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 with a relatively low signal-to-noise ratio. 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. Our analysis is done using subject rs-fMRI correlation matrices generated by two well known brain atlases.