MAPTRAITS 2014: the first audio/visual mapping personality traits challenge: an introduction - perceived personality and social dimensions

Oya Celiktutan, Florian Eyben, Evangelos Sariyanidi, Hatice Gunes, Björn Schuller

MAPTRAITS 2014: The First Audio/Visual Mapping Personality Traits Challenge – An Introduction

Perceived Personality and Social Dimensions

Oya Celiktutan
School of Electronic Eng. and
Computer Science
Queen Mary University of
London
E1 4NS London, UK
o.celiktutandikici@qmul.ac.uk

Florian Eyben
Machine Intelligence & Signal
Processing
Technische Universität
München
80290 München, DE
eyben@tum.de

Evangelos Sariyanidi
School of Electronic Eng. and
Computer Science
Queen Mary University of
London
E1 4NS London, UK
e.sariyanidi@qmul.ac.uk

Hatice Gunes
School of Electronic Eng. and
Computer Science
Queen Mary University of
London
E1 4NS London, UK
h.gunes@qmul.ac.uk

Björn Schuller[*]
Machine Intelligence & Signal
Processing
Technische Universität
München
80290 München, DE
schuller@tum.de

ABSTRACT

The Audio/Visual Mapping Personality Challenge and Workshop (MAPTRAITS) is a competition event that is organised to facilitate the development of signal processing and machine learning techniques for the automatic analysis of personality traits and social dimensions. MAPTRAITS includes two sub-challenges, the continuous space-time sub-challenge and the quantised space-time sub-challenge. The continuous sub-challenge evaluated how systems predict the variation of perceived personality traits and social dimensions in time, whereas the quantised challenge evaluated the ability of systems to predict the overall perceived traits and dimensions in shorter video clips. To analyse the effect of audio and visual modalities on personality perception, we compared systems under three different settings: visual-only, audio-only and audio-visual. With MAPTRAITS we aimed at improving the knowledge on the automatic analysis of personality traits and social dimensions by producing a benchmarking protocol and encouraging the participation of various research groups from different backgrounds.

*Dr. Schuller is also with Department of Computing, Imperial College London, SW7 2AZ London, UK.

Categories and Subject Descriptors

J.4 [**Social and Behavioural Sciences**]: Psychology, Sociology; H.1.2 [**Human/Machine Systems**]: Human Information Processing; I.5 [**Pattern Recognition**]: Applications—*Computer Vision, Signal Processing*

Keywords

Personality; Big Five Model; Attractiveness; Challenge

1. INTRODUCTION

Automatic assesment of people's personality is very important for multiple research domains, particularly those that are related to human-computer or human-robot interaction such as social robotics. Despite a growing interest on personality traits and their effects on human life, systems focusing on automatic analysis of personality traits have started to emerge only recently. Most systems focused on unimodal cues such as written texts, audio, speech, face and body gestures, with some tentative efforts on multimodal analysis. None of the proposed systems to date attempted to assess personality traits continuously in time and space along multiple trait dimensions at a given interaction time.

The MAPTRAITS Challenge aimed at advancing the state of the art in automatic analysis of personality traits and social dimensions by providing a benchmarking protocol that enables evaluation from multiple perspectives, and encouraging the participation of diverse research groups from different disciplines, in particular the audio and video analysis communities and those in the social sciences who study personality, traits, expressive and nonverbal behaviour.

We considered personality in terms of the Big Five model that consists of five traits, namely, extraversion, agreeableness, conscientiousness, neuroticism, openness. As for social dimensions, we considered four dimensions, namely, engagement, facial attractiveness, vocal attractiveness, and lika-

bility. Our benchmarking protocol was based on the well-known SEMAINE dataset, which consists of videos that involve a conversation between a person and an artificial agent [3]. In the scope of MAPTRAITS, we provided a ground truth for all afore-mentiond personality traits and social dimensions. The ground truth has been compiled by combining the annotations of multiple raters who annotated traits and dimensions based on their perception.

In summary, the task in MAPTRAITS was to correctly predict the perceived personality annotations for the audio-visual clips provided. We organised two sub-challenges, i.e. the continuous space-time sub-challenge and the quantised space-time sub-challenge. In continuous space-time challenge the task was to predict the perceived personal traits and social dimensions along with their temporal variation, whereas in the quantised space-time challenge the task was to predict the overall perceived traits and dimensions in shorter video clips. To analyse the effect of audio and visual modalities, we compared the results of each sub-challenge separately for audio and visual modalities, and also for the combination of the two modalities. MAPTRAITS 2014 is the first of an expected series of competitions. We hope that this first version will serve as a useful benchmarking protocol and will push the state of the art in the field by facilitating the development of novel approaches to unimodal/multimodal feature extraction/fusion and prediction.

2. THE MAPTRAITS PROTOCOL

2.1 Dataset

The dataset for quantised space-time challenge comprised 44 clips of 11 different subjects, each rated by 6 annotators along 9 dimensions (the five personality traits and four social dimensions). Videos were rated in a Likert scale ranging from 1 to 10. To promote multi-modal analysis, two sets of annotations were provided; one where annotators used only visual modality and one where annotators used both audio and visual modalities. For both sets of annotations, ground truth labels were generated by taking the average of the annotations per clip per dimension.

The continuous space-time sub-challenge dataset consisted of 30 clips of 10 subjects, each annotated by 21 raters in a scale that ranges from 0 to 100. Three sets of annotations were provided based on three different annotation settings, namely audio only, visual only and audio-visual. To produce the ground truth, we first aligned the annotations of multiple annotators using Dynamic Time Warping (DTW), then eliminated inconsistent annotations based on inter-rater agreement and finally computed the average of the remaining reliable annotations [1].

2.2 Metrics

To enable evaluation from multiple perspectives, we used a variety of evaluation metrics for each sub-challenge. For the quantised space-time challenge, we used the mean square error (MSE), Pearson's correlation coefficient (COR) and coefficient of determination (R^2) and unweighted average recall (UAR). For the continuous space-time challenge we used the MSE and COR metrics as well as the concordance correlation coefficient and cross correlation coefficient.

2.3 Baseline

We provided baseline results for both challenges in three settings: visual-only, audio-only and audio-visual. For the visual-only setting, we first aligned faces with the technique of Xiong and De La Torre [5], then computed the Quantised Local Zernike Moments (QLZMs) of aligned faces [4] and finally obtained predictions using ridge regression. For the audio-only setting, we extracted features such as energy, spectral, cepstral and voicing related low-level features using the openSMILE feature extractor [2] and obtained predictions using SVM (for quantised predictions) and SVR (for continuous predictions).

3. PARTICIPANTS AND RESULTS

For the quantised space-time challenge, the best baseline results in terms of MSE were obtained in audio-visual setting where the MSE of 9 dimensions varied from 0.61 to 3.20. For the continuous space-time sub-challenge, the best baseline results in terms of MSE were generally obtained in visual-only setting where MSE varied between 0.28 and 0.41.

Nine groups from six different countries registered for the MAPTRAITS Challenge. The results obtained by participating systems were not significantly higher than the baseline results, demonstrating the difficulty of the problem. MAPTRAITS 2014 was the first of an expected series of competitions that took the initiative for advancing the state of the art forward and produced a benchmarking protocol that can be used by future studies as well. We expect to see an increasing number of participants in the coming series of competitions with the hope of advancing the state of the art further in this challenging problem.

Acknowledgments

4. REFERENCES

[1] O. Celiktutan and H. Gunes. Continuous prediction of perceived traits and social dimension in space and time. In *Proc. of IEEE Int. Conf. on Image Processing*, Paris, France, 2014.

[2] F. Eyben, F. Weninger, F. Gross, and B. Schuller. Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 835–838. ACM, 2013.

[3] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder. The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Trans. on Affective Computing*, 3(1):5–17, 2012.

[4] E. Sariyanidi, H. Gunes, M. Gökmen, and A. Cavallaro. Local Zernike moment representations for facial affect recognition. In *Proc. of British Machine Vision Conf.*, 2013.

[5] X. Xiong and F. De la Torre. Supervised descent method and its application to face alignment. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2013.