Speech-Based Detection of Alzheimer’s Disease in Conversational German

Jochen Weiner, Christian Herff, Tanja Schultz

Motivation
- Worldwide population is ageing rapidly
- Percentage older than 60 years
  
<table>
<thead>
<tr>
<th>Year</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950</td>
<td>8%</td>
</tr>
<tr>
<td>2013</td>
<td>12%</td>
</tr>
<tr>
<td>2050</td>
<td>21%</td>
</tr>
</tbody>
</table>

Cognitive decline will become a major challenge

- 4 – 8% prevalence of dementia (population aged 60+)
- No known cure, but early therapy delays the disease
  → Large scale cognitive status monitoring
- State of the art diagnosing is costly and takes a lot of expert time

Dementia and Speech
- Dementia affects human speech and language
- Changes occur very early
  → Promising candidate for new approaches to diagnose dementia
- Speech has been used to detect dementia
  - Carefully controlled lab conditions
  - Test recordings
  - Small number of subjects

Database: ILSE
- Biographic interviews
  - Semi-standardized
  - Several hours duration
- Cognitive diagnoses made by psychiatrists
- Four measurements
  - Over time some participants developed dementia
- Real world data set
  - Natural prevalence of dementia
  - Highly unbalanced

Selected data:
- Cognitive diagnoses available
- Transcription available

<table>
<thead>
<tr>
<th>Diagnoses</th>
<th>control</th>
<th>AACD</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>51</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>13</td>
<td>5</td>
</tr>
</tbody>
</table>

Experiment
- Extract features per interview
- Stack features
  → 10 dimensional feature vector
- 98 samples
  - unbalanced data set

- Three-class classification problem:
  control ↔ AACD ↔ AD
- LDA classifier
  - 3-fold cross-validation

Results
- Accuracy: 85.7%
- Unweighted average recall 0.66

<table>
<thead>
<tr>
<th></th>
<th>precision</th>
<th>recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>control</td>
<td>0.88</td>
<td>0.96</td>
<td>0.92</td>
</tr>
<tr>
<td>AACD</td>
<td>0.60</td>
<td>0.23</td>
<td>0.33</td>
</tr>
<tr>
<td>AD</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Voice activity detection (VAD)

Manual transcription

Die mich so scheuchen wollten nicht da bergauf berg ab das habe mir dann nicht bieten lassen ne