Apparent Human Behavior Understanding: Personality Analysis

Sergio Escalera

www.sergioescalera.com

sergio.escalera.guerrero@gmail.com

Chalearn Looking At People

- **ChaLearn** is a non-profit organisation focusing on challenge organization in Machine Learning

- **ChaLearn Looking at People** (ChaLearn LAP) is a branch of ChaLearn focusing on *Human Behaviour Analysis*

http://chalearnlap.cvc.uab.es/ www.sergioescalera.com

Chalearn Looking At People

- **Challenges organized:**
  - **Face analysis:**
    - Apparent age estimation
    - Face attributes analysis
    - Apparent personality analysis
    - Emotion recognition
  - **Body and behaviour:**
    - Human pose estimation
    - Action recognition
    - RGBD gesture recognition
    - Recovering missing data of humans
  - **Scene understanding:**
    - Cultural event recognition
    - Video decaptioning and denoising

- Workshops at **ICMI, CVPR, ECCV, ICCV, NIPS, ICPR2018, ECCV2018, WCCI2018**
- Special issues at **TPAMI, IJCV, IEEE TAC**
- Competitions at **Codalab open source platform**

Papers, codes, and data available:
http://chalearnlap.cvc.uab.es/

Chalearn Looking At People

- **Challenges organized:**
  - **Face analysis:**
    - Apparent age estimation
    - Face attributes analysis
    - Apparent personality analysis
    - Emotion recognition
  - **Body and behaviour:**
    - Human pose estimation
    - Action recognition
    - RGBD gesture recognition
    - **Recovering missing data of humans**
  - **Scene understanding:**
    - Cultural event recognition
    - **Video decaptioning and denoising**

- Workshops at **ICMI, CVPR, ECCV, ICCV, NIPS, ICPR2018, ECCV2018, WCCI2018**
- Special issues at **TPAMI, IJCV, IEEE TAC**
- Competitions at **Codalab open source platform**

Papers, codes, and data available:
http://chalearnlap.cvc.uab.es/

Chalearn Looking At People

- **Challenges organized:**
  - **Face analysis:**
    - Apparent age estimation
    - Face attributes analysis
    - **Apparent personality analysis**
    - Emotion recognition
  - **Body and behaviour:**
    - Human pose estimation
    - Action recognition
    - RGBD gesture recognition
    - **Recovering missing data of humans**
  - **Scene understanding:**
    - Cultural event recognition
    - **Video decaptioning and denoising**

- Workshops at **ICMI, CVPR, ECCV, ICCV, NIPS, ICPR2018, ECCV2018, WCCI2018**
- Special issues at **TPAMI, IJCV, IEEE TAC**
- Competitions at **Codalab open source platform**

Papers, codes, and data available:
http://chalearnlap.cvc.uab.es/

Apparent Human Behavior Understanding

- **Human attributes** analysis in Computer Vision involves **subjectivity and user experience** in some scenarios (many variables are involved)
- Although **“real” labels of personality** could be approximated based on sophisticated questionnaires and exercises it is **difficult** to be automatically computed by **machine perception analysis**
- In some cases the “**apparent**” attributes recognition can be enough and useful for several application scenarios
- We also want machines to perceive with the subjectivity humans do
- **Apparent personality**: how do we show to others?

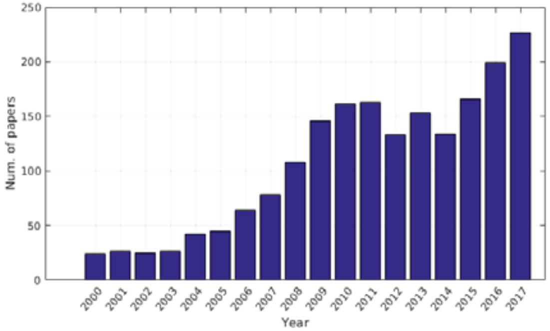
Personality Trait Computing

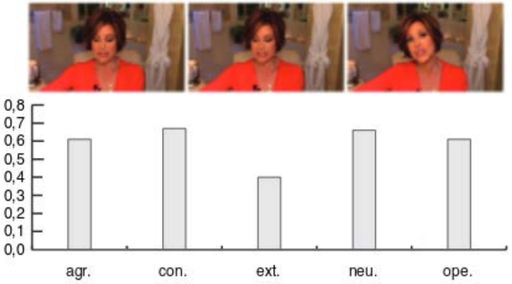
Personality Trait Computing

Personality Trait Computing. The topic has **attracted a lot of attention** from the computer vision and machine learning communities in the last years

- Real personality computing
- Apparent personality computing
- Personality synthesis
    - Affective interfaces, social robotics
    - Adaptive marketing and advertising
    - Recommendation systems
    - Personality-adaptive tutoring systems (social and job skills)
    - Psychological therapy

Searching for "personality" on IEEEXplore





*"The person has a **positive interview recommendation** because of high apparent **conscientiousness** and high apparent **neuroticism**"*

Güçlütürk et al., TAC'17

Apparent Personality Trait Analysis

- Recent studies indicate that **video interviews** are starting to modify the way in which applicants get hired
    - *L. S. Nguyen, D. Frauendorfer, M. S. Mast, and D. Gatica-Perez, “**Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior**,” IEEE transactions on multimedia, 2014*

Apparent Personality Trait Analysis

- For instance, **rapid judgments of competence based solely on the facial appearance of candidates** was enough to allow participants from an experiment to *predict the outcomes* of gubernatorial and Senate elections in the United States in 68.6% and 72.4% of the cases, respectively

- For example, ***politicians who simply look more competent are more likely to win elections***

*C. C. Ballew and A. Todorov, "**Predicting political elections from rapid and unreflective face judgments**," in Proceedings of the National Academy of Sciences of the USA, 2007*

Apparent Personality Trait Analysis

Trait theory

- According to *Vinciarelli and Mohammadi*, the models that most **effectively predict measurable aspects in the life of people** are those based on traits

- <u>*Trait theory*</u> is an approach based on the ***definition and measurement of traits***: habitual patterns of behavior and emotion relatively stable over time

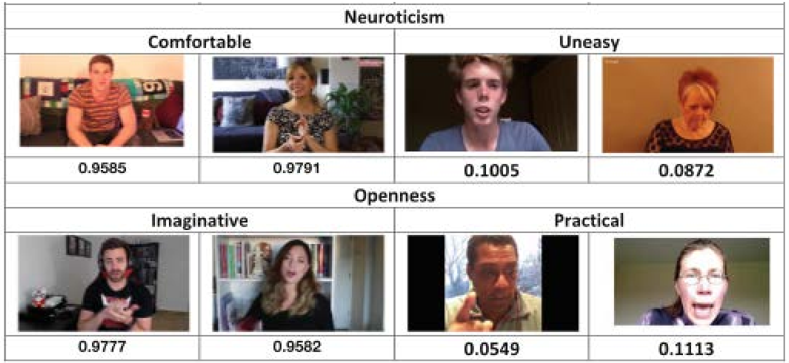
- <u>*Trait models*</u> are built upon human **judgments about *semantic similarity*** and ***relationships between adjectives*** that people use to describe themselves and the others

*A. Vinciarelli and G. Mohammadi, "**A survey of personality computing**," IEEE Transaction on Affective Computing, 2014.*

Apparent Personality Trait Analysis

Trait models

- Different **trait models** have been proposed and broadly studied: the ***Big-Five*** or *Five-Factor Model*, Big-Two (*Agency*: competence, instrumentality, power; and *Communion:* warmth, morality, expressive-ness), 16PF, among others



Screenshot of sample videos labeled

*V. Ponce-Lopez, et al., "**Chalearn lap 2016: First round challenge on first impressions-dataset and results,**" ECCVW, 2016.*

Apparent Personality Trait Analysis

Real *vs* Apparent Personality Trait

- Despite its known limitations, the **self-report questionnaire** has become the dominant method for assessing the *real personality* of an individual

Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree Strongly
1	2	3	4	5

I see Myself as Someone Who...

____1. Is talkative

____2. Tends to find fault with others

____23. Tends to be lazy

____24. Is emotionally stable, not easily upset

- *On the other hand,* **the prediction of the personality others attribute to a given individual** is referred as
    - Personality **Perception**
    - **Apparent** Personality Trait Analysis or
    - First **Impression** analysis

Apparent Personality Trait Analysis

First impressions

- According to *Vernon et al.*, people make first impressions about others **from a glimpse** as brief as 100ms or less, and brain activity appears to track social traits even when no explicit evaluation is required
    - *R. J. Vernon, C. A. Sutherland, A. W. Young, and T. Hartley, "**Modeling first impressions from highly variable facial images**," Proc. of the National Academy of Sciences, 2014.*

"You never get a second chance to make a First Impression"

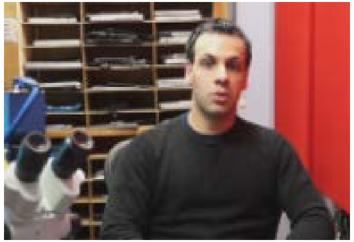
Personality Trait Computing: Labeling

Apparent Personality Trait Analysis

Evaluating first impressions

- Generally, it is measured in terms of **how close** the outcomes of the approach to the **judgments made by the external observers** are

- Making automatic apparent personality trait analysis a **subjective and complex** task



	True	Pred.
Open.	0.57	0.54
Cons.	0.70	0.71
Extr.	0.38	0.40
Agre.	0.60	0.57
Neur.	0.54	0.56

- Raising further questions
  - **How many raters** should be involved in an experiment?
  - **How much they should agree** with one another?
  - **How we should ask** to label this kind of complex and subjective problem?

- The **low agreement** should not be considered the result of low quality judgments or data, but an **effect of the inherent ambiguity of the problem**

Apparent Personality Trait Analysis

Continuous Variables and Bias problems

- In many works, external observers were asked to **label data** with ***continuous values*** or using values mapped into a specific range scale
- Ex.: **How Extroverted is the person in the image?**
  - *Choose:* Low ← (1), (2), (3), (4), (5), (6), (7), (8), (9), (10) → High
- But, **what does a *Extroversion* score of 0.3 mean?**
- Now, imagine **how challenging and subjective can be** to provide the labels, in a similar manner, for the other Big-Five dimensions

Apparent Personality Trait Analysis

Some Attempts to Address Data Labeling

- *Joo et al.* asked annotators from AMT to **“compare” a pair of images (650 total images)** in 14 given dimensions *rather than evaluating each image individually*
    - Annotators **do not need to establish the absolute baseline** or scales in these social dimensions, **which would be inconsistent**
    - The comparison-based ratings **naturally identify the strength of each example** in the context of relational distance from the other examples
    - From these pairwise ratings, the global ranking orders of all examples is retrieved using ***HodgeRank*** (*Jiang et al.*)

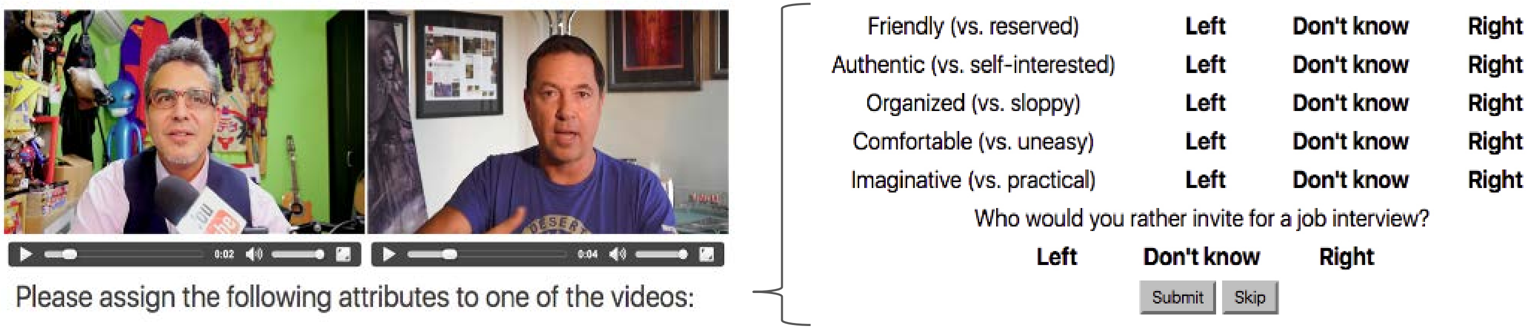
*X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye. “**Statistical ranking and combinatorial hodge theory**,” Mathematical Programming, 127(1):203–244,2011*

*J. Joo, F. F. Steen, and S. C. Zhu, “**Automated facial trait judgment and election outcome prediction: Social dimensions of face**,” in ICCV, 2015.*

Apparent Personality Trait Analysis

Pairwise Comparison in Videos

- To alleviate calibration problems, *Chen et al.* used **pairwise comparisons between videos**, and variable levels were reconstructed by fitting **a *Bradley-Terry-Luce* model with maximum likelihood**



*B. Chen, S. Escalera, I. Guyon, V. Ponce-López, N. Shah, and M. Oliu Simón, "**Overcoming Calibration Problems in Pattern Labeling with Pairwise Ratings: Application to Personality Traits**". Springer Int. Publishing, 2016.*

Apparent Personality Trait Analysis

Subjectivity in data labeling

- The subjectivity of data labeling can have its origin in several sources (**data variability-context and rater opinion-experience-knowledge**):
    - Facial appearance → may change the perceived personality
    - Context → change my behavior
    - *Halo* effect → cognitive bias of the rater
    - Cultural and Gender aspects → labeling bias based on gender and culture of raters
    - etc.

Personality Trait Computing:
Datasets

Apparent Personality Trait Analysis

- Several **databases**, **challenges** and **workshops** have been proposed in order to advance the research in the field
  - *O. Celiktutan et al., "**Maptraits 2014: The first audio/visual mapping personality traits challenge**," in Mapping Personality Traits Challenge and Workshop, 2014*
  - *F. Celli et al., "**The workshop on computational personality recognition**," in Int. Conf. on Multimedia, 2014.*
  - *V. Ponce-Lopez, et al., "**Chalearn lap 2016: First round challenge on first impressions-dataset and results**," in ChaLearn LAP Workshop on Apparent Personality Analysis, ECCV Workshops, 2016.*
  - *Escalante, H.J et al., "**Design of an Explainable Machine Learning Challenge for Video Interviews**," IJCNN, 2017.*

Apparent Personality Trait Analysis

Challenge	Year	DB	Task	Samples	Event
ChaLearn LAP	2017	Job candidate screening	Invite to interview	10,000	CVPR/IJCNN
	2016	First impressions	P. traits (Big-5)	10,000	ECCV
	2016	First impressions	P. traits (Big-5)	10,000	ICPR
MAPTRAITS	2014	Semaine	P. traits (Big-5)	44	ICMI
WCPR	2014	YouTube *vlog*	P. traits (Big-5)	404	MM
Speaker Trait	2012	I-ST	P. traits (Big-5 - audio)	640	Interspeech
Related challenges					
EmotionNet	2017	Emotionet	Emotions (Composed,AUs)	1 million	NA
EmotionNet	2016	Emotionet	Emotions (Composed,AUs)	1 million	NA
AVEC	2016	DAIC / RECOLA	Depression, Emotion	Dozens	MM
	2015	RECOLA	Emotion	Dozens	MM
	2014	AVEC	Depression	150	MM
	2013	AVEC	Affect, depression	340	MM
	2012	AVEC	Emotion	12	ICMI
	2011	AVEC	Emotion	12	ACII
EmotiW	2016	AFEW6 / HAPPEI	Emotions (clip)	$\approx$ 1,500	ICMI
	2015	AFEW5 / SFEW2	Emotions (clip+image)	$\approx$ 1,500	ICMI
	2014	AFEW4	Emotions (clip)	$\approx$ 1,500	ICMI
	2013	AFEW1	Emotions (clip)	$\approx$ 1,500	ICMI

Apparent Personality Trait Analysis

Dataset	Year	Short description	Focus	Labels
ChaLearn First Impression v2	2017	Extended version, *Multimodal*	Apparent personality trait and hirability analysis	Big-Five impressions, job interview variable and transcripts
ChaLearn First Impression	2016	10,000 short videos (15sec) collected from YouTube, *Audiovisual*	Apparent personality trait analysis	Big-Five impressions
SEMAINE	2012	959 conversations (~5min), controlled environment, *Multimodal*	Face-to-face conversations	Metadata*, transcripts, 5 affective dimensions and 27 associated categories
Emergent LEAder (ELEA)	2012	40 meetings (~10h), 27 having both audio and video, controlled environment, *Audiovisual*	Small group interactions	Metadata*, Big-Five (self-report) and social impressions
YouTube *vlog*	2011	442 *vlogs* (~50 to 70sec), *Audiovisual*	Conversational *vlogs* and apparent personality traits	Metadata* and Big-Five impressions

* *Metadata* can be gender, age, number of "likes" on social media, presence of laughs, FACS, etc, and vary for each dataset.

Personality Trait Computing: Subjectivity and biases

Apparent Personality Trait Analysis

Context

- Recent studies show that *first impression* **personality perception based on the face** analysis of a person ***can vary with different photos*** (i.e., the ratings vary w.r.t. context), making the evaluation process even harder

Extraversion

*A. Todorov and J. Porter, “**Misleading first impressions different for different facial images of the same person**,” Psychological Science, 2014.*

- According to *Sutherland et al.*, the most important source of within-person variability, which index the main dimensions in theoretical models of facial impressions, is the **emotional expression** of the face, but ***the viewpoint of the photograph*** also affects impressions

*C. A. M. Sutherland, A. W. Young, and G. Rhodes, “**Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties**,” British Journal of Psychology, 2016.*

Apparent Personality Trait Analysis

Context

*Most people see **the face on the <u>left</u>** as **more attractive** than the face on the right (top row).*

Intra-person variability

*Most people see the **face on the <u>right</u>** as **more attractive** than the face on the left (bottom row)*

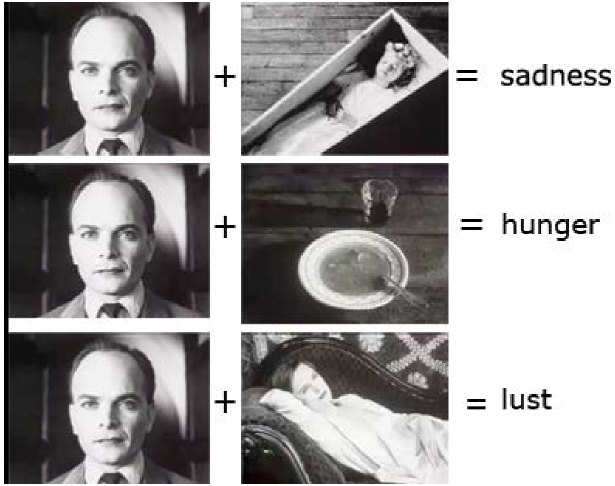
R. Jenkins, D. White, X. Van Montfort, A. M. Burton,
*"**Variability in photos of the same face**", Cognition, 2011*

Apparent Personality Trait Analysis

Context

Kuleshov Effect

- According to film mythology, the Soviet filmmaker *Lev Kuleshov* conducted an experiment (in the 1910s and 1920s) in which he **combined a close-up of an actor's neutral face with three different emotional contexts**: lust, sadness, and hunger
    - Viewers **perceived the actor's face as expressing an emotion congruent with the given context**



*D. Barratt, A. C. Rédei, A. Innes-Ker, and J. van de Weijer, "**Does the kuleshov effect really exist? revisiting a classic film experiment on facial expressions and emotional contexts**," Perception, 2016.*

Apparent Personality Trait Analysis

Social

- Within-person variance in **behavior** is likely to be a ***response to variability*** in **relevant situational cues**
    - ***In general, people are more extraverted in large groups than in small groups*** - even though some individuals may not increase or may even decrease their level of Extraversion with the size of the group
    - **Context affect the personality state** of a person: at home, in a party or in a job interview

*W. Fleeson, "**Toward a structure- and process-integrated view of personality: Traits as density distributions of states**," Journal of Personality and Social Psychology, 2001.*

*Jyoti Joshi, Hatice Gunes, Roland Goecke, **"Automatic Prediction of Perceived Traits Using Visual Cues under Varied Situational Context"**, ICPR 2014*

Apparent Personality Trait Analysis

Halo Effect

- According to *Todorov et at.*, the effects of appearance on **social outcomes** may be partly attributable to *halo* effect, which might have a **strong impact on how observers make their first impressions** about the others
  - *It is the tendency to use global evaluations to make judgments about specific traits*
  - *It is a form of **cognitive bias** in which the brain allows <u>specific positive traits</u> to <u>positively influence the overall evaluation</u> of a person*
  - For example, **attractiveness correlates with perceptions of intelligence**

Rohner JC, Rasmussen A. "Recognition bias and the physical attractiveness stereotype," Scand J Psychology, 2012

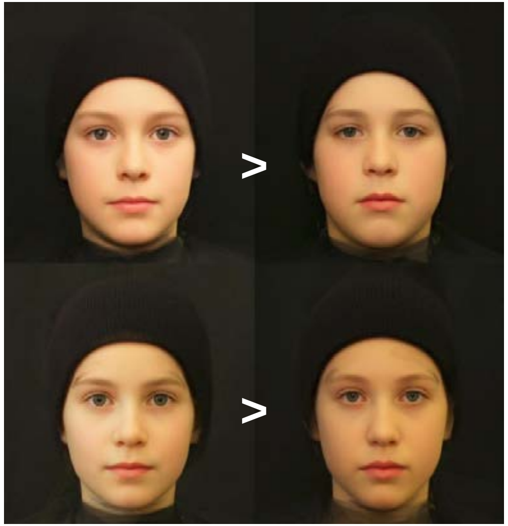
*A. T. Todorov, C. C. Said, and S. C. Verosky, "**Personality Impressions from Facial Appearance**". Oxford Handbook of Face Perception, 2012.*

Apparent Personality Trait Analysis

Attractiveness-Halo Effect

- **A subtle change** in eyelid-openness and mouth curvature in the same person ***can change their perceived intelligence***

Perception of Intelligence

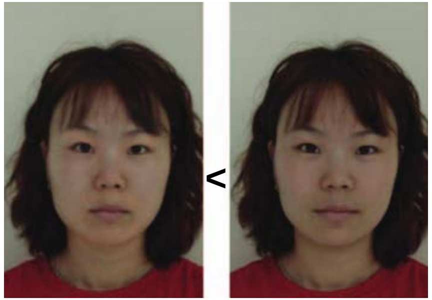


*Talamas, S. N., Mavor, K. I., Axelsson, J., Sundelin, T., & Perrett, D. I. "**Eyelid-openness and mouth curvature influence perceived intelligence beyond attractiveness**". Journal of Experimental Psychology: General, 2016.*

Apparent Personality Trait Analysis

Cultural Aspects

Extraversion



- *Walker et al.* found some cross **cultural consensus**, as well as some ***differences***, while addressing the problem of universal and cultural differences in forming personality trait judgments from faces
- **Aggressiveness, extroversion, likeability**, among others, were analyzed in **photographs of Asian and Western faces**
- **Asian and Western participants** tried to identified the enhanced salience of all different personality traits in the faces
- The results suggest that the associations between mere static facial information and certain personality traits are **highly shared among participants from different cultural backgrounds**
- However, **Asian participants needed more time for this task**
- Faces with enhanced salience of **aggressiveness, extroversion, social skills, and trustworthiness were better identified by Western** than by Asian participants

*M. Walker, F. Jiang, T. Vetter, and S. Sczesny, "**Universals and cultural differences in forming personality trait judgments from faces**," Social Psychological and Personality Science, 2011.*

Apparent Personality Trait Analysis

Gender

- *Jenkins et al.* observed that ***female raters tended to be*** rather harsh on the male faces

  *R. Jenkins, D. White, X. V. Montfort, and A. M. Burton, "**Variability in photos of the same face**," Cognition, 2011.*

- *Mattarozzi et al.* results suggests **women tend to judge trustworthy-looking faces** as significantly more trustworthy than men do, and this is particularly *pronounced for judgments of female faces*

  *K. Mattarozzi, A. Todorov, M. Marzocchi, A. Vicari, and P. M. Russo, "**Effects of gender and personality on first impression**," PLOS ONE, 2015.*

Personality Trait Computing: ChaLearn competitions

Apparent Personality Trait Analysis

Apparent Personality Analysis Challenges

- Automatic evaluation of *apparent personality traits* **from videos** of subjects speaking to the camera (Round 1 - **ECCV'16**, Round 2 - **ICPR'16**)

- Data
  - A data set consisting of **10,000 short clips** from YouTube videos was released
  - Labeled using AMT through pairwise comparisons to alleviate calibration problems
  - Variable levels were reconstructed by fitting **a *Bradley-Terry-Luce* model with maximum likelihood**

- Evaluation
  - Submission and scoring through **CodaLab**, an open-source platform http://codalab.org
  - Accuracy as 1 – regression_value_difference

- Participation
  - Both rounds attracted ~150 participants, from May 15th to July 15th (ECCV) and from June 30 to 16 August, 2016 (ICPR).

Apparent Personality Trait Analysis

Codalab - opensource

- It is a powerful framework for ***running competitions***, which can provide **real-time feedback** for the participants
- ***Organizers*** have several tools and options to **create, setup and monitor** their competitions

http://codalab.org/

Results

#	User	Entries	Date of Last Entry	Team Name	<Rank>	AVG	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness	Coopetition
1	frkngrpnr	3	08/16/16		1.50	0.913 (1)	0.918 (1)	0.907 (2)	0.915 (1)	0.911 (1)	0.914 (1)	0.000 (2)
2	vismay	1	08/15/16	evolgen	1.50	0.912 (2)	0.916 (2)	0.911 (1)	0.914 (2)	0.910 (2)	0.911 (2)	1.000 (1)
3	pandora	3	08/16/16	pandora	2.00	0.903 (3)	0.904 (3)	0.905 (3)	0.901 (3)	0.900 (3)	0.904 (3)	1.000 (1)
4	berkayaydin	2	08/16/16	PILAB	3.00	0.898 (4)	0.895 (4)	0.904 (4)	0.896 (4)	0.894 (4)	0.901 (4)	0.000 (2)

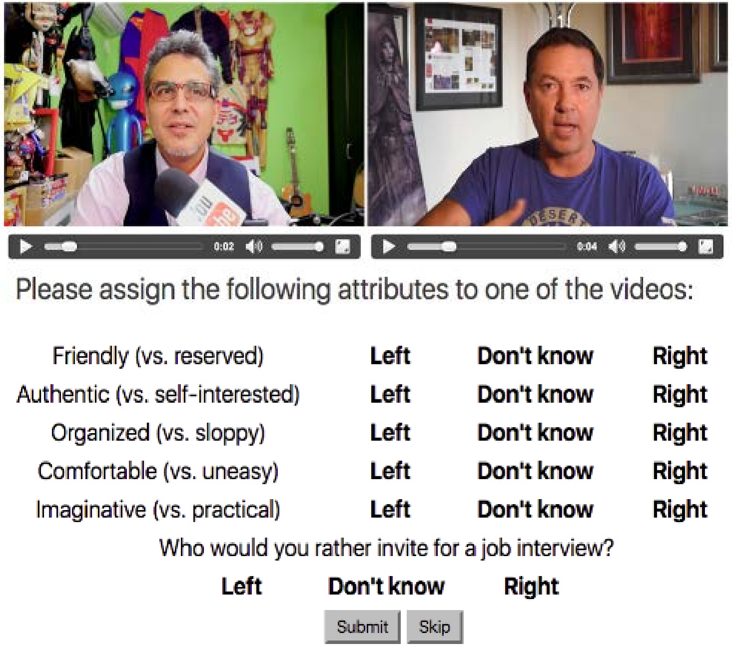
Leaderboard *can provide real-time feedback*

Apparent Personality Trait Analysis
ChaLearn First Impression (ECCV’16) - Round 1

- **10,000 15-second videos** collected from **YouTube**, annotated with personality traits (**Big-Five**) by AMT workers **(creative commons, sanity check)**

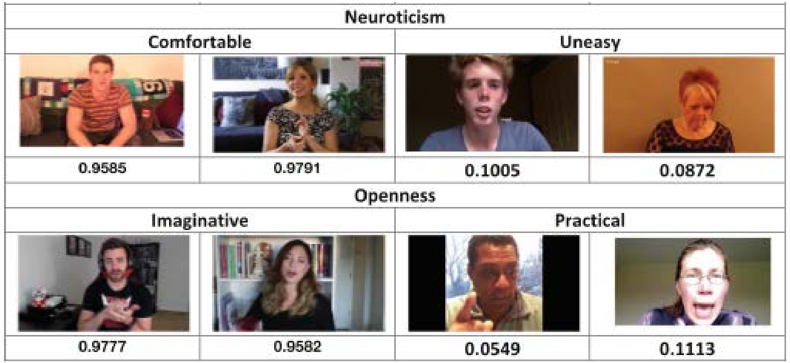
- 86 participants grouped in several teams.

- 9 teams entered the final phase



Apparent Personality Trait Analysis

ChaLearn First Impression (ECCV’16) - Round 1

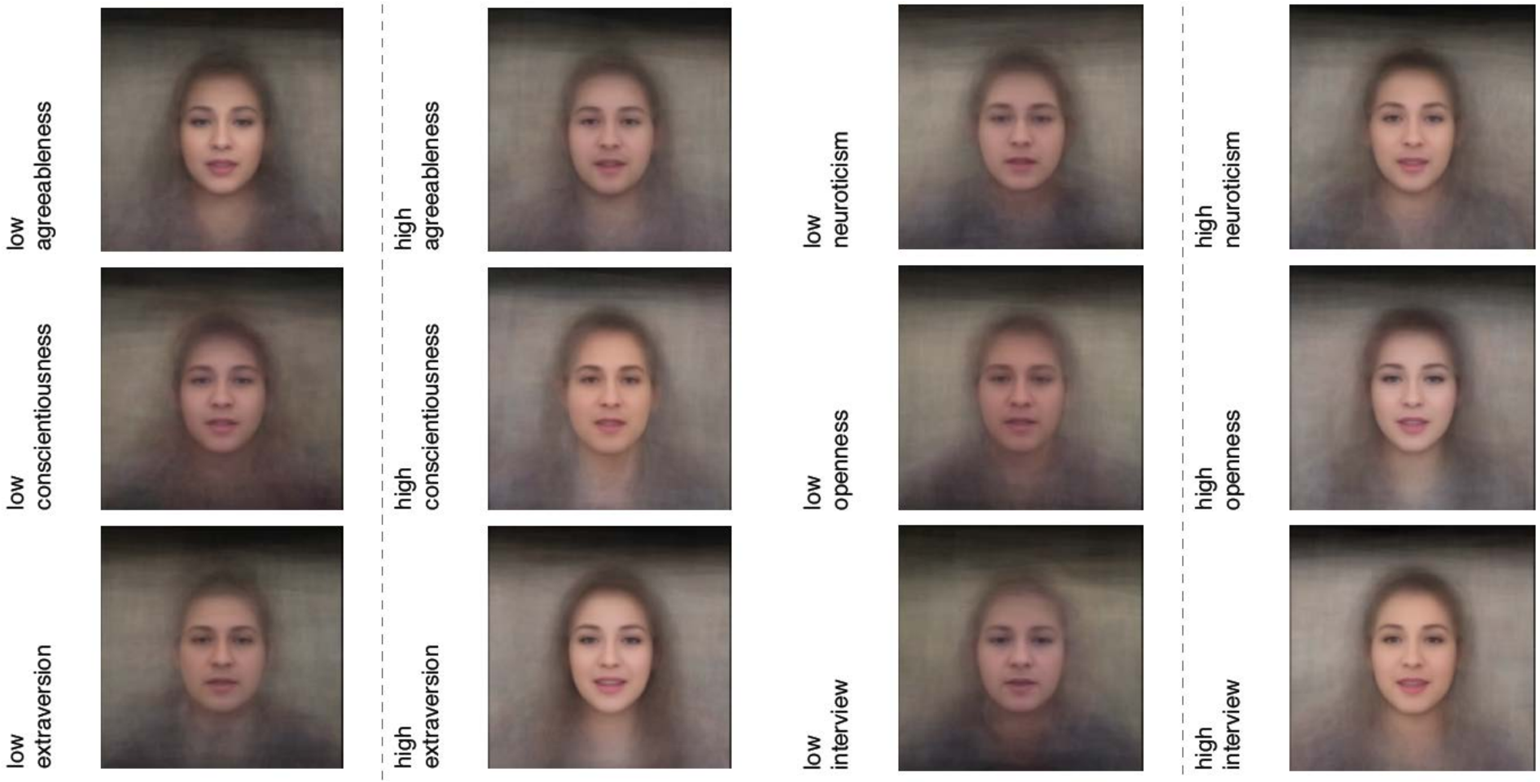


Screenshot of sample videos labeled

*V. Ponce-Lopez, et al., “**Chalearn lap 2016: First round challenge on first impressions-dataset and results**,” ECCVW, 2016.*

Apparent Personality Trait Analysis

ChaLearn First Impression (ECCV'16) - Round 1



Apparent Personality Trait Analysis

ChaLearn First Impression (ECCV’16) - Round 1

	Pretraining	Preprocessing	Modality				Fusion
			Audio		Video		
			R[a]	L[b]	R[a]	L[b]	
NJU-LAMDA [40]	VGG-face	-	Logfbank[c]	NN	CNN	CNN	Late
Evolgen [41]	-	Face alignment	Spectral	RCNN[j]	RCNN[j]	RCNN[j]	Early
DCC [42]	-	-	ResNet	ResNet+FC	ResNet	ResNet+FC	Late
ucas	VGG, AlexNet, ResNet	Face alignment	Spectral	PSLR[d], SVR[e]	CNN(face/scene)	PSLR[d], SVR[e]	Late
BU-NKU	VGG-face, FER2013	Face alignment	-	-	CNN(face/scene)	KELM[f]	Early
Pandora	-	Face alignment	LLD[h]	Bagged regressor	CNN(face/scene)	CNN	Early
Pilab	-	-	Spectral	RF regressor	-	-	-
Kaizoku	-	-	MFCC[i]/CNN	CNN	CNN	CNN	Late
ITU-SiMiT	VGG-face, VGG-16	Face detection	-	-	CNN(face/scene)	SVR[e]	Late

[a]R = Representation, [b]L = Learning Strategy, [c]logfbank = Logarithm Filterbank Energies, [d]PSLR = Partial Least Square Regressor, [e]SVR = Support Vector Regression, [f]KELM = Kernel Extreme Learning Machine, [g]FER = Facial Expression Recognition Dataset, [h]LLD = Low Level Descriptor, [i]MFCC = Mel Frequency Cepstral Coefficient, [j]RCNN = Recurrent Convolutional Neural Networks.

Apparent Personality Trait Analysis

ChaLearn First Impression (ICPR'16) - Round 2

- This is a follow up of the First impressions challenge at ECCV 2016
- Coopetition setting (**collaborate and compete**)
    - The coopetition feature allowed participants to **download other participant codes (from rounds 1 and 2)**
- 51 participants grouped in several teams. 4 teams entered the final phase
- Accuracy as 1 – regression_value_difference

Apparent Personality Trait Analysis

ChaLearn First Impression (ICPR’16) - Round 2

	Pretraining	Preprocessing	Modality				Fusion
			Audio		Video		
			R[1]	L[2]	R[1]	L[2]	
BU-NKU	VGG-face (FER2013), VGG-VD-19 (ILSVRC12)	face alignment	LLD[8]	-	CNN(face/scene), LGBPTOP (face)	KELM[6]	early
evolgen	-	face alignment	spectral	RCNN[10]	RCNN[10]	RCNN[10]	early
pandora	VGG-Net	face alignment	LLD[8]	Bagged Regressor	CNN(face/scene)	CNN	early
Pilab	-	-	spectral	RF regressor	-	-	-

[1] R = Representation [2] L = Learning Strategy [3] logfbank = Logarithm Filterbank Energies [4] PSLR = Partial Least Square Regressor [5] SVR = Support Vector Regression [6] KELM = Kernel Extreme Learning Machine [7] FER = Facial Expression Recognition Dataset [8] LLD = Low Level Descriptor [9] MFCC = Mel Frequency Cepstral Coefficient [10] RCNN = Recurrent CNN.

Improved results respect to first round

Apparent Personality Trait Analysis

Apparent Personality Trait Analysis

Job Candidate Screening Challenge

- Workshop on Explainable Computer Vision Multimedia and Job Candidate Screening Coopetition at **CVPR’17** and **IJCNN’17**

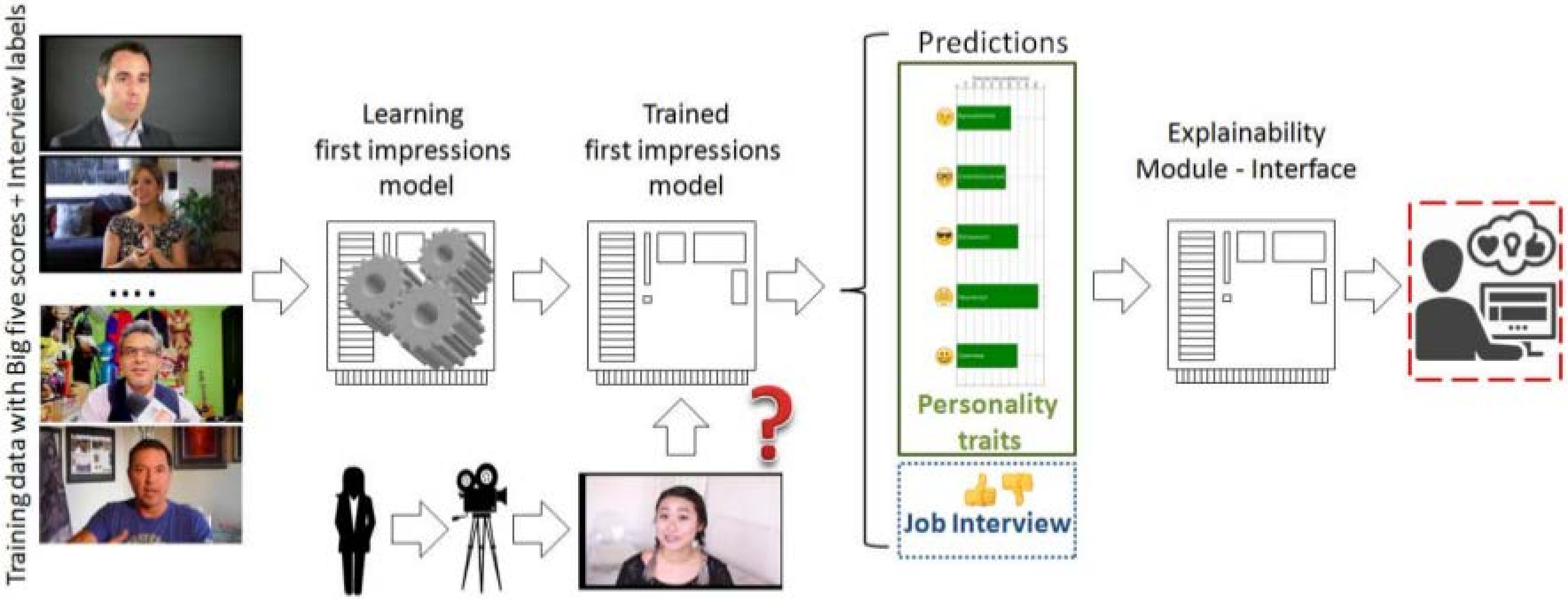
- ***First Impressions dataset*** <u>with the inclusion of</u>

    - **Transcriptions** (transcribed by the professional transcription service Rev.)
    - **Interview annotations** (whether the person should be invited or not to a job interview)

- ***Quantitative stage:*** predict 5 traits values plus whether the candidates are promising enough that the recruiter wants to invite him/her to an interview **(~70 participants)**

- ***Qualitative stage (coopetition, codes from quantitative stage available):*** justify/explain with a <u>text description</u> the recommendation made such that a human can understand it **(2 teams)**

Apparent Personality Trait Analysis
Job Candidate Screening Challenge



Apparent Personality Trait Analysis

Job Candidate Screening Challenge

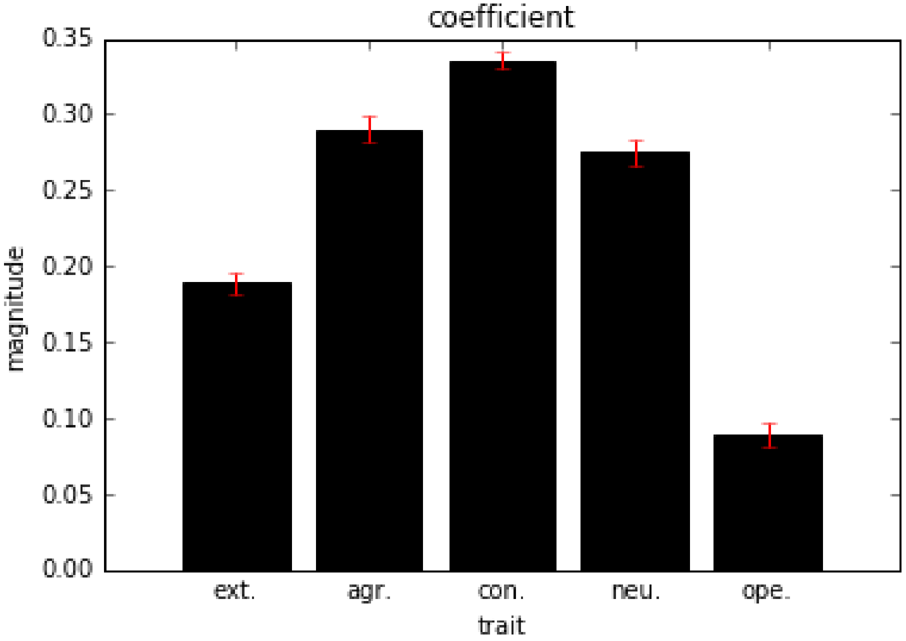
Evaluation metrics

- **Quantitative stage**
    - Same used in First Impression challenges, but now considering the interview recommendation variable
- **Qualitative stage (explanatory mechanisms)**
    - Performance will be evaluated in terms of the **creativity** and **explanatory effectiveness** of the mechanisms-interface (clarity, explainability, soundness, interpretability, creativity)
    - For this evaluation we invited a set of experts in the fields of psychological behavior analysis, recruitment, machine learning and computer vision

Apparent Personality Trait Analysis

Job Candidate Screening Challenge

Interview recommendation variable



- **Regression coefficients.** Bootstrapping was used to estimate means and standard deviations

- The trait annotations are highly predictive of the interview annotations, and the **coefficient of determination is significantly above chance. Conscientiousness** has the largest and **openness** has the smallest contributions to the predictions

Apparent Personality Trait Analysis

Job Candidate Screening Challenge

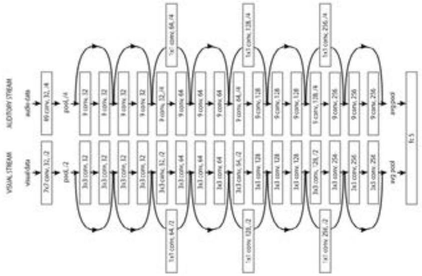
Baseline

Baseline results. Results are reported in terms of 1 - relative mean absolute error on the test set. AGR: Agreeableness; CON: Conscientiousness; EXT: Extroversion; $\overline{NEU}$: (non-)Neuroticism; OPE: Openness; AVE: average over trait results; INT: interview.

Model	AGR	CON	EXT	$\overline{NEU}$:	OPE	AVE	INT
language							
bag-of-words	0.8952	0.8786	0.8815	0.8794	0.8875	0.8844	0.8845
Skip-Thought Vec.	0.8971	0.8819	0.8839	0.8827	0.8881	0.8867	0.8865
sensory							
audio	0.9034	0.8966	0.8994	0.9000	0.9024	0.9004	0.9032
visual	0.9059	0.9073	0.9019	0.8997	0.9045	0.9039	0.9076
Audio-Visual	0.9102	0.9138	0.9107	0.9089	0.9111	0.9109	0.9159
language+sensory							
STV + AV	*0.9112*	*0.9152*	*0.9112*	*0.9104*	*0.9111*	*0.9118*	*0.9162*

Embedding by recurrent encoder-decoder neural network pretrained on the BookCorpus dataset

Resnets



Yagmur Gucluturk, Umut Guclu, Xavier Baro, Hugo Jair Escalante, Isabelle Guyon, Sergio Escalera, Marcel A. J. van Gerven, and Rob van Lier, Multimodal First Impression Analysis with Deep Residual Networks, IEEE Transactions on Affective Computing, TAC, 2017.

Apparent Personality Trait Analysis
Job Candidate Screening Challenge

Results (quantitative stage)

Rank	Team	Invite-Interview *	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
1	BU-NKU	0.920916 (1)	0.913731 (1)	0.919769 (1)	0.921289 (1)	0.914613 (1)	0.917014 (1)
-	baseline	0.916202 (2)	0.911230 (2)	0.915228 (2)	0.911220 (3)	0.910378 (2)	0.911123 (2)
2	PML	0.915746 (3)	0.910312 (3)	0.913775 (3)	0.915510 (2)	0.908297 (3)	0.910078 (3)
3	ROCHCI	0.901859 (4)	0.903216 (4)	0.894914 (4)	0.902660 (4)	0.901147 (4)	0.904709 (4)
4	FDMB	0.872129 (5)	0.891004 (5)	0.865975 (5)	0.878842 (5)	0.863237 (5)	0.874761 (5)

Yagmur Gucluturk, Umut Guclu, Xavier Baro, Hugo Jair Escalante, Isabelle Guyon, Sergio Escalera, Marcel A. J. van Gerven, and Rob van Lier, Multimodal First Impression Analysis with Deep Residual Networks, IEEE Transactions on Affective Computing, TAC, 2017.

Apparent Personality Trait Analysis

Job Candidate Screening Challenge

Winner approach (quantitative stage)

- BU-NKU team extracted **LGBP and deep convolutional network** based features from facial images. **Audio descriptors** were extracted using the openSMILE library. Finally, the authors used **a DCNN to extract scene information**, in an effort to extract useful contextual information

- This team also performed preliminary experiments including **ASR transcripts**, however text descriptors were not included in the winning entry

- Features were combined via linear **kernel extreme learning machines** (ELM). Then, a **random forest model combined the outputs of the two ELMs**. This method was trained to predict the interview variable and the 5 personality traits

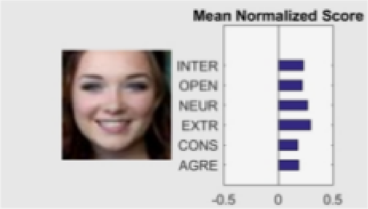
Apparent Personality Trait Analysis
Job Candidate Screening Challenge
Results (qualitative stage)

- The two teams completing the final phase of the qualitative stage
    - Other teams also developed solutions to the explainability track, but did not succeed in submitting predictions for the test videos

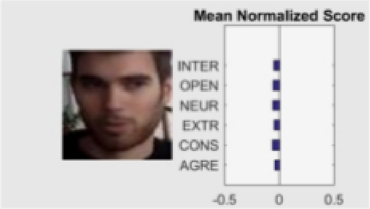
Rank	Team	Clarity	Explainability	Soundness	Interpretability	Creativity	Mean score
1	BU-NKU	4.31	3.58	3.4	3.83	2.67	3.56
1	TUD	3.33	3.23	3.43	2.4	3.4	3.16

Apparent Personality Trait Analysis
Job Candidate Screening Challenge
Results (qualitative stage)

- Sample verbal and visual explanations from qualitative stage for the BU-NKU entry



This lady is invited for an interview due to her high apparent agreeableness and non-neuroticism impression. The impressions of agreeableness, conscientiousness, extroversion, non-neuroticism and openness are primarily gained from facial features.

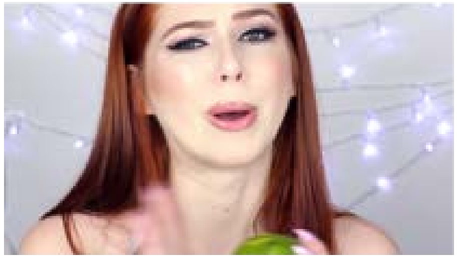


This gentleman is not invited due to his low apparent agreeableness, non-neuroticism, conscientiousness, extroversion and openness scores. The impression of conscientousness is modulated by facial expressions. Furthermore, the impressions of agreeableness, extroversion, non-neuroticism and openness are gained particularly through vocal features.

Apparent Personality Trait Analysis
Job Candidate Screening Challenge

Results (qualitative stage)

- Sample verbal and visual explanations from qualitative stage for the TUD entry



****
* ASSESSMENT REPORT FOR VIDEO 2c42A4Z7qPE.001.mp4: *
****

On a scale from 0.0 to 1.0, I would rate this person's interviewability as 0.497947.
Below, I will report on linguistic and visual assessment of the person.
Percentiles are obtained by comparing the person against scores of 6000 earlier assessed people.

****
* USE OF LANGUAGE *
****

Here is the report on the person's language use:

** FEATURES OBTAINED FROM SIMPLE TEXT ANALYSIS **
Cognitive capability may be important for the job. I looked at a few very simple text statistics first.

*** Amount of spoken words ***
This feature typically ranges between 0.000000 and 90.000000. The score for this video is 29.000000 (percentile: 25).
In our model, a higher score on this feature typically leads to a higher overall assessment score.

****
* VISUAL FEATURES *
****

Here is the report on what I could 'see':

*** Action Unit 12: how often was the lip corner pulled? ***
This feature typically ranges between 0.000000 and 1.000000. The score for this video is 0.148148 (percentile: 82).

*** Action Unit 12: how much was the lip corner pulled on average? ***
This feature typically ranges between 0.000000 and 2.880709. The score for this video is 0.333867 (percentile: 81).

Apparent Personality Trait Analysis

The subjectivity contained in the data

- The labels used in both challenges *may reflect the bias towards the person in the videos*, even though it **may be *unintentional* and *subconscious***

- We analyzed, in a *post-challenge* stage, the existence of bias towards **gender** and **ethnicity** (from new labeled data)

- Some points could be observed:

    - There is an **overall positive attitude towards females** in both *personality traits* (except Agreeableness) and *job interview* invitation

    - There is an **overall positive bias towards Caucasians**, and a **negative bias towards African-Americans**. There is no discernible bias towards Asians in either way

    - ***Gender bias*** is **stronger** compared to ethnicity bias

Apparent Personality Trait Analysis

How to deal with these subjectivities?

- **Explainability** could be an ***effective way to overcome data biases***, or at least to point out potential bias so that decision takers can take them into account
  - Explainable mechanisms could use data-bias information to provide ***explanations* on their recommendations**
- Future research, already addressed someway in the field of psychology, could focus on the **external observers** rather than on the variable predictions
  - Providing **correlation analysis from the observer point of view** (i.e., what he/she is labeling/looking at) in order to identify potential patterns and bias (explanations)
  - **Bias analysis in terms of real vs apparent personality**
  - All previous is challenging and time consuming task, and may **requires new data definition and protocols**

Final remarks

- Analysis of humans involve **subjective evaluation of attributes (so many variables involved)**
- **Apparent personality** is a hot topic with a lot of work to be done:
  - Behavioral models
  - Labels and quantification
  - Raters
  - Real from apparent personality analysis
  - Applications
  - Multi-disciplinary (vision, ML, psychology) and Multi-modality (vision and language, plus questionnaires and particular tests may be required…)
  - …

http://chalearnlap.cvc.uab.es/

Thank you!

Sergio Escalera
www.sergioescalera.com
http://chalearnlap.cvc.uab.es/
sergio.escalera.guerrero@Gmail.com