Machine learning approaches for clinical neuroimaging data

London, UK
28th September 2012

Janaina Mourao-Miranda
Jane Rondina
Maria Joao Rosa
Machine Learning & Neuroimaging Lab

• We are interested in developing and applying novel machine learning techniques to the analysis of neuroimaging data.

• We focus on the diagnosis and prognosis of psychiatric disorders and on understanding affective processing in normal and patients' groups.
Neuroimaging data

• Neuroimaging includes the use of various techniques to either directly or indirectly image the structure or function of the brain.
  – **Structural neuroimaging** is used to investigate brain structure (e.g. shows contrast between different tissues: cerebrospinal fluid, grey matter, white matter).
  – **Functional neuroimaging** is used to indirectly measure brain functions (e.g. neural activity)

• Example of Neuroimaging techniques:
  – Computed Tomography (CT),
  – Positron Emission Tomography (PET),
  – Single Photon Emission Computed Tomography (SPECT),
  – Structural Magnetic Resonance Imaging (sMRI),
  – Functional Magnetic Resonance Imaging (fMRI).

• Among other imaging modalities sMRI/fMRI became largely used due to its low invasiveness, lack of radiation exposure, and relatively wide availability.
Examples of brain scans

sMRI
- one image
- high resolution
  (1 mm)

fMRI
- many images
  (e.g., every 2 sec for 5 min)
- low resolution
  (~3 mm but can be better)

Data Properties:
- High dimensionality: 50,000-500,000
- Few sample: tens to hundreds
- Highly correlated features/voxels
- Structure/Brain regions
Clinical Questions

1. **Diagnosis:**
   - Can we classify groups of subjects (e.g. patients vs. controls) using structural sMRI/fMRI scans?
   - Can we combine information from different imaging modalities and/or clinical information?
   - Are patients outliers with respect to a “normal population”?

2. **Prognosis:**
   - Can we predict who will develop a disease based on a baseline scan (e.g. fMRI, sMRI)?

3. **Treatment Response:**
   - Can we predict treatment response based on brain scans?

4. **Interpretation:**
   - How to interpret classifiers/regression weights?
   - Can we find the most relevant brain regions for diagnosis/prognosis?
Machine Learning Models

• **Classification/Regression Models**
  – Support Vector Machine (SVM)
  – Gaussian Process (GP)

• **Outlier Detection Models**
  – One-class SVM

• **Sparsity/Structured Sparsity**
  – LASSO/Elastic-Net
  – Total Variation/Sparse Total Variation
  – Sparse Laplacian

• **Multiple Kernel Learning (MKL)**
  – Elastic-Net MKL

• **Stability Selection**
General Framework

Automated diagnosis of neurological and psychiatric diseases
(modified from Kloppel et al., 2011)
Example of Applications
Patient classification as an outlier detection problem: An application of the One-Class Support Vector Machine

Janaina Mourão-Miranda\textsuperscript{a,c,*}, David R. Hardoon\textsuperscript{a,c}, Tim Hahn\textsuperscript{b,d}, Andre F. Marquand\textsuperscript{c}, Steve C.R. Williams\textsuperscript{c}, John Shawe-Taylor\textsuperscript{a}, Michael Brammer\textsuperscript{c}

Results:

- Correlation between OC-SVM predictions and Hamilton Rating Scale for Depression (HRSD) = -0.81
- 79\% of healthy subjects were detected as non-outlier.
- 52\% of depressed patients were detected as outlier.
- 89\% patients classified as non-outliers responded to treatment
- 30\% of patients classified as outliers responded to treatment
Results:

- Gaussian Process Classifier (GPC) based on the whole brain activity to neutral faces differentiated at-risk adolescents from healthy controls with 75% accuracy (sensitivity = 75%, specificity = 75%).

- GPC predictive probabilities were significantly higher for the at-risk adolescents that subsequently developed depression or anxiety than for the at-risk adolescents who remained healthy at follow-up.
In the second stage of the analysis we classified patients with an intermediate course according to the discriminating pattern obtained in the continuous versus episodic analysis.

We found that 78% of those who did not go on to develop other episodes were classified as episodic, and 65% of those who developed further episodes were classified continuous.
Task: Classify whole patterns of brain activity to pleasant versus unpleasant stimuli.

An extension of the Total Variation method was presented and assess several other structured sparsity models on accuracy, sparsity and stability.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Sparsity</th>
<th>Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Least Squares</td>
<td>83.0 ± 5.9%</td>
<td>44 ± 1%</td>
<td>41 ± 1%</td>
</tr>
<tr>
<td>T-Test (10%) + LS</td>
<td>78.6 ± 5.8%</td>
<td>4.43 ± 0.02%</td>
<td>62 ± 3</td>
</tr>
<tr>
<td>Lasso</td>
<td>85.8 ± 6.6%</td>
<td>6.4 ± 1.2%</td>
<td>64 ± 15%</td>
</tr>
<tr>
<td>Elastic Net</td>
<td>85.9 ± 6.8%</td>
<td>44.4 ± 0.2%</td>
<td>39 ± 9%</td>
</tr>
<tr>
<td>TV</td>
<td>85.0 ± 6.4%</td>
<td>42 ± 2%</td>
<td>34 ± 19%</td>
</tr>
<tr>
<td>Sparse TV</td>
<td>87.4 ± 6.2%</td>
<td>9.4 ± 0.4%</td>
<td>71 ± 3%</td>
</tr>
<tr>
<td>Laplacian (α = 0)</td>
<td>83.2 ± 5.7%</td>
<td>44.2 ± 0.1%</td>
<td>40 ± 1%</td>
</tr>
<tr>
<td>SLAP</td>
<td>85.5 ± 6.2%</td>
<td>7 ± 10%</td>
<td>52 ± 22%</td>
</tr>
</tbody>
</table>
PROBABILISTIC PREDICTION OF NEUROLOGICAL DISORDERS WITH A STATISTICAL ASSESSMENT OF NEUROIMAGING DATA MODALITIES*

By M. Filippone, A.F. Marquand C.R.V. Blain S.C.R. Williams J. Mourão-Miranda and M. Girolami

- **Task:** Discriminate three Parkinsonian neurological disorders from one another and healthy controls.

- A multinomial logit model with Gaussian process priors is proposed to: (i) predict disease state based on whole-brain neuroimaging data and (ii) analyze the relative informativeness of different image modalities and brain regions.

*Predictive accuracy (multi-source classifier). Min and max values refer to minimum and maximum values across CV folds.*

<table>
<thead>
<tr>
<th>Input data</th>
<th>Accuracy (min, max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 GM only</td>
<td>0.627 (0.321, 0.854)</td>
</tr>
<tr>
<td>2 WM only</td>
<td>0.603 (0.350, 0.771)</td>
</tr>
<tr>
<td>3 T2 only</td>
<td>0.545 (0.500, 0.604)</td>
</tr>
<tr>
<td>4 FA only</td>
<td>0.569 (0.442, 0.688)</td>
</tr>
<tr>
<td>5 MD only</td>
<td>0.623 (0.533, 0.750)</td>
</tr>
<tr>
<td>6 Weighted sum</td>
<td>0.598 (0.350, 0.708)</td>
</tr>
<tr>
<td>7 Unweighted sum</td>
<td>0.610 (0.400, 0.708)</td>
</tr>
<tr>
<td>8 SimpleMKL</td>
<td>0.418 (0.143, 0.625)</td>
</tr>
</tbody>
</table>
A new feature selection method based on stability theory - exploring parameters space to evaluate classification accuracy in neuroimaging data

Jane M Rondina\textsuperscript{1,2}, John Shawe-Taylor\textsuperscript{2}, and Janaina Mourao-Miranda\textsuperscript{1,2}
On-going work
**Elastic-Net Multiple Kernel Learning**

Dr. Janaina Mourao-Miranda

---

**Task:** Find the optimal combination of brain regions to classify depressed patients versus healthy subjects

<table>
<thead>
<tr>
<th>Regions</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postcentr_R</td>
<td>0.52632</td>
</tr>
<tr>
<td>Precentral_R</td>
<td>0.089159</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Regions</th>
<th>Accuracy (%)</th>
<th>TP (%)</th>
<th>TN (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postcentr_R</td>
<td>86.84</td>
<td>84.21</td>
<td>89.47</td>
</tr>
<tr>
<td>Precentral_R</td>
<td>84.21</td>
<td>78.95</td>
<td>89.47</td>
</tr>
</tbody>
</table>

---

Table 1: MKL Results

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole brain</td>
<td>71.05</td>
<td>68.42</td>
<td>73.68</td>
</tr>
<tr>
<td>MKL - Elastic Net*</td>
<td>86.84</td>
<td>84.21</td>
<td>89.47</td>
</tr>
</tbody>
</table>

* optimized using nested cross-validation

---

Table 2: Single region accuracies

---

Anatomical template for segmenting the brain into regions
Multi-center Classifier
Mireille Nieuwenhuis, University Medical Center Utrecht

Task: Develop a multi-center model that can predict outcome of first episode schizophrenia patients from multiple centers.

Data:
Structural MRI brain scan from first episode patients separated into clinical sub groups (continuous and episodic).
- London Sample: 28 continuous and 28 episodic.
- Utrecht Sample: 23 continuous and 23 episodic.
- Melbourne Sample: 14 continuous and 14 episodic.
Predicting clinical scores from brain images
Dr. Liana Portugal, CS/UCL

Task: Predict elevated symptoms of mania (ESM) from patterns brain activation/anatomy.

Data:
• Brain scans and neurocognitive testing acquired as part of a Longitudinal Assessment of Manic Symptoms (LAMS) study at 4-sites in USA.
• The participants in the study are children identified with elevated symptoms of mania (ESM) at the time of their first clinical presentation and age/race/sex matched comparison children without ESM.
• During the study, these participants were evaluated at baseline and every 6 months thereafter regarding their psychiatric diagnoses, psychiatric symptomatology.
SCoRS (Survival Count on Random Subspaces)
Dr. Jane Rondina, CS/UCL

- A new method based on stability selection to detect distributed patterns in neuroimaging
Generative embedding for neuroimaging
Dr. Maria Joao Rosa, CS/UCL

Generative model
(e.g. dynamic causal models, MAR models, Riccican mixtures)

Generative embedding:
(e.g. Fisher kernels, free energy scores, TOP kernel)

Generative based kernel
(e.g. linear kernel, RBF kernel, information theoretic kernels)

Advantages:
• Better interpretability of classification/regression results
• Classification/Regression based on hidden (ideally physiological) quantities
• More biologically meaningful (model based) feature extraction.

Supervised learning
(Classification (e.g. SVM) / Regression (e.g. KRR))
for clinical neuroimaging data application
PRoNTo Project
Group Members

- Dr. Janaina Mourao-Miranda (principal investigator)
- Dr Maria Joao Rosa (postdoctoral fellow)
- Dr Jane Rondina (visiting postdoctoral fellow)
- Dr Liana Portugal (visiting postdoctoral fellow)
- Dr Christophe Phillips (honorary research fellow)

PRoNTTo Team

- Janaina Mourao-Miranda (University College London)
- Jessica Schrouff (University of Liege)
- Christophe Phillips (University of Liege)
- Maria Joao Rosa (University College London)
- Jane Rondina (University College London)
- Andre Marquand (Kings College London)
- John Ashburner (University College London)
- Jonas Richiardi (University of Geneva)
- Carlton Chu (NIH)
Sponsors

- Professor John-Shawe Taylor (Computer Science Department, UCL)
- Professor Steven Williams (Centre for Neuroimaging Sciences, KCL)