Predicting Cognitive Change from Short-Term Longitudinal Neuroimaging

Michael H. Coen
Department of Biostatistics and Medical Informatics
Department of Computer Sciences
University of Wisconsin-Madison

Joint work with Hidayath Ansari* and Barbara Bendlin
NIPS—MLINI’13
December 10, 2013
Our Goals

• Detect subtle changes in neural structure
  – Indicative of preclinical disease
  – Reveal etiology of pathology
  – Previously undetectable

• Connect neuroimaging and clinical data

• Predict change in individuals as opposed to populations
  – From short-term longitudinal neuroimaging
Challenges

1) Datasets are *wide*

2) Naturally occurring variation
   – Overwhelmingly non-pathological

3) Lack *ground truth* to validate results
   – Waiting is not exactly practical...

4) Increasingly common to look at very short term changes
   – Is there even a signal to find?
Our Data

- Diffusion Tensor Imaging (DTI) data
  - Wisconsin Registry for Alzheimer's Prevention (WRAP)

- 75 (now 150) subjects, middle-aged, cognitively normal
- Imaged twice
  - 1.5-2 years apart, staggered groups

- 27 males, 48 female
- 49 ApoE ε4 negative, 26 ApoE ε4 positive
- 16 Fam Hx negative, 59 Fam Hx positive
Data Preprocessing

1) Registration using FMRIB’s Software Library (FSL)
2) Transform to a common space (MNI coordinates)
3) JHU white-matter tractography atlas to extract voxels
Our First Problem: Which Came First?

- Give two DTI white matter (WM) scans from the *same* patient
  - Taken roughly 2 years apart

- **Can we tell the earlier from the latter?** No other information.
Basic Strategies

• Look at them?
  – Nope

• Compare the “region-wide” means

  Accuracy: 57%

• Need more sophisticated techniques
  – Focus on sub-regions
Too Simple a Problem?

- Human experts cannot solve it
- Introduces a broader theme
  Elucidating etiology...
- Cast this as a machine learning problem

Solve this problem with in-situ statistics

Identify “responsible” voxels

Identify “responsible” regions

New Question: Do these regions predict with cognitive function?

Postulate causality from correlation
Examine the Corpus Callosum
A Simple Summary of Our Approach
Identify Interesting Voxels

• A cross-sectional and longitudinal approach
  – We select voxels that undergo significant temporal changes that are consistent and similar across subjects.

• For each voxel, consider its fractional anisotropy (FA) (or mean diffusivity (MD)).

• We define:

\[
Q(\hat{v}_i) = \frac{\text{mean}(\hat{v}_i^{t_2} - \hat{v}_i^{t_1})}{\text{variance}(\hat{v}_i^{t_2} - \hat{v}_i^{t_1}) \times \text{entropy}(\hat{v}_i^{t_2} - \hat{v}_i^{t_1})}
\]

where functions calculated cross-sectionally over all patients.
Identify Interesting Voxels

• Note $Q$ is signed. Magnitude $|Q|$ is important

• So, we want to find the set of voxels with high (or low) $Q$

$$\max \arg \{v_i\} \quad |Q(v_i)| \geq \tau$$

where $\tau$ is some threshold

• These voxels are informative \textit{longitudinally} and validated \textit{cross-sectionally}
Point Set Lifting – An Introduction

1) We treat voxels as **point sets** in a metric space.

2) We project them to **vectors in a high-dimensional reproducing kernel Hilbert space** (RKHS)

\[ \mathbb{R}^d \rightarrow \mathbb{R}^{2\rho}, \quad 2\rho \gg d \]
Random Fourier Features

Random Fourier features via projection map:

$$\phi_\rho(x) = \left[ \cos(\omega_1 x), \ldots, \cos(\omega_D x), \sin(\omega_1 x), \ldots, \sin(\omega_D x) \right]$$

where $\omega_i$'s are normally distributed

$$\phi_\rho : \mathbb{R}^d \rightarrow \mathbb{R}^{2\rho}$$

and

$$\langle \phi_\rho(x), \phi_\rho(y) \rangle \approx e^{\frac{-||x-y||^2}{2}}$$

Approximates Gaussian RBF kernel arbitrarily well for sufficient $\rho$ (Rahimi and Recht 2007)
Useful for building fast linear SVMs.
Point Sets to Points using Random Fourier Features

For a point set $C_i \subset \mathbb{R}^d$, the lifted representation is

$$\phi_\rho(C_i) = \sum_{x \in C_i} w_x \phi_\rho(x) \in \mathbb{R}^{2\rho}$$

where $w_x$'s are the weights of each voxel.

Then, the distance between point sets is:

$$d_\rho(C_1, C_2) = \left\| \frac{\phi_\rho(C_1)}{\|\phi_\rho(C_1)\|} - \frac{\phi_\rho(C_2)}{\|\phi_\rho(C_2)\|} \right\|$$

Can always select $\rho$ for given error level
(Phillips, Raman, and Venkatasubramanian 2011)
Lifting Point Sets – Example
Towards the Lift Kernel

DTI Data
Results

• We used $\tau = 0.70$
  – Provided 463 voxels for the splenium
• Obtained 97% accuracy
  – Checked with 10-fold cross validation
• Smoothing with an 8mm Gaussian kernel makes little difference

| Region                      | $|\hat{V}_{\text{CONS}}(\tau = 0.7)|$ | Accuracy |
|-----------------------------|-------------------------------------|----------|
| Corpus Callosum (whole)     | 3429 voxels                         | 96%      |
| Corpus Callosum (splenium)  | 463 voxels                          | 97.3%    |
| Corpus Callosum (genu)      | 364 voxels                          | 90.7%    |
| Cingulum bundle (R & L)     | 776 voxels                          | 97.3%    |
Visualizing the Lift Kernel on Our Data
Discussion

- Are we able to detect the kinds of WM differences we’re looking for?  Yes

- Are we able to separate between signal and noise?  Yes
Where Are the Voxels?

View of the splenium
What Do the Signs of $Q$ Mean?

View of the splenium
Along the Corpus Callosum
Examining Different Thresholds $\tau$
Correlation with Cognition

- WRAP Cohort Cognitive Tests
  1) Verbal Ability
  2) Visual Spatial
  3) Speed Flexibility
  4) Working Memory
  5) Verbal Language
  6) Immediate Memory
Predicting Cognitive Change

• Goal: construct a regression model based on FA changes to predict individual’s cognitive test scores.

• Even thresholding voxels with our $Q$ measure, it remains the case that $p \gg N$ (wide data)

• Used coordinate descent for lasso and ridge to perform logistic regression on *signs* of factors.
Why is this difficult?

Speed and Flexibility Test

Time 2 vs. Time 1

$y = 0.92x + 0.02$
A Well-Defined Problem

Can changes in neuroimaging data predict whether a subject's score for some neuropsychological test has increased or decreased?

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameters</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso logistic regression [41]</td>
<td>$\lambda = .011$</td>
<td>70%</td>
</tr>
<tr>
<td>SVM, Lifted kernel</td>
<td>$2D = 500, C = 1$</td>
<td>58%</td>
</tr>
<tr>
<td>SVM, Gaussian kernel</td>
<td>$\sigma = 1, C = 1$</td>
<td>57%</td>
</tr>
<tr>
<td>Baseline Accuracy</td>
<td></td>
<td>54%</td>
</tr>
</tbody>
</table>
Results in more detail
Towards a more sparse model

- We prefer fewer explanatory variables in a model
- Clustered voxels based on spatial proximity and $Q$

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameters</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ridge logistic regression [41]</td>
<td>$\lambda = 0.013$</td>
<td>75%</td>
</tr>
<tr>
<td>SVM, Lifted kernel</td>
<td>$2D = 500, C = 1$</td>
<td>55.7%</td>
</tr>
<tr>
<td>SVM, Gaussian kernel</td>
<td>$\sigma = 1, C = 1$</td>
<td>58.5%</td>
</tr>
<tr>
<td>Baseline Accuracy</td>
<td></td>
<td>54%</td>
</tr>
</tbody>
</table>
Summary

• First evidence that very small changes in white matter structure over a two year period can predict change in cognitive function in healthy adults.

• Introduces a new machine learning approach to longitudinal neuroimaging analysis.