# The Workshop on Computational Personality Recognition 2014

Fabio Celli
University of Trento,
Trento, Italy
fabio.celli@unitn.it

Bruno Lepri
FBK,
Trento, Italy
lepri@fbk.eu

Joan-Isaac Biel
IDIAP research institute, EPFL
Martigny, Switzerland
jibiel@idiap.ch

Daniel Gatica-Perez
IDIAP research institute,
Martigny, Switzerland
gatica@idiap.ch

Giuseppe Riccardi
University of Trento,
Trento, Italy
riccardi@disi.unitn.it

Fabio Pianesi
FBK,
Trento, Italy
pianesi@fbk.eu

## ABSTRACT

The Workshop on Computational Personality Recognition aims to define the state-of-the-art in the field and to provide tools for future standard evaluations in personality recognition tasks. In the WCPR14 we released two different datasets: one of Youtube Vlogs and one of Mobile Phone interactions. We structured the workshop in two tracks: an open shared task, where participants can do any kind of experiment, and a competition. We also distinguished two tasks: A) personality recognition from multimedia data, and B) personality recognition from text only. In this paper we discuss the results of the workshop.

## Categories and Subject Descriptors

5.5 [**Emotional and Social Signals in Multimedia**]: Novel methods for the classification and representation of interactive social and/or emotional signals

## Keywords

personality recognition; WCPR; multimedia; youtube; evaluation

## 1. INTRODUCTION

Personality Recognition is an emerging research field that consists in the automatic classification of users' personality traits from various sources, including text [8] and multimedia, [2], [4]. Gold standard labels for an objective evaluation can be obtained by means of the Big5 personality tests [5], which are well-known and widely accepted in psychology and other research fields. The Big5 factor model defines personality along 5 bipolar scales: extraversion (sociable vs. shy); emotional stability (secure vs. neurotic); agreeableness (friendly vs. ugly); conscientiousness (organized vs. careless); openness to experience (insightful vs. unimaginative).

The *Workshop on Computational Personality Recognition*[1] aims

---

[1]Website: https://sites.google.com/site/wcprst/home/wcpr14

to allow participants to compare the performance and quality of different approaches in personality recognition tasks, as well as defining the state-of-the-art. To do that, we stuctured the workshop in two tracks. One track is an open shared task, to test or compare resources, techniques and approaches to the automatic prediction of personality. The other one is a competition, to define a state-of-the-art in the field. We also distinguished two tasks. In **task A** participants aree allowed to use text and multimedia data, while in **task B** they are allowed to use only text. We provided two gold-standard labelled datasets: **Transcriptions of Youtube Video Blogs** (Youtube dataset) [1] and **Mobile Phones interactions** (Mobile dataset) [9]. The Youtube dataset contains 404 users/videos, with transcriptions and observed personality labels. We provided text (246.5k tokens) and feature vectors extracted from videos (including values like pitch, eye gaze direction, camera proximity, energy, time speaking, voice rate etc. See [1] for details), For the Mobile dataset, collected inside the Friends and Family longitudinal study by MIT Human Dynamics group, we provided a sample of 50 users, labeled with self-assessed personality traits, and data about users' proximity, couples, sms and calls. For both datasets we provided training/test splits, official baselines, personality scores and binary personality labels. We also provided instructions for the format of the prediction files and the official personality scorer. Participants to the competition had to deliver the prediction files to test the validity of their experiments.

## 2. CONTRIBUTIONS AND RESULTS

Participants had about three months to develop their own systems for personality recognition and evaluate the results on the datasets released. We released training and test sets at the same time, because we included the shared task in the workshop. We had 6 papers accepted, all of them ran experiments on the Youtube dataset. Among these, 3 teams participated to the competition, 2 to the shared task and 1 to both.

Farnadi *et al.* tested 6 different regression models (3 variants of Target Stacking, 2 of Ensamble Regressor Chains and one Random Forest) to predict the personality of YouTube video bloggers. They use a feature vector that contains gender, audio/visual and linguistic features, extracted from text by means of psychologic (LIWC, MRC, SPLICE), sentiment and emotions (NRC, SentiStrength) lexical resources. Sarkar *et al.* presented an analysis of audio-visual, sentiment, word statistics and gender features to predict personality traits. they report results with a variety of feature combinations. Verhoeven *et al.* tried to classify personality traits for Youtube

text transcriptions using features for gender identification. In the shared task track they ran 10-fold cross validation on the training set to detect which personality traits are predicted best. In the competition they obtained an f1-score of 0.54, averaged over the five personality traits, using gender, liwc and words as features for the prediction of personality from text. Alam & Riccardi exploited a

| | set | alg. | task | feat. | avg-f1 | avg-rmse |
|---|---|---|---|---|---|---|
| baseline | yt | maj | A/B | - | 0.39 | 0.84 |
| Farnadi | yt | st | A | e,l,s,av | - | **0.76** |
| Verhoeven | yt | svm | B | n,l,g | 0.54 | - |
| Sarkar | yt | lr | A | n,e,s,g,av | 0.57 | - |
| Alam | yt | svm | A | s,l,p,e,av | **0.67** | - |

**Table 1: Results of the competition track. Tasks: A=personality prediction from text and multimedia data; B=personality predictions only from text. Algorithms: st=target stacking, svm=support vector machines, lr=logistic regression, svmr=support vector machine regression. Features: e=emotions/sentiment, l=liwc g=gender, s=statistics on word usage, av=audio/video, n=ngrams/raw text, p=part-of-speech.**

large feature space including psycholinguistic, emotional and part-of-speech features, in addition to the audio-visual features provided with the dataset. They obtained very high results automatically combining the best models for each trait. Results of the competition are reported in table 1.

In the shared task, Nowson & Gill trained a model of personality in the Youtube dataset and tested it into the EAR corpus [7], that contains trascriptions of dialogues annotated with personality traits by external observers. They obtain an accuracy of 62% over a baseline of 55% using Naïve Bayes as algoritm and LIWC features. They also report the features that are effective in the different domains. Gievska *et al.* correlated Ekman's 6 emotions [6] with

| | set | alg. | task | feat. | result | base |
|---|---|---|---|---|---|---|
| Verhoeven | yt | svm | B | n,l,g | f1=0.555 | f1=0.546 |
| Gievska | yt | svm | A | av,g,e,s | f1=0.661 | f1=0.508 |
| Nowson | yt+ear | nb | B | l | acc=62% | acc=55% |

**Table 2: Results of the shared task track. Tasks: A=personality prediction from text and multimedia data; B=personality predictions only from text. Algorithms: nb=naïve bayes, svm=support vector machines, Features: e=emotions, l=liwc g=gender, s=statistics on word usage, av=audio/video, n=ngrams, p=part-of-speech.**

the Big Five personality traits, and used those features to augment audio-visual and gender for personality trait classification. They report an F1-score of 0.661 over a baseline of 0.508. Results of the shared task are reported in table 2.

## 3. DISCUSSION AND CONCLUSION

The results of the WCPR14 confirm once again that personality recognition is a very hard task. The scores achieved by the participants in this edition of the workshop reveal that the state-of-the-art in computational personality recognition from multimedia is around f1=0.67 for the prediction of classes and rmse=0.76 for the prediction of z-scores (normalized between 0 and 1). In task B (personality recognition from text) the state-of-the-art is even lower, but more work is required to test this. Comparing this edition of the shared task to the previous one [3], we see that participants successfully exploited emotions and gender for the prediction of personality traits. The variability of performances among traits

is the main problem for the classification, averaged over the five traits. Alam & Riccardi succesfully addressed this problem by voting automatically the best model for each personality trait to obtain a combined model.

In this edition of the Workshop on Computational Personality Recognition, we made available corpora and tools for future evaluations of personality recognition systems. In addition, participants to the workshop have tested many different techniques, feature sets and algoritms for future reference in this and related research fields.

## 4. ACKNOWLEDGEMENTS

## 5. REFERENCES

[1] J. Biel and D. Gatica-Perez. The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *Multimedia, IEEE Transactions on*, 15(1):41–55, 2013.

[2] J.-I. Biel, O. Aran, and D. Gatica-Perez. You are known by how you vlog: Personality impressions and nonverbal behavior in youtube. In *ICWSM*, 2011.

[3] F. Celli, F. Pianesi, D. Stillwell, and M. Kosinski. Workshop on computational personality recognition: Shared task. In *WCPR in conjuction to ICWSM 2013*, 2013.

[4] G. Chittaranjan, J. Blom, and D. Gatica-Perez. Who's who with big-five: analyzing and classifying personality traits with smartphones. In *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*, pages 29–36. IEEE, 2011.

[5] P. T. Costa and R. R. McCrae. The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment*, 2:179–198, 2008.

[6] P. E. Ekman and R. J. Davidson. *The nature of emotion: Fundamental questions.* Oxford University Press, 1994.

[7] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30(1):457–500, 2007.

[8] J. Oberlander and S. Nowson. Whose thumb is it anyway?: classifying author personality from weblog text. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 627–634. Association for Computational Linguistics, 2006.

[9] J. Staiano, B. Lepri, N. Aharony, F. Pianesi, N. Sebe, and A. Pentland. Friends don't lie: inferring personality traits from social network structure. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 321–330. ACM, 2012.