Prediction of Negative Symptoms of Schizophrenia from Objective Linguistic, Acoustic and Conversational Cues

Debsubhra Chakraborty
Institute for Media Innovation, Interdisciplinary Graduate School
Nanyang Technological University, Singapore

20th November 2018
Outline

- Introduction
- Design of Experiment
- System Overview
- Results
- Discussion & Conclusion
One in four people suffer from mental illness at some point in their lives\(^1\)

Nearly two-thirds of them never seek treatment\(^1\)

Current clinical assessments and treatments require highly trained clinicians

Methods are subjective and time-consuming

Can we move towards automated objective assessments?

---

\(^1\) 2001 WHO Press Release on Mental Illness
A chronic, mental disorder
Three types of symptoms: Positive, Negative and Cognitive¹
Few clinical treatments for negative symptoms²
Negative symptoms: Reduction of non-verbal behavior, such as speech, gestures, and facial expressions
Negative symptoms contribute to poor function and quality of life for patients³

Aim:
Identify objective biomarkers: disorganized speech is a key symptom

Speech Disorder in Schizophrenia

- Problems of **syntactic complexity and semantic coherence** common in schizophrenia
- Indicators of **early onset; insight into symptoms and thought disorders, and trajectory of recovery**¹
- Linguistic analysis a popular tool for investigation of schizophrenia speech², ³
- However they utilize **manual transcriptions**, which is tedious
- Comparatively, analysis of schizophrenia speech with **non-verbal features** is lesser⁴
- Combining verbal and non-verbal features together for holistic understanding of schizophrenia, akin to clinicians

Outline

- Introduction & Literature Review
- Design of Experiment
- System Overview
- Results
- Discussion & Conclusion
### Study in collaboration with Institute of Mental Health in Singapore

**Patients with Schizophrenia (Subjects) (50)**

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Patients (50)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (years)</td>
<td>30.3</td>
</tr>
<tr>
<td>Range (years)</td>
<td>20-46</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>25</td>
</tr>
<tr>
<td>Female</td>
<td>25</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>42</td>
</tr>
<tr>
<td>Malay</td>
<td>5</td>
</tr>
<tr>
<td>Indian</td>
<td>3</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>7</td>
</tr>
<tr>
<td>Diploma/Vocational</td>
<td>27</td>
</tr>
<tr>
<td>High School</td>
<td>16</td>
</tr>
</tbody>
</table>
Each participant undergoes cognitive tasks, a semi-structured interview, and functional tasks.

- No pre-determined time limit for the interview.
- No role-playing during the interview.
- On average, the interview lasts for 30 minutes.
- Each interview (audio and video) is analysed in its entirety.
The psychometric scale used is Negative Symptoms Assessment-16 (NSA-16)

One of few rating instruments specifically designed for negative symptoms

16 item scale with scores for speech behaviour, emotional behaviour, affect, movement and daily activity etc.

Ratings are on a scale of 1-6

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully normal behaviour</td>
<td>1</td>
</tr>
<tr>
<td>Behaviour minimally reduced</td>
<td>2</td>
</tr>
<tr>
<td>Behaviour mildly reduced, but noticeable by trained rater</td>
<td>3</td>
</tr>
<tr>
<td>Behaviour moderately reduced, noticeable even by untrained rater</td>
<td>4</td>
</tr>
<tr>
<td>Behaviour is markedly reduced, hampering social function</td>
<td>5</td>
</tr>
<tr>
<td>Behaviour severely reduced or absent</td>
<td>6</td>
</tr>
</tbody>
</table>
Outline

- Introduction & Literature Review
- Design of Experiment
- System Overview
- Results
- Discussion & Conclusion
- Lapel microphones records the audio of the participant and psychologists
- Audio captured in 2-channel audio file
- Patient and Psychologist seated 2m apart, cross-talk is minimal
- Speech-to-text using Kaldi toolkit

Audio Features: Linguistic

- Linguistic Inquiry and Word Count (LIWC2015)
- Extract number of linguistic words related to thoughts, feelings, personality, and motivations
- Captures people’s social and psychological state
- 78 output categories in LIWC2015

<table>
<thead>
<tr>
<th>Category name</th>
<th>Abbrev</th>
<th>Examples</th>
<th>Words in category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Adverbs</td>
<td>adverb</td>
<td>very, really, basically</td>
<td>140</td>
</tr>
<tr>
<td>Netspeak</td>
<td>netspeak</td>
<td>coz, thx, yeah</td>
<td>209</td>
</tr>
<tr>
<td>Female references</td>
<td>female</td>
<td>Mom, sister, her, she</td>
<td>124</td>
</tr>
<tr>
<td>Feel</td>
<td>feel</td>
<td>Feel, touch, warm</td>
<td>128</td>
</tr>
</tbody>
</table>

- Each category we normalized by the duration of the interview

1. https://liwc.wpengine.com/
Audio Features: Prosodic

- Prosodic cues based on openSMILE’s ‘emobase’ set
- openSMILE toolkit - a modular and flexible audio feature extractor for signal-processing applications¹
- 19 different statistics for the 26 signals and their delta values
- Total 988 features

Conversational cues capture the dynamics of conversation. Account for *who* is talking, *when* and *how much*. 14 features related to Natural Turns, Interruption, Interjection, Speaking Percentage, Mutual Silence, Response Time etc.
Outline

- Introduction & Literature Review
- Design of Experiment
- System Overview
- Results
- Discussion & Conclusion
Prediction of NSA-16 scores from Speech Features

- NSA-16 ratings are from 1-6, but not all ratings are equally frequent
- Ratings regrouped into 2 classes
  - Ratings 1 & 2 → class 0 (‘Low’ symptoms: unobservable)
  - Ratings 3, 4, 5 & 6 → class 1 (‘High’ symptoms: observable)
- Classification
  - Features: Linguistic LIWC features + openSMILE Prosody features, Conversational features
  - Labels: NSA-16 ratings as target
  - Classifier: Linear SVM with Stochastic Gradient Descent Optimization
  - Feature Selection: Kruskal-Wallis test based feature selection on validation set, nested inside leave-one-out cross-validation

NSA ratings (High/Low)
**Prediction: Prosodic Signals**

- Several NSA-16 items related to *speech* and *emotion* can be predicted with reasonably high accuracy

<table>
<thead>
<tr>
<th>NSA item</th>
<th>Confusion Matrix</th>
<th></th>
<th>Accuracy</th>
<th>Baseline Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSA 1: Prolonged Time to Respond</td>
<td>True class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>8</td>
<td>8</td>
<td>82.00%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>1</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>NSA 2: Restricted Speech Quantity</td>
<td>True class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>14</td>
<td>7</td>
<td>78.00%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>NSA 3: Impoverished Speech Content</td>
<td>True class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>15</td>
<td>9</td>
<td>64.00%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>9</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>NSA 6: Affect Reduced Modulation of Intensity</td>
<td>True class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>17</td>
<td>8</td>
<td>72.00%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>6</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>NSA 15: Reduced Expressive Gestures</td>
<td>True class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>10</td>
<td>8</td>
<td>72.00%</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>6</td>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>
Outline

- Introduction & Literature Review
- Design of Experiment
- System Overview
- Results
- Discussion & Conclusion
Discussion and Conclusion

- Items related to even emotion and gestures are accurately predicted
- Persons who speak less, also make less gestures
- Symptoms are highly inter-related → disruption of cognitive processes in patients
- Mentally ill individuals often have (subtle) differences in social behavior
- This paves the way for holistic objective assessment of negative symptoms, and can be translated to practical tools for (apps, social robots)
  - Long-term assessments of mentally ill people
  - Evaluations of treatments.
- Plan: We have just started a similar 3-year study related to depression (funded by Rehabilitation Research Institute Singapore--RRIS)
Thank You
Objective:
Automated linguistic analysis of the interviews with schizophrenia patients

Speech recognition (\textit{Speech to Text})

\textbf{Kaldi} is an open source speech recognition toolkit which could enable the recognition and translation of spoken language into text by computers. (offline)
Features (Dictionary based)

- **Linguistic Inquiry and Word Count** (LIWC2015) to extract number of linguistic words related to thoughts, feelings, personality, and motivations.
  - 78 output features (e.g., assent, work, feeling, religion)
  - capture people’s social and psychological states
  - Normalized by the duration of interviews

- **Diction 7.0**: 42 features
  - 31 sub-features (e.g., Praise, Satisfaction, Denial)
  - 5 core features (Activity, Optimism, Certainty, Realism and Commonality)
  - 4 calculated features (e.g., Insistence = [Number of Eligible Words x Sum of their Occurrences] ÷ 10)
  - 2 common features (avg word size, num of unique words)