Prediction of Negative Symptoms of Schizophrenia from Facial Expressions and Speech Signals

Debsubhra CHAKRABORTY
PhD student
Institute for Media Innovation/ Interdisciplinary Graduate School

Supervisor: Assoc. Prof. Justin Dauwels (EEE)
Co-supervisor: Prof. Nadia Magnenat-Thalmann (IMI)

21st February, 2018
Outline

• Introduction
• Design of Experiment
• System Hardware and Features
• Results and Discussion
• Conclusion
Assessment and Treatments in Psychiatry

- 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
Schizophrenia

• A chronic, mental disorder
• Two types: Positive and negative symptoms
• Few clinical treatments for negative symptoms
• 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 patients

Aim:
Identify objective cognitive/sociological biomarkers: speech/facial expression signals

Research Questions

1. Can the **objective** measurements be used to **predict** the **subjective** ratings?
   
   *This would enable automated long-term assessments of negative symptoms*

2. Can **objective** measures **differentiate** between **patients** and **healthy** individuals?
   
   *This would enable more objective diagnosis of mental illnesses*
Outline

• Introduction

• Design of Experiment

• System Hardware and Features

• Results and Discussion

• Conclusion
Demographics

- Study in collaboration with Institute of Mental Health in Singapore
- 2 types of participants
  – *Schizophrenia patients* (56)
  – *Healthy (Controls)* (26)

<table>
<thead>
<tr>
<th></th>
<th>Patient (56)</th>
<th>Controls (26)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (years)</td>
<td>31.3</td>
<td>29.6</td>
</tr>
<tr>
<td>Range (years)</td>
<td>20-51</td>
<td>19-47</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>26</td>
<td>12</td>
</tr>
<tr>
<td>Female</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>48</td>
<td>22</td>
</tr>
<tr>
<td>Malay</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Indian</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Diploma/ Vocational</td>
<td>29</td>
<td>15</td>
</tr>
<tr>
<td>High School</td>
<td>19</td>
<td>7</td>
</tr>
</tbody>
</table>
Study Protocol

- Each participant undergoes cognitive tasks, a **semi-structured interview**, and functional tasks
- **No** pre-determined **time-limit** for the interview
- **No role-playing** during interview
- On average, interview lasts for **25 minutes**
- Each interview (audio and video) analysed in its entirety
Assessment Scale

- 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>
NSA-16 Items: Summary

- **NSA 1**: Prolonged time to respond
- **NSA 2**: Restricted speech quantity
- **NSA 3**: Impoverished speech content
- **NSA 4**: Inarticulate speech
- **NSA 5**: Emotion Reduced Range
- **NSA 6**: Affect: Reduced modulation of intensity
- **NSA 7**: Affect: Reduced display on demand
- **NSA 8**: Reduced social drive
- **NSA 9**: Poor rapport with interviewer
- **NSA 10**: Interest in Emotional and Physical Intimacy
- **NSA 11**: Poor grooming and hygiene
- **NSA 12**: Reduced sense of purpose
- **NSA 13**: Reduced interests
- **NSA 14**: Reduced daily activity
- **NSA 15**: Reduced expressive gestures
- **NSA 16**: Slowed movements
- **NSA 17**: Global negative symptoms rating
Outline

• Introduction

• Design of Experiment

• **System Hardware and Features**

• Results and Discussion

• Conclusion
System Hardware

- Lapel microphones record the audio of the participant and psychologists.
- Microsoft Kinect® captures RGB+D of the participant during interviews.
- Camera placed 2m in front of participants.
- Helps identify and track the body skeleton.
Audio Signals: Prosodic

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

Audio Signals: Conversational

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

- Affectiva – facial emotion recognition toolkit
- Analyses expressions through ‘landmarks’ on the face (FACS)
- Best performance with frontal head pose, rotation within ±25°
- 45 emotions from each frame, 90 features with Mean and SD
Outline

• Introduction

• Design of Experiment

• System Hardware and Features

• Results and Discussion

• Conclusion
Prediction of NSA-16 scores from Prosodic signals

- NSA-16 ratings are from 1-6, but not all ratings are equally frequent
- Ratings regrouped into 2 classes
  - Ratings 1 & 2 → class 0 (unobservable)
  - Ratings 3, 4, 5 & 6 → class 1 (observable)
- Classification
  - Features: openSMILE Prosodic signals
  - Labels: NSA-16 ratings as target
Prediction: Prosodic Signals

- Several NSA-16 items can be predicted from Prosodic signals

<table>
<thead>
<tr>
<th>NSA-dimension</th>
<th>Confusion Matrix</th>
<th>Accuracy</th>
<th>Baseline Accuracy</th>
<th>Algorithm</th>
<th>Feature-Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSA 15: Reduced Expressive Gestures</td>
<td></td>
<td></td>
<td></td>
<td>Adaboosted Decision Trees</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>True class</td>
<td>1 14 4</td>
<td>84.6%</td>
<td>65.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 4 30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSA 1: Prolonged Time to Respond</td>
<td></td>
<td></td>
<td></td>
<td>kNN</td>
<td>F-score</td>
</tr>
<tr>
<td>True class</td>
<td>1 10 7</td>
<td>82.7%</td>
<td>67.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 2 33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSA 2: Restricted Speech Quantity</td>
<td></td>
<td></td>
<td></td>
<td>Gaussian SVM</td>
<td>PCA</td>
</tr>
<tr>
<td>True class</td>
<td>1 17 4</td>
<td>82.7%</td>
<td>59.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 5 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSA 3: Impoverished Speech Content</td>
<td></td>
<td></td>
<td></td>
<td>Linear SVM</td>
<td>Linear SVM</td>
</tr>
<tr>
<td>True class</td>
<td>1 20 5</td>
<td>80.8%</td>
<td>51.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 5 22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSA 6: Affect Reduced Modulation of Intensity</td>
<td></td>
<td></td>
<td></td>
<td>Adaboosted Decision Trees</td>
<td>Decision Trees</td>
</tr>
<tr>
<td>True class</td>
<td>1 21 5</td>
<td>78.8%</td>
<td>50.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 6 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Accepted in ICASSP 2018.*
Classification of Participants

- Objective:
  Detect from the audio/video recordings whether patient or healthy (control)
- Prosodic, conversational and facial expression signals as features of classifier
- Participant class as target label (Patient vs. Control)
Classification Results

- Participant class (Healthy/Patient) as targets
- Audio & facial expression signals individually can distinguish participants
- Combined signals leads to the best results

<table>
<thead>
<tr>
<th>Feature-set</th>
<th>True class</th>
<th>Confusion Matrix</th>
<th>Accuracy</th>
<th>Baseline Accuracy</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Predicted class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Patient</td>
<td>40</td>
<td>6</td>
<td>73.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>12</td>
<td>11</td>
<td>66.7% Lin. SVM w/ Gradient Descent</td>
</tr>
<tr>
<td>Facial Emotions</td>
<td>True class</td>
<td>Patient</td>
<td>40</td>
<td>12</td>
<td>76.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>6</td>
<td>20</td>
<td>66.7% Lin. SVM w/ Gradient Descent</td>
</tr>
<tr>
<td>Speech: Prosodic</td>
<td>True class</td>
<td>Patient</td>
<td>43</td>
<td>11</td>
<td>71.2% Multilayer Perceptron</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>12</td>
<td>14</td>
<td>67.5%</td>
</tr>
<tr>
<td>Speech: Conversational</td>
<td>True class</td>
<td>Patient</td>
<td>37</td>
<td>11</td>
<td>84.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>4</td>
<td>19</td>
<td>66.7% Lin. SVM w/ Gradient Descent</td>
</tr>
<tr>
<td>All combined</td>
<td>True class</td>
<td>Patient</td>
<td>37</td>
<td>6</td>
<td>66.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>4</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

*Submitted to EMBC 2018.*
Outline

- Introduction
- Design of Experiment
- System Hardware and Features
- Results & Discussion
- Conclusion
Conclusions and Future Work

- Mentally ill individuals often have (subtle) differences in social behavior
- Objective audio/video signals can be utilized to detect such variations
- 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)

Mental Illness
- Schizophrenia
- Depression

Symptoms
- Negative symptoms
- Neurocognitive impairments
- Social cognitive deficits

Objective cues
- Prosody
- Conversation
- Facial expressions
THANK YOU for listening - ANY QUESTIONS?