Global PCA of Local Moments

With Applications to MRI Segmentation

Jacob M. Maronge[1, 2], John Muschelli[3], Ciprian M. Crainiceanu[3]

[1]University of Wisconsin - Madison, Department of Statistics

[2]University of Wisconsin - Madison, Department of Biostatistics and Medical Informatics

[3]Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Contact Information:
Jacob M. Maronge
Department of Statistics
University of Wisconsin - Madison
1300 University Ave, Madison, WI, 53706

Email: jmmaronge@gmail.com
Website: https://jmmaronge.github.io

Abstract

We are interested in describing the information contained in local neighborhoods, and higher moments of local neighborhoods, of complex multimodal imaging techniques at the population level. This is problematic because of the size of medical imaging data. We propose a simple, computationally-efficient approach for representing the variation in multimodal images using the spatial information contained in all local neighborhoods across multiple subjects. This method achieves 3 goals: 1) decomposes the observed variability images at the population level; 2) describes and quantifies the main directions of variation; 3) uses these directions of variation to improve segmentation and studies of association with health outcomes. To achieve this, we efficiently decompose the observed variation in local neighborhood moments. In order to assess the quality of this method we present results using the 2015 Ischemic Stroke Lesion Segmentation (ISLES) Challenge.

Introduction

Every image can be vectorized. However, its meaning, interpretation, and complexity is encapsulated in the collection of all neighborhoods of all locations. We present a framework for studying the information stored within these neighborhoods. Such matrices are very large and store information inefficiently, but they provide a useful theoretical framework for representation of imaging information. Here we propose to exploit this theoretical framework to introduce simple methods to quantify the variation in multimodal images based on the shared information across local spatial neighborhoods and subjects.

Challenges

- Imaging data is very large.
- When we consider local neighborhoods the size of these data are immediately increased by a factor of the size of the neighborhoods we consider.
- It is difficult to concisely describe this complicated data structure in a way that is useful for studies of association with health outcomes.

Here we consider the case when multimodal imaging is available for multiple subjects. Such data are routinely available in MRI studies, where FLAIR, T1-weighted and T2-weighted images are collected for multiple subjects.

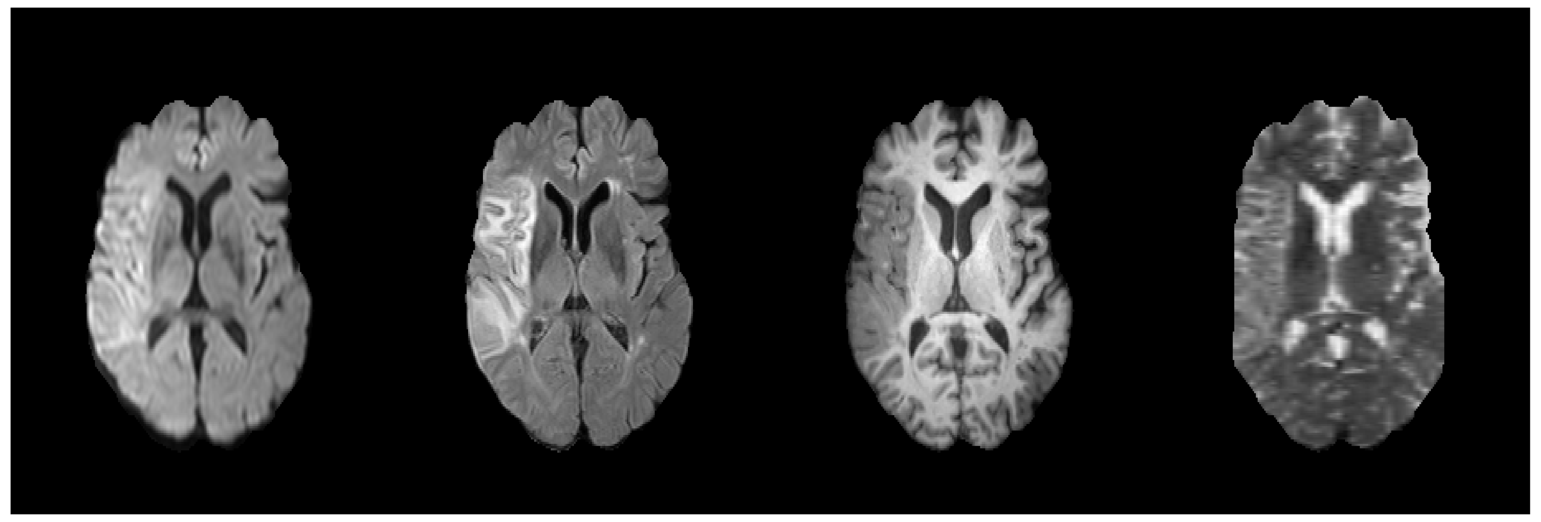

Figure 1: From left to right: DWI, FLAIR, T1w, and T2w images for Subject 05, 2015 ISLES Challenge

Objectives

1. Decompose observed variability images at the population level.
2. Describe and quantify the main directions of variation.
3. Use these directions of variation to improve segmentation and studies of association with health outcomes.

Materials and Methods

To achieve these objectives we propose to decompose the local variability of various moments of the image intensities: the image intensity, image intensity square, and so on. Consider the following mock example in 2-D:

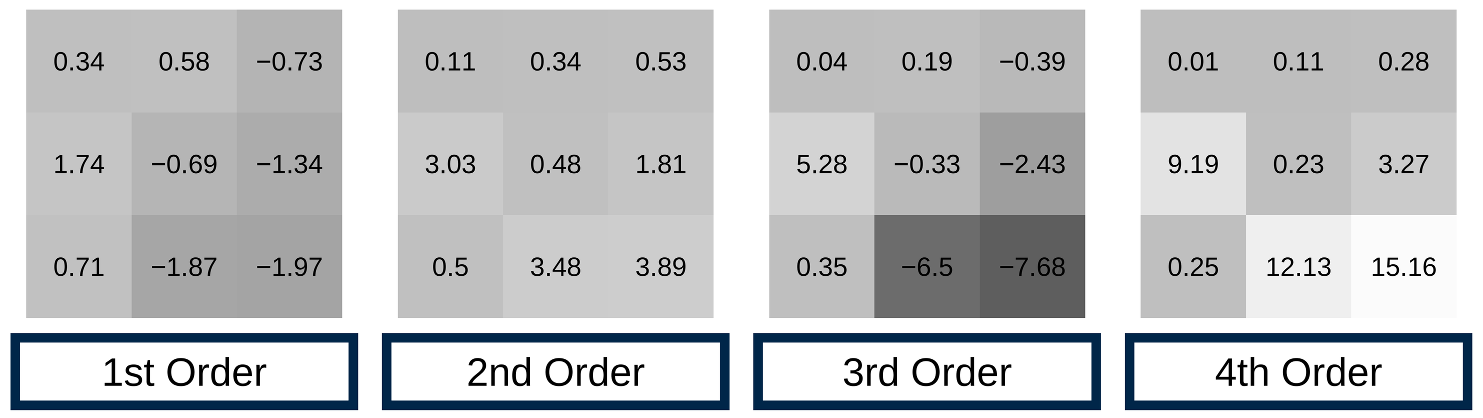

Figure 2: 2-D mock example of moments of local neighbors

$$X_{ij} = (0.34, 0.58, \ldots, -1.97, 0.11, \ldots, 3.89, 0.04, \ldots, -7.68, 0.01, \ldots, 15.16).$$

This gives the jth row for subject i. If we are interested in considering additional imaging modalities, we simply add those as additional columns in the matrix. We preform this operation for each voxel for each subject and stack these vectors into a (potentially large) matrix $\mathbf{X}$. Next,

- Center and scale the columns of $\mathbf{X}$
- Perform PCA on centered and scaled matrix.
- Calculate principal component scores.
- Use principal component scores as predictors and expert manual segmentation as outcome to train a model to perform segmentation.

Computational Issues

Since the size of the matrix $\mathbf{X}$ is large, we need to decompose this matrix without loading it into memory. To overcome this challenge, we iteratively read in each subject and calculate sufficient statistics for the principal component analysis of the correlation matrix between the columns of $\mathbf{X}$. This is implemented in the R package MEDALS (Memory Efficient Decomposition for Analysis of Local neighborhood moments for Segmentation) and is available at https://github.com/JMMaronge/medals.

Results

All data shown are from the 2015 ISLES SISS Challenge. Below we show partial ROC (pROC) curves for each subject, as well as overall training and overall testing results. We also show an example of how our method performed on a particular subject.

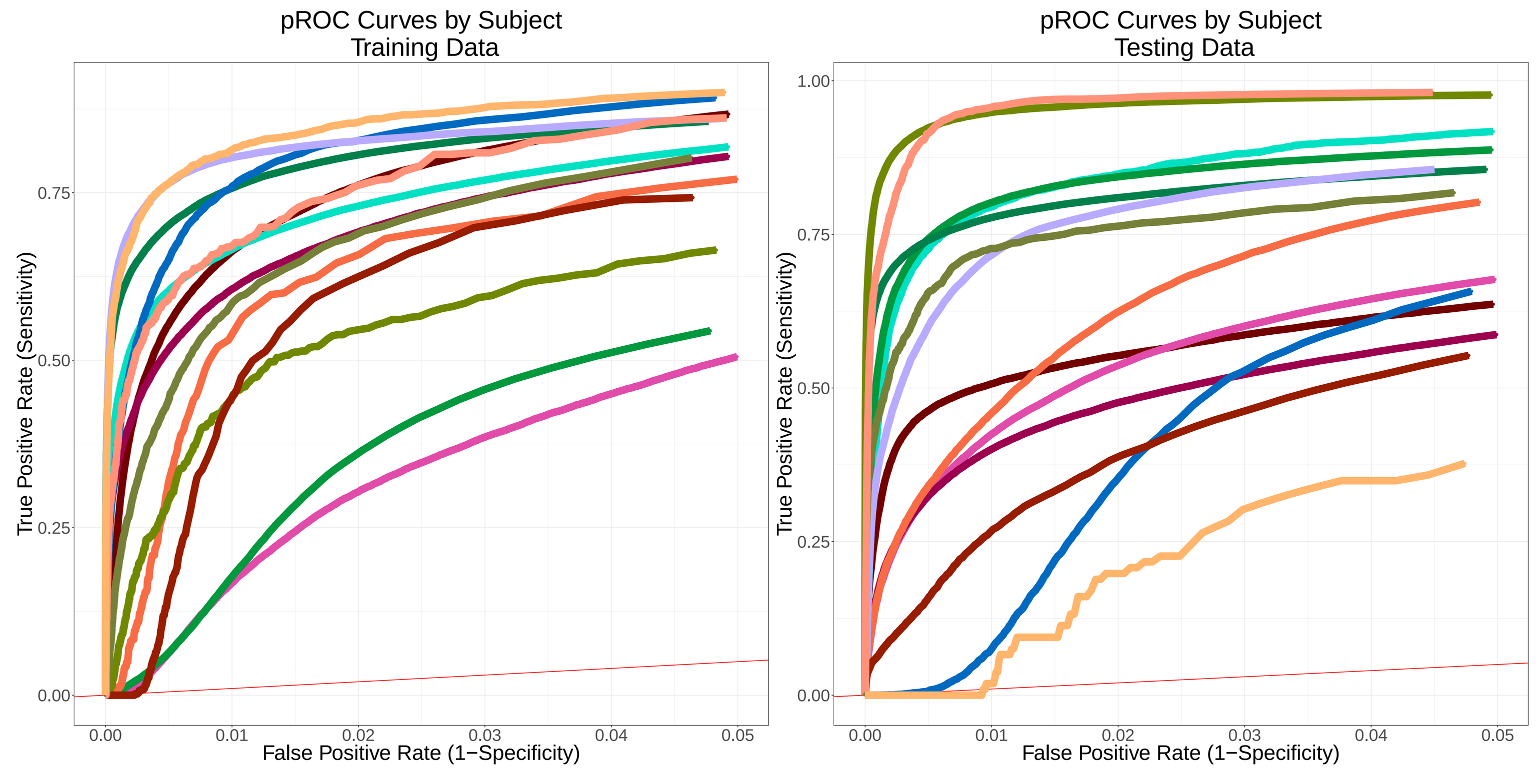

Figure 3: Subject level pROC curves for training set and testing set

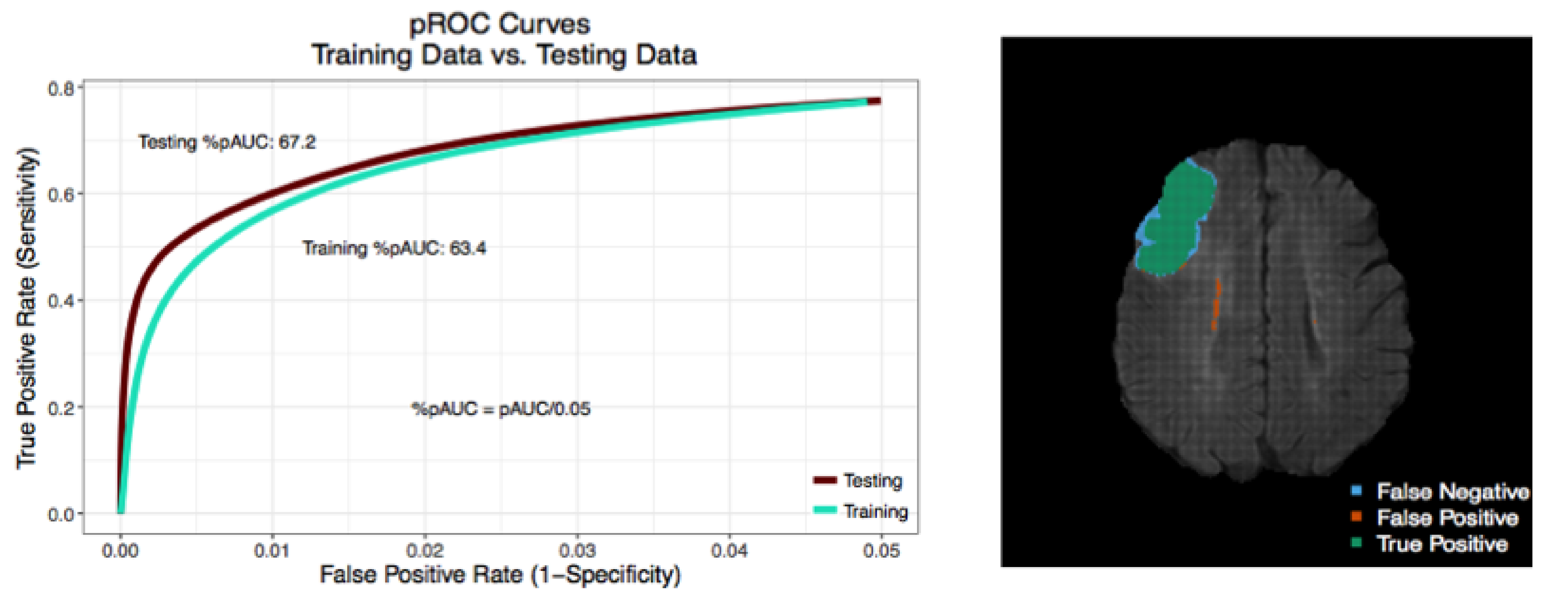

Figure 4: (left) Overall training results and overall testing results (right) Example of performance on one subject

Conclusions

- Carefully using information contained within neighborhoods allows for a richer insight of imaging data structures.
- Using local neighborhoods creates difficulties with storing the data, but these can be overcome by leveraging structure of the data.
- Characterizing variation in neighborhoods allows for interesting applications in studies of association with health outcomes.