Motivation

• Recognize personality
  – From written language
  – From conversations

➤ Improve user modeling in computer systems
  • Dialogue systems
  • Virtual agents
  • Intelligent tutoring systems
The Big Five Personality Traits

- Most essential personality traits?
- Factor analysis of descriptors
  → 5 dimensions (Norman, 1963)

- **Extraversion**
  - Sociability, assertiveness vs. quietness

- **Emotional stability**
  - Calmness vs. neuroticism, anxiety

- **Agreeableness**
  - Kindness vs. unfriendliness

- **Conscientiousness**
  - Need for achievement, organization vs. impulsiveness

- **Openness to experience**
  - Imagination, insight vs. conventionality
Personality Correlates for Recognition

- **Attitude toward machines** (Sigurdsson, 1991)
  - E.g. neurotics have problems using computers

- **Academic motivation** (Komarraju & Karau, 2005)
  - Extravert and open students are more engaged in learning, conscientious achieve more
  - Training systems

- **Leadership** (Hogan et al., 1994)
  - High on extraversion, stability, conscientiousness and openness
  - Leader identification in meetings

- **Relationship success** (Donnellan et al., 2004)
  - E.g. both partners high on openness to experience
  - Partner matching in dating websites
Language and Personality

- **Linguistic markers of extraversion** (Furnham, 1990)
  - Talk more, faster, louder and more repetitively
  - Lower type/token ratio
  - More positive emotion words (Pennebaker & King, 1999)
    - E.g. happy, pretty, good

- **Emotional instability** (Pennebaker & King, 1999)
  - 1st person singular pronouns

- **Conscientiousness** (Pennebaker & King, 1999)
  - Fewer negations and negative emotion words

- **Low but significant correlations**
  - What about non-linear relations?
  - No-one has tried to recognize personality on unseen subjects
Data driven approach:

1. Collect individual corpora
2. Collect associated personality ratings
3. Extract features from the texts
4. Build statistical models of the personality ratings
5. Test the models on unseen individuals
Methodology

Data driven approach:

1. Collect individual corpora
2. Collect associated personality ratings
3. Extract features from the texts
4. Build statistical models of the personality ratings
5. Test the models on unseen individuals
Corpus 1: Stream of Consciousness Essays  
(Pennebaker & King, 1999)

- 2,479 essays over 7 years (1.9M words)
- Self-report personality assessment
  - Five Factor Inventory questionnaire (John et al., 1991)

<table>
<thead>
<tr>
<th>Introvert</th>
<th>Extravert</th>
</tr>
</thead>
<tbody>
<tr>
<td>I’ve been waking up on time so far. What has it been, 5 days? Dear me, I’ll never keep it up, being such not a morning person and all. But maybe I’ll adjust, or not. [...]</td>
<td>I feel like I was born to do BIG things on this earth. But who knows... There is this Persian party today. My neck hurts. [...]</td>
</tr>
</tbody>
</table>

I see myself as someone who...

1. ...Is talkative

Disagree 1  2  3  4  5 Agree
96 participants recorded for 2 days, wearing an Electronically Activated Recorder (EAR)

- Self-report personality ratings
- Averaged personality ratings from 7 observers ($r = 0.84, p < 0.01$)

<table>
<thead>
<tr>
<th>Introvert</th>
<th>Extravert</th>
</tr>
</thead>
</table>
| - I don't know man, it is fine I was just saying I don't know.  
- I was just giving you a hard time, so.  
- I don't know.  
- I will go check my e-mail.  
- I said I will try to check my e-mail, ok. | - Oh, this has been happening to me a lot lately. Like my phone will ring. It won't say who it is. It just says call. And I answer and nobody will say anything. So I don't know who it is.  
- Okay. I don't really want any but a little salad. |
Datasets Comparison

- Essays or conversations?
- Self reports or observer reports?

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Self reports</th>
<th>Observer reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written language</td>
<td>Yes</td>
<td>?</td>
</tr>
<tr>
<td>Spoken language</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Methodology

Data driven approach:

1. Collect individual corpora
2. Collect associated personality ratings
3. Extract features from the texts
4. Build statistical models of the personality ratings
5. Test the models on unseen individuals
Automatic Feature Extraction

• Utterance type (initiative)
  – Utterance tags based on parse tree
    • Command, back-channel, question or assertion (Walker & Whittaker, 1990)

• Content and syntax
  – LIWC categories (Pennebaker & Francis, 2001)
    • E.g. Positive emotion words, swear words, 1st person pronouns
  – MRC Psycholinguistic database (Coltheart, 1981)
    • E.g. Familiarity, age of acquisition, concreteness

• Prosody
  – Voice pitch, intensity and speech rate
Data driven approach:

1. Collect individual corpora
2. Collect associated personality ratings
3. Extract features from the texts
4. Build statistical models of the personality ratings
5. Test the models on unseen individuals
Statistical Personality Modelling

- Regression problem?
  - E.g. extraversion = 4.3 on a 1-5 scale
  - Linear regression, regression trees

- Classification problem?
  - E.g. introvert vs. extravert
  - Decision tree, Naïve Bayes, Nearest Neighbour, SVM

- Depends on task and adaptation capabilities
• Ranking problem?
  – E.g. X is more extravert than Y

  ➢ RankBoost (Freund et al. 2003)
  – Non-linear model using boosting
  – Computes a ranking score for each instance
  – Minimizes the ranking error in the training data
    • percentage of misordered instance pairs

Extravert

Introvert

A

B

C

33.3% ranking error

Ranking model

A

B

C
Data driven approach:

1. Collect individual corpora
2. Collect associated personality ratings
3. Extract features from the texts
4. Build statistical models of the personality ratings
5. Test the models on unseen individuals
Regression Results - Essays

- **Baseline:** average personality score
- **Accuracy metric:** improvement (%) over the baseline’s absolute error
- **10 fold cross validation**
  - 90% of the data for training / 10% for testing
- **Results with self-reports:**
  Models outperform the baseline for all traits \((p < 0.05)\)
- **BUT very small improvement**
  - Between 0.7% (Extraversion) and 6.2% (Openness)
- ➢ What if we model spoken language?
Regression Results - Conversation

• Conversation data with self-reports
  – Never significantly outperform the baseline

• Conversation data with observer ratings

<table>
<thead>
<tr>
<th></th>
<th>Improvement</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraversion</td>
<td>23.20%</td>
<td>M5’ model tree</td>
</tr>
<tr>
<td>Emotional stability</td>
<td>3.92%</td>
<td>M5’ regression tree</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>14.75%</td>
<td>M5’ regression tree</td>
</tr>
<tr>
<td>Openness</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>
Regression Tree for Conscientiousness

Swear words (E.g. damn, f**k, sh*t)

- ≤ 0.93
- > 0.93

Pronouns

- ≤ 16.7
- > 16.7

Sexuality words (E.g. lust, horny)

- ≤ 0.62
- > 0.62

Comm. words

- ≤ 1.46
- > 1.46

Syllables per word

- ≤ 1.14
- > 1.14

Body states words (E.g. ache, heart, cough)

- ≤ 0.59
- > 0.59

- 2.96
- 2.98
Binary Classification Results – Conversation

- Observer reports
- Accuracy metric: correct classifications (%)
- Baseline: majority class (~ 50%)
- Naïve Bayes best model for all traits

<table>
<thead>
<tr>
<th>Trait</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraversion</td>
<td>73.20●</td>
</tr>
<tr>
<td>Emotional stability</td>
<td>70.71●</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>55.08</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>65.68●</td>
</tr>
<tr>
<td>Openness</td>
<td>56.53</td>
</tr>
</tbody>
</table>

● significantly better than the baseline (two-tailed, p < 0.05)
Decision Tree for Extraversion

- 67.26% accuracy
- Better than baseline (p < 0.05)

![Decision Tree Diagram]

- Word count
  - ≤ 1284
  - > 1284

- Metaphysical issues (E.g. God, heaven, coffin)
  - ≤ 0.25
  - > 0.25

- Commas
  - ≤ 8.72
  - > 8.72

  - Eating
    - ≤ 0.51
    - > 0.51
      - Introvert
      - Sad (E.g. grief, cry, sad)
        - ≤ 0.15
        - > 0.15
          - Introvert
          - Extravert

  - Articles
    - ≤ 3.51
    - > 3.51
      - Extravert

- Space (E.g. around, over, up)
  - ≤ 3.22
  - > 3.22

- Frequency of use
  - ≤ 6072
  - > 6072

Francois Mairesse & Marilyn Walker, University of Sheffield
Ranking Results

- Baseline: random ranking (ranking error = 0.50)
- Paired t-test on a 10 fold cross-validation (two-tailed, p < 0.05)
- Self-reports models never outperform the baseline

- Observer models perform significantly better for all traits!

<table>
<thead>
<tr>
<th>Trait</th>
<th>Ranking error</th>
<th>Feature set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraversion</td>
<td>0.26</td>
<td>Prosody</td>
</tr>
<tr>
<td>Emotional stability</td>
<td>0.39</td>
<td>MRC</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>0.31</td>
<td>All</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>0.33</td>
<td>All</td>
</tr>
<tr>
<td>Openness</td>
<td>0.37</td>
<td>LIWC</td>
</tr>
</tbody>
</table>
• Observed extraversion with prosodic features
  – Extraverts speak more, faster, with higher pitch
  – Introverts’ voice pitch and intensity vary a lot

<table>
<thead>
<tr>
<th>Condition</th>
<th>$\alpha_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words-per-sec $\geq 0.73$</td>
<td>1.43</td>
</tr>
<tr>
<td>Pitch-mean $\geq 194.6$</td>
<td>0.41</td>
</tr>
<tr>
<td>Voiced-time $\geq 647.4$</td>
<td>0.41</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Pitch-deviation $\geq 118.1$</td>
<td>-0.15</td>
</tr>
<tr>
<td>Intensity-deviation $\geq 6.3$</td>
<td>-0.18</td>
</tr>
<tr>
<td>Pitch-deviation $\geq 119.7$</td>
<td>-0.47</td>
</tr>
</tbody>
</table>

Sum $\rightarrow$ Extraversion ranking score
Features of extraversion
Features of introversion
RankBoost Models

- **Observed conscientiousness with all features**
  - **Conscientious people**
    - Talk about their occupation (e.g. *work, class, boss*)
    - Use insight words (e.g. *think, know, consider*)
  - **Unconscientious people**
    - Swear a lot (e.g. *damn, f*ck, *p*ss*)
    - Talk loud

<table>
<thead>
<tr>
<th>Condition</th>
<th>$\alpha_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupation $\geq 1.21$</td>
<td>0.37</td>
</tr>
<tr>
<td>Insight $\geq 2.15$</td>
<td>0.36</td>
</tr>
<tr>
<td>Positive feelings $\geq 0.30$</td>
<td>0.30</td>
</tr>
<tr>
<td>Intensity-deviation $\geq 7.83$</td>
<td>0.29</td>
</tr>
<tr>
<td>Num letters $\geq 3.29$</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
<tr>
<td>Swearing $\geq 0.93$</td>
<td>-0.21</td>
</tr>
<tr>
<td>Swearing $\geq 0.17$</td>
<td>-0.24</td>
</tr>
<tr>
<td>Religion $\geq 0.32$</td>
<td>-0.27</td>
</tr>
<tr>
<td>Swearing $\geq 0.65$</td>
<td>-0.31</td>
</tr>
<tr>
<td>Intensity-max $\geq 86.84$</td>
<td>-0.50</td>
</tr>
</tbody>
</table>
Conclusion

• Models performance better than baseline for extraversion, emotional stability, and conscientiousness

• Observed personality easier to model
  – Self-reports are influenced by many factors, e.g. desirability of the trait

• Spoken language with observer ratings produce best models
  – Less constrained?

• Regression results: (improvement over baseline)

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Self reports</th>
<th>Observer reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written language</td>
<td>0.7% - 6.2%</td>
<td>?</td>
</tr>
<tr>
<td>Spoken language</td>
<td>N.S.</td>
<td>3.9% - 23.2%</td>
</tr>
</tbody>
</table>
References


• Try the online demo!

http://www.dcs.shef.ac.uk/~francois/personality/demo.html

Thank you
Essays – Self Reports Distributions
Essays – Self Reports Distributions

![Histograms for ZCONSC and ZOPEN](chart.png)
• Standard deviations between 0.5 and 1.0
EAR - Observer Ratings Distributions

![Graphs showing distributions of ratings for EMSJ and OPEJ](image_url)