A Machine Learning Analysis of the Progression of Alzheimer’s Disease Using the Alzheimer's Disease Neuroimaging Initiative (ADNI) Cohort

JSM 2014,
August 6th, 2014,

bruno.jedynak@jhu.edu

Engineering team: Andrew Lang, Bo Liu, Jerry Prince, Brian Caffo, Murat Bilgel, Runze Tan, Chun-Guang, and Zhou Ye;

Medical team: Kostas Lyketsos, Susan Resnik, Sterling Johnson, and Pierre Jedynak
Longitudinal databases of hundreds of subjects observed during several years with tens of validated biomarkers are becoming available, allowing the creative use of computational methods in neurology:

- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Parkinson Progression Marker Initiative (PPMI)
- Predict-HD, TrackOn-HD for Huntington disease
- .. Much more
- Would it be possible that a discovery in neurology come from an innovative analysis of the data?
Potential contributions

1. Insight about the disease process
   1. Is there a single mechanism under a given disease name?
   2. Validate a biomarker of disease progression

2. Provide instruments for more efficient drug discovery process
   1. Entry criteria: Define subjects which are likely to benefit from a given drug
   2. Measure precisely the disease progression

3. Help the neurologist
   1. Which biomarkers are the most informative at a given stage of a disease
   2. What can be expected if the patient is untreated
# Table of content

1. The "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade"
2. Using the ADNI dataset and statistical modeling to build an Alzheimer’s disease Progression Score (ADPS)
3. The ADPS as a composite biomarker of disease progression
4. Simultaneous experiments with the cognitive measurements of the BLSA and WRAP study
5. A first take at modeling the progression of the amyloid burden within the brain in the BLSA
6. Conclusion
Hypothetical model of dynamic biomarkers of the Alzheimer’s pathological cascade.
Principles for building an Alzheimer’s Disease Progression Scale (ADPS) for each subject in the ADNI

1. Qualitatively, there is a single disease progression for late onset AD.

2. Along the disease progression, each biomarker change continuously and monotonically from Normal status to Abnormal status, following a sigmoid curve.

3. Each subject in the ADNI is progressing at constant speed relative to its age during the time it is observed.
One dimensional model

\[ y_{ijk} = f_k(\alpha_i t_{ij} + \beta_i, \theta_k) + \sigma_k \epsilon_{ijk} \]

with

\[ f_k(s, \theta = (a, b, c, d)) = \frac{a}{1 + e^{-b(s-c)}} + d \]

- \( t_{ij} \) is the age of subject \( i \) at visit \( j \)
- \( s_{ij} = \alpha_i t_{ij} + \beta_i \) is the ADPS for subject \( i \) at visit \( j \)
- \( k \) is the index of a biomarker
- \( \epsilon_{ijk} \) are independent, standard, normal
- Ref: self-modeling \((K=1)\)
**Puzzle analogy**

Think of a puzzle which is finished. It provides
1. A picture, a scene, which was invisible when the pieces were scrambled
1. A localization for each piece

Similarly, computing the ADPS provides
1. A visualization of the biomarker values along the time-line of AD
2. A score for each time-point of each subject

**Note:**

The ADPS requires calibration in translation and scale. This calibration is performed with the Normal subjects in ADNI.
Algorithm for building an Alzheimer’s Disease Progression Score (ADPS) for each subject in the ADNI

1. Correct each biomarker for the age effect. Initialize the ADPS of each subject with its age.

2. Repeat
   A. Fit a sigmoid (=4 parameters) to each biomarker, fixing the ADPS for each subject.
   B. Fit the ADPS of each subject (=2 parameters) fixing the sigmoid for each biomarker.
   C. Fit the variance of the noise (=1 parameter) for each biomarker.

3. Standardize the ADPS of all subjects, such that the median of the normal subjects is 0 and the median absolute deviation (mad) of the normal subjects is 1
ADNI I dataset

1. 687 Subjects have MRI volumetric data and 2 to 6 visits (5 in average)

2. Expert selected biomarkers:
   A. Dementia ratings: Alzheimer’s Disease Assessment Scale (ADAS), Mini Mental State Examination (MMSE), Clinical Dementia Rating Sum of Boxes (CDRSB)
   B. CSF measurements: proteins: $A\beta_{42}$, tau
   C. MRI measurement: Hippocampus volume over intra-cranial volume (Hippo)
   D. Memory rating: Rey Auditory Verbal Learning Test, 30 min (RAVLT_30min)
Biomarkers as function of the Alzheimer’s disease Progression Score

Progression of ADNI biomarkers as function of the Alzheimer’s Disease Progression Score (ADPS)

- Normal
- MCI
- AD
- MCI-AD
- MCI-N
- sigmoids

**Biomarkers**:
- HIPPO
- MMSE
- TAU (pg/ml)
- ABETA (pg/ml)
- CDRSB
- RAVLT30
- ADAS

**Axes**:
- ADPS (Alzheimer’s Disease Progression Score)
Bootstrap analysis of the variability of the sigmoids

100 Bootstrap replicates of the estimated biomarker sigmoids
Standardized biomarkers

(a) ADPS Densities Conditioned on Clinical Status

(b) Timing of Biomarker Dynamics

- Normal
- Abnormal

- N
- MCI
- AD

- HIPPO
- ADAS
- MMSE
- TAU
- ABETA
- CDRSB
- RAVLT30

ADPS
Rate of change of the ADPS as function of the ADPS
Prediction of conversion from MCI to AD in 2 years
## Cognitive tests used in WRAP analyses

<table>
<thead>
<tr>
<th>WRAP</th>
<th>BLSA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AVLT-sum</strong></td>
<td><strong>CVLT-sum</strong></td>
</tr>
<tr>
<td>- Auditory verbal learning test</td>
<td>- California verbal learning test</td>
</tr>
<tr>
<td>- Immediate recall/total learning score summed across 5 trials</td>
<td>- Immediate recall/total learning score summed across 5 trials</td>
</tr>
<tr>
<td>- Range 0-75</td>
<td>- Range 0-80</td>
</tr>
<tr>
<td><strong>AVLT-delayed</strong></td>
<td><strong>CVLT-frs and frl</strong></td>
</tr>
<tr>
<td>- Delayed free recall score (~20 mins delay)</td>
<td>- Delayed free recall scores with short delay (after List B) and long delay (~20 mins later)</td>
</tr>
<tr>
<td>- Range 0-15</td>
<td>- Range 0-16</td>
</tr>
<tr>
<td><strong>AVLT-recognition</strong></td>
<td><strong>Benton visual retention</strong></td>
</tr>
<tr>
<td>- Recognition of list words in a paragraph</td>
<td>- Scored for errors in drawing replication</td>
</tr>
<tr>
<td>- Range 0-15</td>
<td></td>
</tr>
<tr>
<td><strong>MMSE</strong></td>
<td><strong>BMS</strong></td>
</tr>
<tr>
<td>- Mini-mental state exam</td>
<td>- Blessed information memory concentration score</td>
</tr>
<tr>
<td>- Range 0-30</td>
<td>- Scored for errors</td>
</tr>
<tr>
<td></td>
<td><strong>MMSE</strong></td>
</tr>
<tr>
<td></td>
<td>- Mini-mental state exam</td>
</tr>
<tr>
<td></td>
<td>- Range 0-30</td>
</tr>
</tbody>
</table>
Estimates for $A_k$

Estimates for $B_k$

18 subjects
48 visits total

$2 \times 2 \times 2$ mm
PiB-DVR images
(amyloid)

231k brain voxels
Some observations

- $A \approx 0$ in cerebellum.
  - Supports choice of reference tissue
  - Validity of longitudinal DVR estimates
- Sparing of pre- and postcentral gyri.
  - Literature indicates these regions accumulate amyloid in late stages.
- Precuneus, frontal, lateral parietal & temporal.
  - Usual suspects

Estimates for $A_k$
Conclusion

We have explored a statistical modeling technique aimed at better understanding the progression of AD.

We have experimented with ADNI, BLSA, and WRAP.

Thank you
References


2. Bruno M. Jedynak, Bo Liu, Andrew Lang, Yulia Gel and Jerry L. Prince, "A computational method for computing an Alzheimer's Disease Progression Score; experiments and validation with the ADNI dataset", accepted for publication in Neurobiology of Aging.

Compositional models of disease progression for all subjects, including patients and control

- $y_{ij} = f(g(t_{ij}, c_{ij}, v_{ij})) + \epsilon_{ij}$
- $i$: subject
- $j$: visit index
- $y_{ij}$: collection of measurements (features, markers, biomarkers) available
- $t_{ij}$: age of subject $i$ at visit $j$
- $v_{ij}$: treatment of subject $i$ at visit $j$
- $(t, c, v) \mapsto g(t, c, v) \in \mathbb{R}^k$
- $k$: the intrinsic dimensionality of the “disease space”
- $f: \mathbb{R}^k \mapsto \mathbb{R}^m$: dynamic of the measurements
- $c_{ij} \in \mathbb{R}^l$: vital statistics of subject $i$. Might include weight, height, intracranial volume, ...
- $\epsilon_{ij}$: centered noise.
\[ \rho = \{ \rho = (a, b, \alpha, \beta, \sigma); I^{-1} \sum_{i=1}^{I} \alpha_i = \alpha_0, I^{-1} \sum_{i=1}^{I} \beta_i = \beta_0, \\
   b_k > 0, a_k \neq 0 \text{ for all } k \in \mathcal{I} \} \]

**Theorem 1.** The model \( \{ P_\rho; \rho \in \rho \} \) is identifiable as long as the following 2 conditions are verified:

1. For each biomarker, there is at least 1 subject \( i \) with \( \alpha_i \neq 0 \) and with at least 4 distinct time-points at which this biomarker is available.
2. For each subject, there is at least 1 biomarker which is available at 2 time points.
Hypothetical progression of Parkinson's Disease: What For?

- **C** Presymptomatic Diagnosis
  - Neuroprotection

- **B** Positive Diagnosis
  - Differential Diagnosis
  - Precursor symptoms:
    - Hyposmia
    - Constipation
    - REM-Sleep Behavior
  - Onset of neurodegeneration: -5 to 10 years

- **A** Diagnosis of Progression
  - Prognosis & Treatment Optimization
  - Progression symptoms:
    - Cognitive impairment
    - Postural imbalance
    - Dysarthria
  - Motor Symptoms
  - Non-Motor Symptoms

- Time (years)
- Diagnosis
- Motor Symptoms
- Non-Motor Symptoms
- Dopaminergic neuronal loss

Premotor phase
Symptomatic PD