Others and me: Detecting personality traits and attachment orientations in Online Social Networks

Karanatsiou, D., Vakali, A., Kafetsios, K.

GEC 2019

Athens, 7 June 2019
Motivation - Challenge

Dynamics of user behavior in social networks reveal phenomena that are closely related to human factor and psychology.

People carry their real self online and not a fake persona.

User Generated Content (UGC) is enhanced with several metadata which reflect users inner psychological world (language, emotions, behavior).

Uncover a holistic model that explains un-expectable behavior and decisions (hidden psychological factor).

Applications on niche marketing, job matching, management, politics and more.
Building a holistic personality prediction model

Holistic personality prediction methodology:
- **Self traits and relationships with others** (crowdsource ground truth data)
- **Extracting language, emotion and behavioral features** as expressed on Twitter
- **Training a Regression Chain model** to exploit intercorrelations between traits
Experimentation

- Crowdsourced ground truth data based on mini-IPIP, ECR-R and NARQ-S scales (Amazon Mturk)
- 26,000 Tweets and Twitter profiles of 105 quality users (non-spammers) were used for features extractions
- Language Features: Open vs Closed vocabulary approach (N-grams, POS-tag vectors and BoW vs LIWC)
- Emotion Features: Primary emotion detection with WordNet-Affect extension
- Behavioral Features: Build-in platform metrics and extracted features
- Regression chains with Random Forest base estimator were utilized to exploit hidden intercorrelations between traits
Results and Conclusions

**Single trait prediction**

*Anxiety* orientation and *neuroticism* are the easiest and *narcissism* the most difficult trait to predict.

**Random Forest** performed better for most of the traits.

**Language** features performed better for most of the traits.

**Table 1.** Experiment results for single trait prediction

<table>
<thead>
<tr>
<th>Trait</th>
<th>Best Model</th>
<th>Best Features</th>
<th>MSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>Gaussian Processes</td>
<td>Language (Tf-Idf)</td>
<td>0.038</td>
<td>0.136</td>
</tr>
<tr>
<td>Avoidance</td>
<td>Random Forest</td>
<td>Language (Tf-Idf)</td>
<td>0.061</td>
<td>0.195</td>
</tr>
<tr>
<td>Openness</td>
<td>Random Forest</td>
<td>Language (Trigrams-POS)</td>
<td>0.057</td>
<td>0.172</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>Random Forest</td>
<td>All features combined</td>
<td>0.054</td>
<td>0.183</td>
</tr>
<tr>
<td>Extraversion</td>
<td>Random Forest</td>
<td>Language (POS)</td>
<td>0.092</td>
<td>0.262</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>Random Forest</td>
<td>All features combined</td>
<td>0.050</td>
<td>0.199</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>Random Forest</td>
<td>Language (Tf-Idf)</td>
<td>0.044</td>
<td>0.176</td>
</tr>
<tr>
<td>Narcissism</td>
<td>Gaussian Processes</td>
<td>Language (Tf-Idf)</td>
<td>0.097</td>
<td>0.275</td>
</tr>
</tbody>
</table>

**Figure 2.** Performance comparison between multioutput regressor and regression chain.
This research has been financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (Project Code: T1EDK-03052). Year: 2018