



ARISTOTLE
UNIVERSITY
OF THESSALONIKI



DATA & WEB
SCIENCE
LABORATORY

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

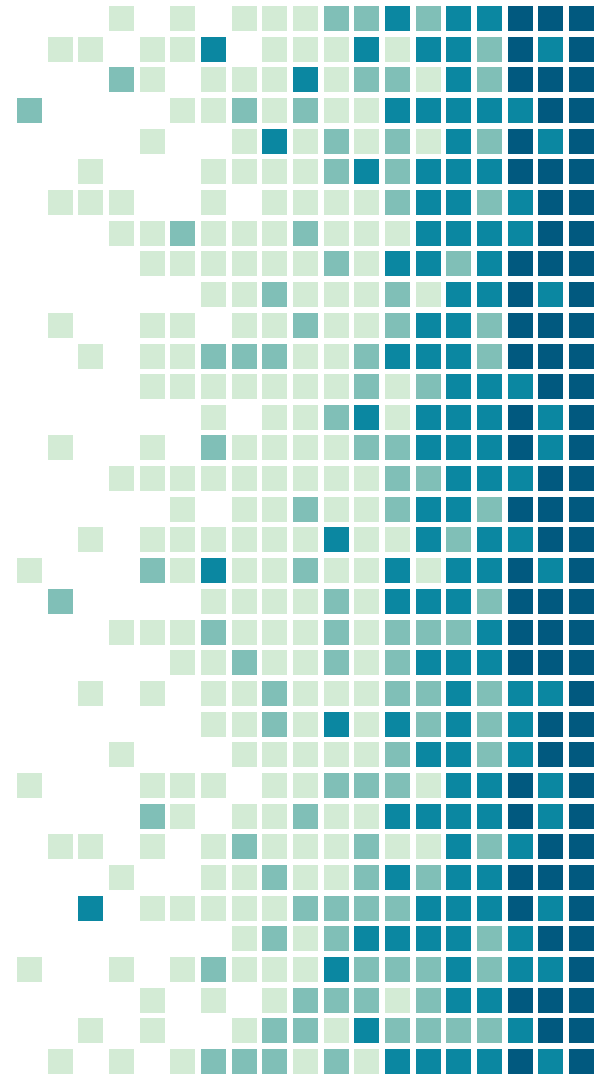
Dynamics of user behavior in social networks reveal phenomena that are closely related to human factor and psychology

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

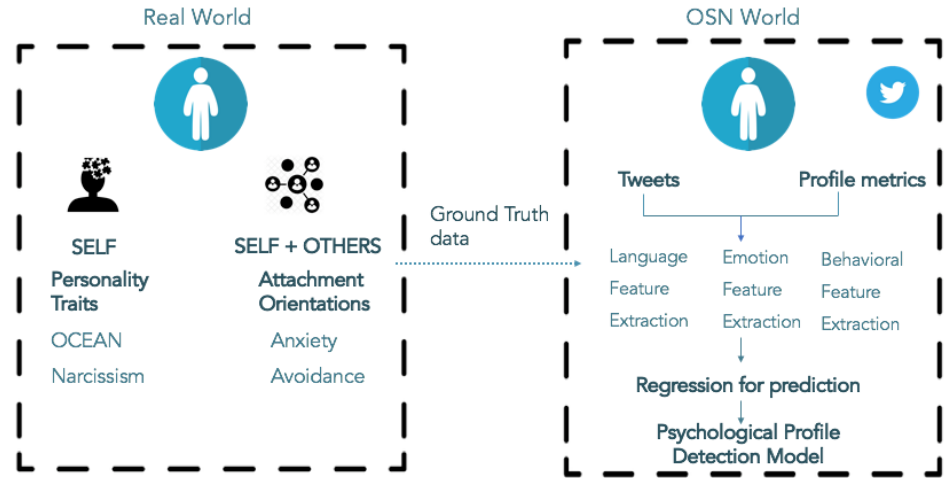
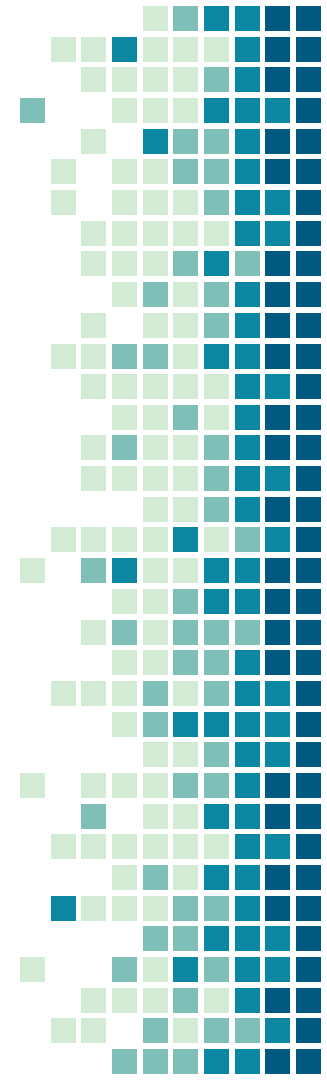


Fig. 1 Process of training a personality prediction model

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 feature 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

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

Regression Chains performance

For holistic personality detection, general language features and the intercorrelations between personality characteristics can lead to higher accuracy than independent multioutput regression models

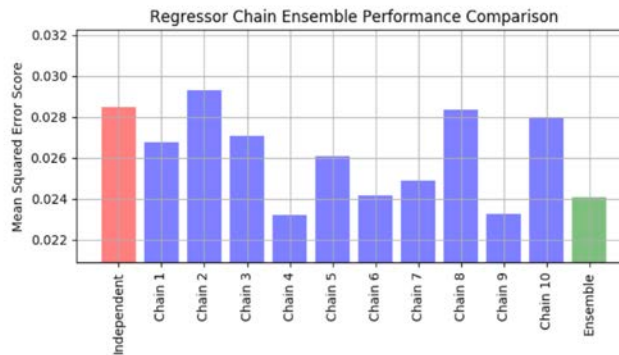
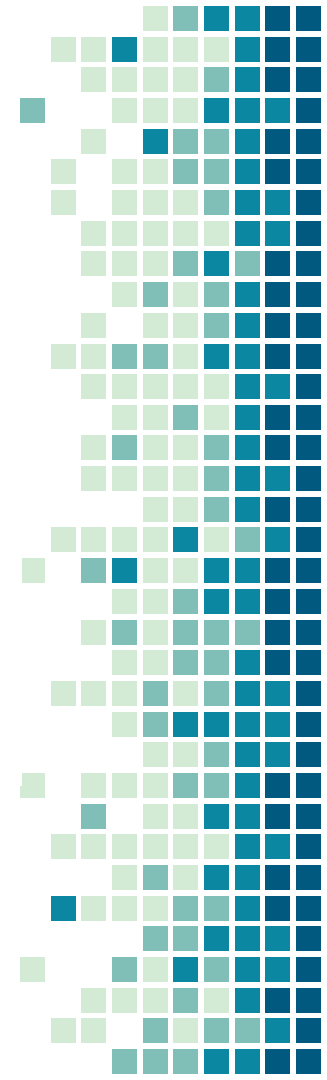


Figure 2. Performance comparison between multioutput regressor and regression chain





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: TIEDK-03052).Year:2018

