Digital Phenotyping

Jukka-Pekka “JP” Onnela
Department of Biostatistics
Harvard T. H. Chan School of Public Health

July 20, 2016
INTRODUCTION

• Onnela Lab

• Statistical network science

• Digital phenotyping

www.hsph.harvard.edu/onnela-lab/
• 64% of American adults own a smartphone in 2015
• Up from 35% in 2011


U.S. Smartphone Use in 2015, Pew Research Center; 2015
DIGITAL PHENOTYPING

• **Digital Phenotyping Project** (funded by 2013 NIH Director’s New Innovator Award):
  1. Develop *customizable, scalable, open, research platform* for collecting smartphone data
  2. Develop *statistical methods* for analyzing and modeling the data

• Digital phenotyping
  • Definition: “Moment-by-moment quantification of the individual-level human phenotype *in situ* using data from personal digital devices”
  • Goal: Capture the lived experiences of subjects, and their interactions with the surrounding world, with minimal interference
  • Active data & passive data
Overall goal

• Tools → Data → Theory → Understanding

Scientific goals

• Precise disease phenotypes for psychiatric, neurological, and surgical patients
• Classification of psychopathologies based on observable behavior
• Monitor response to treatment or intervention
• More efficient drug trials
• Exposure to environment
• Depth of genotype vs. depth of phenotype (GWAS)
• Deep phenotyping (longitudinal)
• Completed development of the **Beiwe research platform**:

  1. Manage study and data collection using the web-based study portal
  2. Download app (Android & iOS)
  3. Store hashed and encrypted data on cloud server
  4. Model and analyze data using open source Beiwe data tools
  5. Share “data protocol” and publish results

DATA STREAMS

Active data
• Survey responses and metadata
• Voice / audio recordings
• Cognitive tests (under development)

Passive data
• GPS
• Accelerometer
• Phone and screen state
• WiFi routers
• Bluetooth devices
• Phone call logs
• Text message logs
• Magnetometer
• Proximity
• Map: https://mkiang.cartodb.com/viz/c67b3202-2023-11e5-96ef-0e853d047bba/public_map
• Animation: http://cdb.io/1GvZefN
Surveys (like PHQ-9) are the standard approach in mental health
- Completed during office visits
- Difficult to collect high frequency longitudinal data
- Rely on retrospective recollection and may be inaccurate
- Conformation to expectations or avoidance of responses
- This is an “in vitro” as opposed to an “in vivo” approach
PILOT STUDY

- **Used Mindful Moods app by John Torous**
- Micro-surveys only
- 13 subjects run for 30 days (29 days for 2 subjects)
- Outpatients with a diagnosis of depression
- 3 micro-surveys per day, 3 questions each, with replacement
- Own phones
- Incentive ($50 for 30 days)

PILOT: PHQ-9 ESTIMATES

- Average app score is on average 3.0 points higher than paper based score
- Average paper and application PHQ-9 scores strongly correlated: Pearson correlation coefficient of 0.84 with 95% CI: (0.55, 0.95)

• *Existing approach 1: Use dedicated GPS receivers*
  • Data collected continuously and therefore no missingness
  • Poor scalability and poor long-term adherence

• *Existing approach 2: Use smartphone GPS*
  • Needs to be sampled to avoid battery drainage
  • Ignore missingness or use linear interpolation

• *Our approach: Use smartphone GPS and deal with missingness*
  • Scientific opportunity: scalable for medical and public health applications
  • Statistical opportunity: principled way to deal with missingness
MOBILITY TRACES

- High frequency GPS trajectory converted to a mobility trace: (1) flights, (2) pauses, (3) time, (4) spatial scale
MOBILITY TRACES
MISSING DATA

• Complete mobility trace vs. simulated missingness
• Typical sampling cycle on our platform: on-cycle = 2 mins, off-cycle = 10 mins; 83.3% of mobility trace missing
SIMULATED TRACES
### MOBILITY METRICS

<table>
<thead>
<tr>
<th>Measures</th>
<th>TL.1</th>
<th>TL.10</th>
<th>TL.20</th>
<th>GL.1</th>
<th>GL.10</th>
<th>GL.20</th>
<th>GLC.1</th>
<th>GLC.10</th>
<th>GLC.20</th>
<th>LI</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hometime</td>
<td>831.5±2.3</td>
<td>832.3±2.4</td>
<td>833.4±2.2</td>
<td>830.5±2.8</td>
<td>829.8±1.9</td>
<td>829.1±2.1</td>
<td>832.1±2.2</td>
<td>831.3±2.5</td>
<td>826.7</td>
<td>882.8</td>
<td></td>
</tr>
<tr>
<td>DistTravelled</td>
<td>22184±969.7</td>
<td>22446±843.5</td>
<td>22569±811.6</td>
<td>18801±337.5</td>
<td>18779±369.4</td>
<td>21791±969.9</td>
<td>22380±712.1</td>
<td>22444±645.6</td>
<td>17236</td>
<td>19344</td>
<td></td>
</tr>
<tr>
<td>RoG</td>
<td>2787.3±2.3</td>
<td>2791.3±2.6</td>
<td>2791.2±1.9</td>
<td>2783.0±1.6</td>
<td>2783.3±2.5</td>
<td>2785.6±1.3</td>
<td>2787.0±1.5</td>
<td>2787.5±1.8</td>
<td>2779.4</td>
<td>2781.3</td>
<td></td>
</tr>
<tr>
<td>MaxDiam</td>
<td>6717±169</td>
<td>6745±129</td>
<td>6727±98</td>
<td>6494±44</td>
<td>6483±8</td>
<td>6496±34</td>
<td>6516±55</td>
<td>6517±55</td>
<td>6562</td>
<td>6479</td>
<td>6467</td>
</tr>
<tr>
<td>MaxHomeDist</td>
<td>6372±165</td>
<td>6410±123</td>
<td>6379±93</td>
<td>6160±49</td>
<td>6147±16</td>
<td>6153±39</td>
<td>6144±30</td>
<td>6152±5</td>
<td>6163</td>
<td>6149</td>
<td>6129</td>
</tr>
<tr>
<td>SigLocsVisited</td>
<td>3.96±0.73</td>
<td>3.20±0.58</td>
<td>3.20±0.71</td>
<td>3.16±0.69</td>
<td>3.00±0.76</td>
<td>2.96±0.79</td>
<td>3.28±0.61</td>
<td>3.12±0.60</td>
<td>3.20±0.65</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>AvgFlightLen</td>
<td>172.7±10.7</td>
<td>160.2±7.6</td>
<td>158.6±7.4</td>
<td>200.2±23.2</td>
<td>193.2±19.2</td>
<td>191.7±18.1</td>
<td>129.9±13.6</td>
<td>122.8±6.1</td>
<td>127.1</td>
<td>478.8</td>
<td>251.2</td>
</tr>
<tr>
<td>StdFlightLen</td>
<td>152.9±30.8</td>
<td>125.8±10.1</td>
<td>123.2±5.5</td>
<td>213.4±51.5</td>
<td>205.8±36.3</td>
<td>202.7±43.5</td>
<td>151.0±30.0</td>
<td>134.2±8.4</td>
<td>137.1</td>
<td>639.6</td>
<td>223.3</td>
</tr>
<tr>
<td>AvgFlightDur</td>
<td>79.0±9.3</td>
<td>69.4±5.8</td>
<td>68.8±5.6</td>
<td>119.0±17.9</td>
<td>115.2±13.4</td>
<td>113.5±13.7</td>
<td>65.4±10.5</td>
<td>57.2±4.1</td>
<td>60.0±5.1</td>
<td>340.6</td>
<td>77.0</td>
</tr>
<tr>
<td>StdFlightDur</td>
<td>131.7±17.0</td>
<td>115.3±9.0</td>
<td>113.5±10.2</td>
<td>170.3±22.0</td>
<td>168.7±14.8</td>
<td>166.7±14.4</td>
<td>103.7±18.2</td>
<td>85.0±10.9</td>
<td>91.7±13.1</td>
<td>289.8</td>
<td>55.2</td>
</tr>
<tr>
<td>FracPause</td>
<td>0.88±0.01</td>
<td>0.89±0.01</td>
<td>0.89±0.01</td>
<td>0.87±0.01</td>
<td>0.87±0.01</td>
<td>0.87±0.01</td>
<td>0.87±0.01</td>
<td>0.88±0.01</td>
<td>0.88±0.01</td>
<td>0.86</td>
<td>0.93</td>
</tr>
<tr>
<td>SigLocEntropy</td>
<td>0.63±0.01</td>
<td>0.63±0.01</td>
<td>0.63±0.01</td>
<td>0.63±0.01</td>
<td>0.63±0.01</td>
<td>0.63±0.01</td>
<td>0.63±0.01</td>
<td>0.63±0.01</td>
<td>0.63±0.01</td>
<td>0.63</td>
<td>0.63</td>
</tr>
<tr>
<td>MinsMissing</td>
<td>1243</td>
<td>1243</td>
<td>1243</td>
<td>1243</td>
<td>1243</td>
<td>1243</td>
<td>1243</td>
<td>1243</td>
<td>1243</td>
<td>1243</td>
<td>92</td>
</tr>
<tr>
<td>CirdnRtn</td>
<td>0.64±0.02</td>
<td>0.63±0.01</td>
<td>0.63±0.02</td>
<td>0.67±0.01</td>
<td>0.67±0.01</td>
<td>0.65±0.02</td>
<td>0.66±0.01</td>
<td>0.66±0.01</td>
<td>0.66±0.02</td>
<td>0.69</td>
<td>0.66</td>
</tr>
<tr>
<td>WkEndDayRtn</td>
<td>0.76±0.02</td>
<td>0.76±0.01</td>
<td>0.76±0.01</td>
<td>0.77±0.01</td>
<td>0.78±0.01</td>
<td>0.76±0.02</td>
<td>0.76±0.01</td>
<td>0.77±0.01</td>
<td>0.77±0.01</td>
<td>0.81</td>
<td>0.79</td>
</tr>
</tbody>
</table>
THANK YOU!

www.hsph.harvard.edu/onnela-lab/