, to compute importance scores for ROIs in each atlas from the real data.

2.1 Real rs-fMRI data

This is where we describe the LIBR data.

2.2 Spatial overlap between brain atlases

Let R_A and R_B represent regions of interest (ROIs) in atlases A and B, respectively. We assume that atlases A and B are in the same coordinate space. Since R_A and R_B are just collections of voxels that have well defined three-dimensional coordinates within an atlas, the spatial overlap between R_A and R_B can be defined as the set intersection between the two ROIs. Spatial dissimilarity between R_A and R_B can be computed with the Jaccard metric, which is given by the following

$$\mathrm{d}^{\mathrm{J}}(R_A, R_B) = \frac{|R_A \cup R_B - R_A \cap R_B|}{|R_A \cup R_B|}, \tag{1}$$

where the $(-)$ sign denotes set complement and $|\cdot|$ represents set cardinality. If the intersection $R_A \cap R_B$ is empty, then the two ROIs do not share any voxels and the Jaccard distance (Eq. 1) between them is 1. On the other hand, the Jaccard distance is 0 if the union $R_A \cup R_B$ and intersection $R_A \cap R_B$ are the same sets, which means the two ROIs have exactly the same voxels. All other possible Jaccard distances between R_A and R_B are strictly within $(0, 1)$. Hence, the Jaccard metric is contained within $[0, 1]$. The reason for division by $|R_A \cup R_B|$ in the denominator of the Jaccard metric (Eq. 1) is specifically to normalized the distance to be within $[0, 1]$. Otherwise, this distance between two ROIs would be affected by the cardinalities of R_A and R_B, respectively.

The Jaccard metric is intuitive in this context because ROIs are not just points in space, but rather they can have irregular three-dimensional shapes. Therefore, a Euclidean metric that gives the straight-line distance between two points does not necessarily indicate 'closeness' between two ROIs. It is possible to compute the Euclidean distance between the centroids of two ROIs, but the ROIs may not share many voxels due to their potentially irregular shapes. Therefore, it is more informative to use a distance metric that uses set operations like the Jaccard metric (Eq. 1). We show an example (Fig. 1) of the Jaccard distance between ROIs R_A and R_B that contain n_1 and n_2 voxels, respectively.

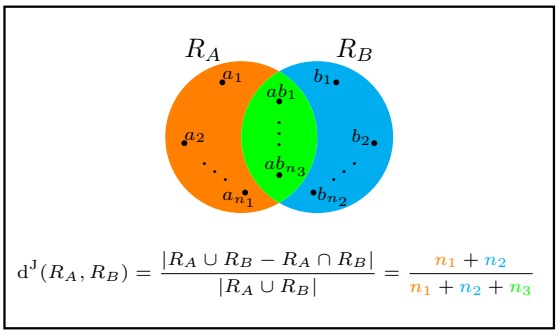

Fig 1. Example computation of Jaccard distance between ROIs R_A and R_B from two atlases A and B, respectively. There are n_1, n_2, n_3 voxels in R_A only, R_B only, and both R_A and R_B, respectively. The numerator gives the number of voxels unique to R_A (n_1) plus the number of voxels unique to R_B (n_2). The denominator contains the total number of voxels in R_A or R_B.

Each ROI in atlas A may overlap many different ROIs in atlas B. On the other hand, some ROIs in A may not overlap any ROIs in B. Furthermore, it is likely that A and B contain different numbers of ROIs. If we want to compute a minimum distance one-to-one mapping between the atlases, it is possible that some ROIs in A will not have a mapped partner in B. In order to efficiently compute this atlas-atlas mapping, we formulate this task as a standard Assignment Problem [5], which has a very concise definition (Fig. 2).

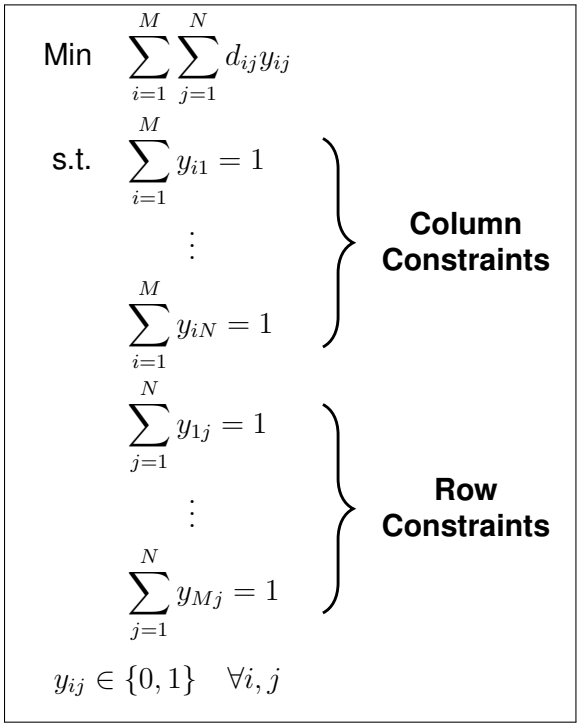

Fig 2. Assignment problem mathematical definition. The assignment matrix Y is binary, where $y_{ij} = 0$ if nodes i and j are assigned to each other and 0 otherwise. The distance matrix D between all nodes is computed to assign costs to arcs (or edges) between nodes. The distance between nodes i and j is denoted by d_{ij}. Therefore, the objective function is the sum of all pairwise distances in the collection of assigned arcs. the column and row constraints dictate that each node is connected to exactly one and only one other node.

The objective function is the sum over all pairwise distances between nodes included

in the mapping, where inclusion is determined by the binary solution matrix Y that has the following definition

$$y_{ij} = \begin{cases} 1 & \text{nodes } i \text{ and } j \text{ connected,} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Each row and each column of Y has a sum-to-one constraint, which means that each node in one collection is connected to exactly one other node in another disjoint collection. This problem assumes that the order (or size) of each collection is equal ($M = N$ in Fig. 2), so that a one-to-one assignment is possible. In the context of brain atlases, we will satisfy this requirement by adding artificial variables to our solution matrix Y. Pairwise distances between actual ROIs in a brain atlas and an artificial variable will be given a large constant value, so that ROIs in atlas A will preferentially map to another ROI in B if a mapping is possible. Absent the possibility of a mapping, an ROI will map to an artificial variable, which implies that this particular ROI goes unmapped in our unconstrained solution.

2.3 Relative importance of ROIs

3 Results

3.1 New brain atlas

4 Discussion

References

1. César Caballero Gaudes and Richard C. Reynolds. Methods for cleaning the BOLD fMRI signal. *NeuroImage*, 154:128–149, December 2017.

2. Trang T. Le, Bryan A. Dawkins, and Brett A. McKinney. Nearest-neighbor Projected-Distance Regression (NPDR) detects network interactions and controls for confounding and multiple testing. *Under Review*, 2019.

3. Jonathan D Power, Alexander L Cohen, Stephen M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran M Miezin, Bradley L Schlaggar, and Steven E Peterson. Functional network organization of the human brain. *Neuron*, 72(4):665–678, November 2011.

4. X. Shen, F. Tokoglu, X. Papademetris, and R. T. Constable. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage*, (0):403–415, November 2013.

5. David W. Pentico. Assignment problems: A golden anniversary survey. *European Journal of Operational Research*, 176:774–793, November 2007.