

---

# How Well Do Your Facebook Status Updates Express Your Personality?

---

Golnoosh Farnadi<sup>1,2</sup>, Susana Zoghbi<sup>2</sup>, Marie-Francine Moens<sup>2</sup>, Martine De Cock<sup>1</sup>  
{Golnoosh.Farnadi, Martine.DeCock}@UGent.be  
{Susana.Zoghbi, Sien.Moens}@cs.kuleuven.be

<sup>1</sup>Department of Applied Mathematics, Computer Science and Statistics, Ghent University

<sup>2</sup>Department of Computer Science, Katholieke Universiteit Leuven

**Keywords:** Personality recognition, Classification, Online Social Network sites, Facebook

User generated content in online social networking sites provides a potentially rich source of information for applications that rely on personalisation, such as on-line marketing. In this study we contribute to this effort by exploring the use of machine learning (ML) techniques to automatically infer users' personality traits based on their Facebook status updates (i.e., text messages to communicate with friends).

Personality traits are commonly described using five dimensions (known as the Big Five), i.e., extraversion (EXT), agreeableness (AGR), conscientiousness (CON), neuroticism (NEU), and openness (OPN). More than one trait can be present for the same user. We train a binary classifier for each trait that separates the users displaying the trait from those who do not. Formally, given a set of statuses of each user, represented as a feature vector  $x \in R^p$ , the task is to obtain the set  $F = \{(x_1, y_1), \dots, (x_l, y_l)\}$ , where  $y_i \in C = \{C_1, \dots, C_5\}$  corresponding to the five traits. We use a variety of features as input for the classifiers: (1) features related to the text of statuses (e.g., vocabulary and writing style), (2) features about the user's social network (e.g., network size and density) and (3) temporal factors (e.g., frequency of updating status).

Our initial results, based on 250 users and 9917 status updates, show that even with a small set of training examples we can outperform the majority baseline for each trait, with SVM with a linear kernel leading over kNN with  $k=1$  and Naive Bayes. Table 1 presents the results obtained based on accuracy.

	EXT	NEU	AGR	CON	OPN
Majority baseline	0.62	0.55	0.54	0.52	0.70
Classification results	0.68	0.66	0.57	0.55	0.71

Table 1. Classification results based on accuracy (Golbeck et al., 2011) have recently done a similar study on personality prediction based on all pub-

licly available information in a user's Facebook profile. They obtain promising results on a dataset of 167 users, which is richer than ours in the sense that they have crawled many more profile features (e.g., gender, religion, list of favorite things,...) which were not available to us. Our experiments were carried out on a 250 user sample of a Facebook dataset from the myPersonality project that was released on Feb 1, 2013 (Celli et al., 2013). More efforts on predicting personality traits using the myPersonality project data are undoubtedly underway, but no work has been published yet based exclusively on Facebook status updates, network properties and time factors, like our work.

We also trained the classifiers on a corpus of 2468 essays labeled with personality traits (Mairesse et al., 2007). These essays are on average much longer than the status updates, and the context is different. Still, our results show that models trained on the essay dataset perform well on the Facebook data, and vice versa. This provides evidence that ML based models for personality trait recognition generalise across different domains. Advantages of this are that training examples from different social media platforms can be used in combination to train more accurate models and that such models are also applicable on social network sites for which no training data is available.

## References

- Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. (2013). Workshop on computational personality recognition (shared task). *Proc. of WCPR13, in conjunction with ICWSM13*.
- Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. *Proc. of CHI*.
- Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*, 457-500.