

**Fact sheet**

**Explainable Computer Vision Workshop**

and

**Job Candidate Screening Coopetition**

**Human-explainable features for Job Candidate Screening Prediction**

9 April 2017

**1. Team details**

1.1 Team name: TUD-MMC

1.2 Team leader name: Achmadnoer Sukma Wicaksana

1.3 Team leader address, phone number and email: Zoomweg 7, 6705DM, Wageningen, The Netherlands

+31 6 533 24536

achmadnoersukmawicaksana@student.tudelft.nl

1.4 Rest of the team members: Cynthia C. S. Liem

1.5 Team website URL (if any): NA

1.6 Affiliation: Delft University of Technology

**2. Contribution details**

2.1 Title of the contribution: Human-explainable features for Job Candidate Screening Prediction

2.2 General method description:

In our method, we wanted to ensure that features were used which are understandable by humans. We focused on several visual and linguistic features.

* Overview of the proposed approach:

The proposed method uses four different features representations and average weighting for prediction of six annotation traits.

**Feature Representation & Extraction**:

1. We use OpenFace [1] (<https://github.com/TadasBaltrusaitis/OpenFace>) to extract 112x112 face videos and recognize Facial Action Unit (AU) for each frame of the videos. These AU’s can be trace back on changes in facial expressions that are present along the video. There are two ways to describe AUs from OpenFace; presence and intensity. We extract percentage of AUs appearing over videos, maximum intensity of AUs, and average of AUs’ intensity over video from this representation.
2. The resulting face videos from first step are used to obtain Motion Energy Image. We can measure facial movement better with this approach because the overall resulting image not necessarily disturb by moving background. Three statistics then are produced from these images; mean, median, and entropy.
3. A simple features of word count and unique word count are produced from the given transcript file.
4. Readability scores are used from transcript file to measure the complexity of spoken words. We produce 8 readability index for this purpose using the readability measures as implemented through NLTK-contrib (<https://github.com/nltk/nltk_contrib/tree/master/nltk_contrib/readability>).

**Trait Prediction**:

Each features representation is trained separately to predict the annotation scores. PCA is used for dimension reduction and use to trace back features influence through the transformation matrix. Final prediction is produced by averaging between each representation’s linear regression prediction results.

**Explainability**:

We on purpose picked features that are human-explainable. The value of each of the features is reported in the descriptive evaluation, together with a short explanation of what the feature should depict. To give a better sense of how ‘unusual’ a value is, the value for each feature is compared to the 6000 earlier assessed people in the training set, and a percentile is reported. Besides, as we use a linear regression model, it is possible to (approximately) indicate what factors contribute negatively or positively to the model’s prediction. By considering the transformation matrix of the PCA, we can determine what original features have the strongest factor weights in the main PCA dimensions, and through the linear model, we know the sign of each dimension’s ultimate contribution to the prediction. For the two features most strongly represented in the two dimensions of the PCA with the strongest linear model coefficients, we report on the influence of the feature on the linear model (e.g. ‘high values of feature X tend to lead to high prediction scores’).

* The proposed method uses / takes advantage of personality traits? No
* Coopetition: did you use any code or idea from other participants (shared in previous stage of the challenge)? No
* Total method complexity: NA
* Which pre-trained or external methods/models have been used (for any stage, if any): NA
* Which additional data has been used in addition to the provided ChaLearn training and validation data (at any stage, if any): NA
* Qualitative advantages of the proposed solution:

1. The use of Facial Action Unit can be traceback on why certain video have certain scores in the traits.
2. Motion features are more detailed on the face which reduce background noise and focused on persons.
3. Text based features can relate to how comprehensive people on their spoken context.

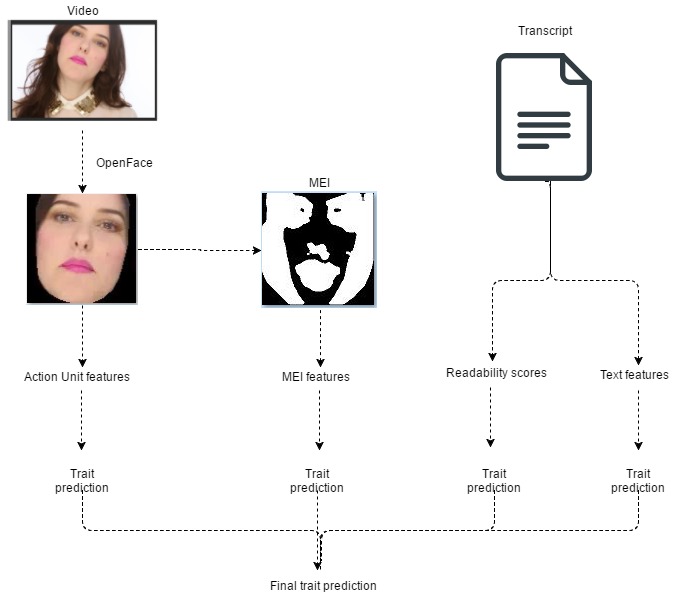
* Novelty degree of the solution and if it has been previously published: we did not publish our setup yet. In terms of novelty, we on purpose took the ‘old-fashioned’ way of picking ‘crafted’ features and a regression model which allow for a reasonably clear human explanation, rather than an automatically learned representation, even though this may not be optimal in terms of prediction accuracy. This was done after local discussions with organizational psychologists, who expressed strongest interest in seeing constructs expressed through traceable and understandable features, and also noted the importance of cognitive ability in many hiring decisions. This was not formally assessed through the personality traits, but we felt that the linguistic features may possibly reveal some of this information.

2.3 GitHub URL for the project: <https://github.com/sukmawicaksana/CVPR2017>

2.4 References:

1. **Cross-dataset learning and person-specific normalisation for automatic Action Unit detection** Tadas Baltrušaitis, Marwa Mahmoud, and Peter Robinson in Facial Expression Recognition and Analysis Challenge, IEEE International Conference on Automatic Face and Gesture Recognition, 2015
2. Biel J, Aran O, Gatica-Perez D (2011) You are known by how you vlog: Personality impressions and nonverbal behavior in youtube. In: Proceedings of international AAAI conference on weblogs and social media (ICWSM)

2.5 Representative image / diagram of the method



2.6 Describe data preprocessing techniques applied:

We use MEI for face videos for one of our features. In order to do that, we use OpenFace to produce 112x112 face videos. Then, these are used to obtain Motion Energy Image for motion features.

**3. Visual Analysis**

3.1 Features / Data representation: AU percentage, AU max intensity, AU average intensity, MEI mean, MEI median, MEI entropy

3.2 Dimensionality reduction: PCA (for AUs and MEI separately)

3.3 Model: NA

3.4 Learning strategy: Linear Regression

3.5 Other techniques: NA

3.6 Method complexity: NA

**4** **Explainability from Visual data**

4.1 Features / Data representation: AU percentage, AU max intensity, AU average intensity, MEI mean, MEI median, MEI entropy

4.2 Dimensionality reduction: PCA (for AUs and MEI separately)­­­­

4.3 Model: NA

4.4 Learning strategy: Linear Regression

4.5 Other techniques: NA

4.6 Method complexity: NA

**5 Explainability from Audio data**

5.1 Features / Data representation: NA

5.2 Dimensionality reduction: NA

5.3 Model: NA

5.4 Learning strategy: NA

5.5 Other techniques: NA

5.6 Method complexity: NA

**6 Explainability from ASR/text data**

6.1 Features / Data representation: Eight Readability indices from NLTK-contrib, word count, and unique word count

6.2 Dimensionality reduction: PCA (for Readability and simple word counts statistics separately)

6.3 Model: NA

6.4 Learning strategy: Linear Regression

6.5 Other techniques: NA

6.6 Method complexity: NA

**7 Multimodal Explainability**

7.1 Data Fusion Strategies: Late fusion by averaging prediction from four feature representations (OpenFace, MEI, Readability, and Text)

**8 Other details**

8.1 Language and implementation details (including platform, memory, parallelization requirements):

OS: Windows-10 64-bit

RAM: 8GB

CPU: Intel® Core™ i7-4710HQ CPU @2.50GHz

Language: Python 2.7 & Matlab R2012b

8.2 Human effort required for implementation, training and validation?: we hard-coded the human explanations for each feature. However, explanations are generated automatically.

8.3 Training/testing expended time?:

8.4 General comments and impressions of the challenge? what do you expect from a new challenge in face and looking at people analysis? It is a good challenge although we missed the quantitative phase which might interesting to enter along with the qualitative phase.

**9 References**

1. **Cross-dataset learning and person-specific normalisation for automatic Action Unit detection** Tadas Baltrušaitis, Marwa Mahmoud, and Peter Robinson in Facial Expression Recognition and Analysis Challenge, IEEE International Conference on Automatic Face and Gesture Recognition, 2015
2. Biel J, Aran O, Gatica-Perez D (2011) You are known by how you vlog: Personality impressions and nonverbal behavior in youtube. In: Proceedings of international AAAI conference on weblogs and social media (ICWSM)