6th Author Profiling task at PAN
Multimodal Gender Identification in Twitter

PAN-AP-2018 CLEF 2018
Avignon, 10-14 September

Francisco Rangel
Autoritas Consulting &
PRHLT Research Center -
Universitat Politècnica de València

Paolo Rosso
PRHLT Research Center
Universitat Politècnica de Valencia

Manuel Montes y Gómez
INAOE - Mexico

Martin Potthast & Benno Stein
Bauhaus-Universität Weimar

Introduction

Author profiling aims at identifying **personal traits** such as age, **gender**, personality traits, native language, language variety… from writings?

This is crucial for:

- Marketing.
- Security.
- Forensics.

Task goal

To investigate the identification of author's **gender** with multimodal information: texts + images.

Three languages:

Arabic **English** **Spanish**

Corpus

- PAN-AP'17 subset extended with images shared in author's timelines:
  - 100 tweets per author.
  - 10 images per author.

	(AR) Arabic	(EN) English	(ES) Spanish	Total
Training	1,500	3,000	3,000	7,500
Test	1,000	1,900	2,200	5,100
Total	2,500	4,900	5,200	12,600

Evaluation measures

The **accuracy** is calculated per *modality* and language:

- Text-based.
- Image-based.
- Combined.

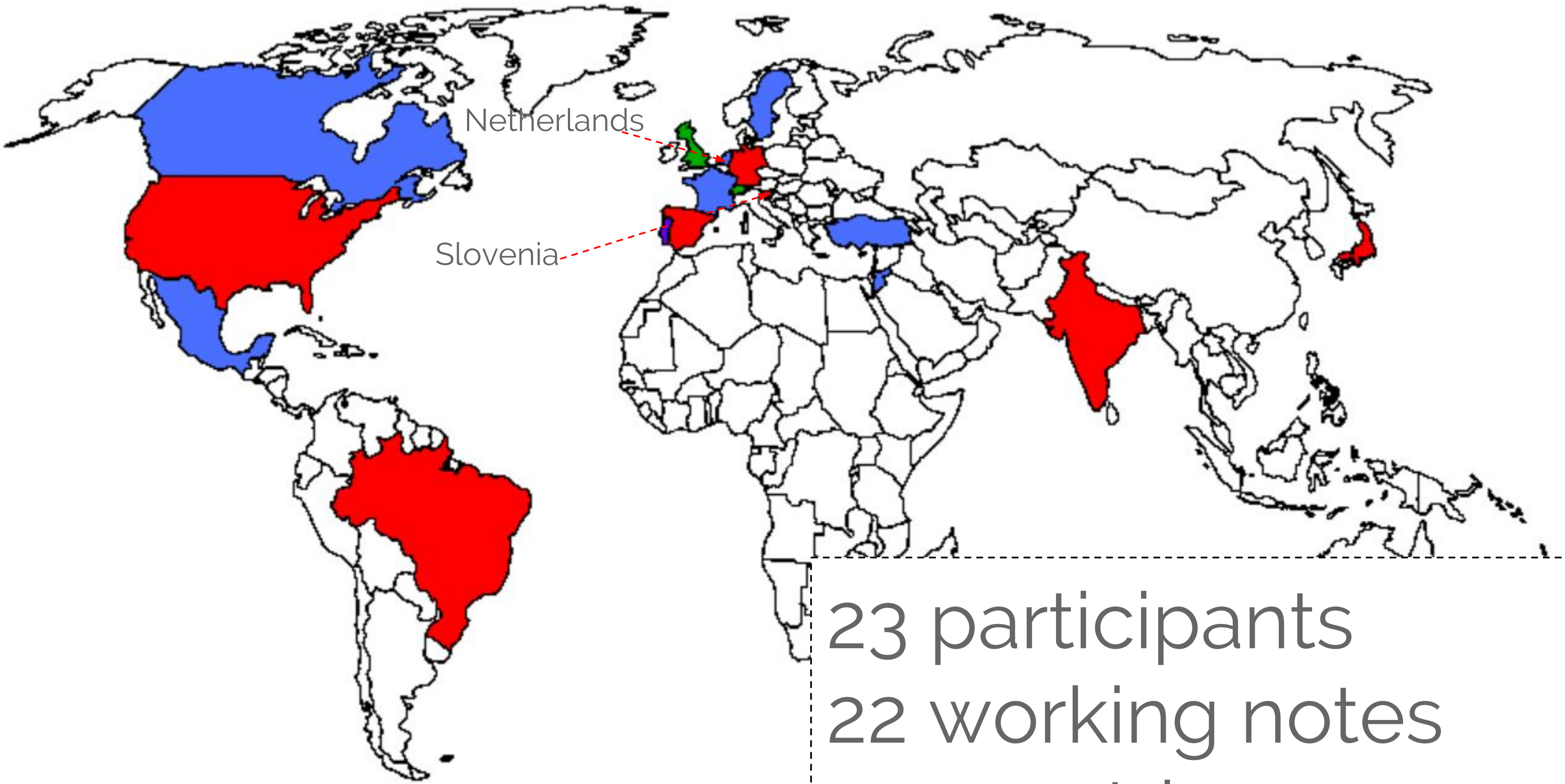
The final ranking is the average of the combined* accuracies per language:

$$ranking = \frac{acc_{ar} + acc_{en} + acc_{es}}{3}$$

* If only the textual approach was submitted, its accuracy has been used

Baselines

- **BASELINE-stat:** A statistical baseline that emulates random choice. Both modalities.
- **BASELINE-bow:**
  - Documents represented as bag-of-words.
  - The 5,000 most common words in the training set.
  - Weighted by absolute frequency.
  - Preprocess: lowercase, removal of punctuation signs and numbers, removal of stopwords.
  - Textual modality.
- **BASELINE-rgb:**
  - RGB color for each pixel in each author images.
  - The author is represented with the minimum, maximum, mean, median, and standard deviation of the RGB values.
  - Images modality.



23 participants
22 working notes
17 countries

Approaches

What kind of ...

Preprocessing

Features

Methods

... did the teams perform?

Approaches - Preprocessing

TEXTS	Punctuation signs	Ciccone *et al.*, Stout *et al.*, HaCohen-Kerner *et al.*, Veenhoven *et al.*
	Character flooding	Ciccone *et al.*, Raiyani *et al.*
	Lowercase	Von Däniken *et al.*, Veenhoven *et al.*, Nieuwenhuis *et al.*, Bayot & Gonçalves, Kosse *et al.*, Stout *et al.*, Schaetti, HaCohen-Kerner *et al.*
	Stopwords	Ciccone *et al.*, Raiyani *et al.*, HaCohen-Kerner *et al.*, Veenhoven *et al.*
	Twitter specific components: hashtags, urls, mentions and RTs	Ciccone *et al.*, Takahashi *et al.*, Stout *et al.*, Raiyani *et al.*, Schaetti, HaCohen-Kerner *et al.*, Von Däniken *et al.*, Martinc *et al.*, Veenhoven *et al.*, Nieuwenhuis *et al.*, Kosse *et al.*
	Contractions and abbreviations	Stout *et al.*, Raiyani *et al.*
	Normalisation and diacritics removal in Arabic	Ciccone *et al.*
IMAGES	Resizing, rescaling	Takahashi *et al.*, Martinc *et al.*, Sierra-Loaiza & González
	Normalisation (subtracting the average RGB value per lang)	Takahashi *et al.*

Approaches - Textual Features

Stylistic features: - Ratios of links - Hashtag or user mentions - Character flooding - Emoticons / laugher expressions - Domain names	Patra *et al.*, Karlgren *et al.* ,HaCohen-Kerner *et al.,* Von Däniken *et al.*
N-gram models	Stout *et al.,* Sandroni-Dias & Paraboni, López-Santillán *et al.,* Von Däniken *et al.,* Tellez *et al.*, Nieuwenhuis *et al.*, Kosse *et al.,* Daneshvar, HaCohen-Kerner *et al.,* Ciccone *et al.,* Aragón & López
LSA	Patra *et al.*
Second order representation	Áragon & López
A variation of LDSE	Gàribo-Orts
Word embeddings	Martinc *et al.*, Veenhoven *et al.*, Bayot & Gonçalves, López-Santillán *et al.*, Takahashi *et al.*, Patra *et al.*
Character embeddings	Schaetti

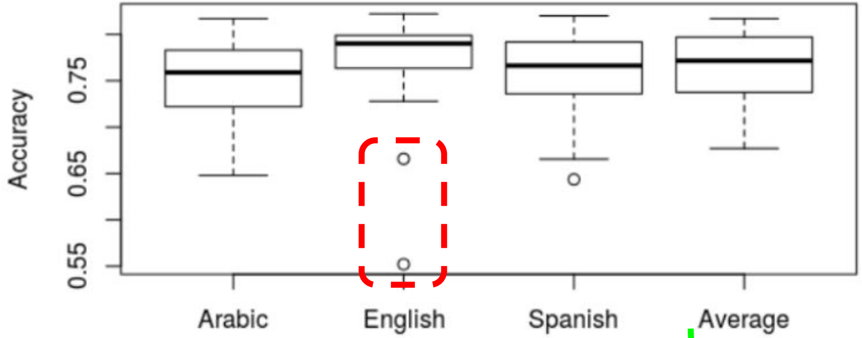
Approaches - Image Features

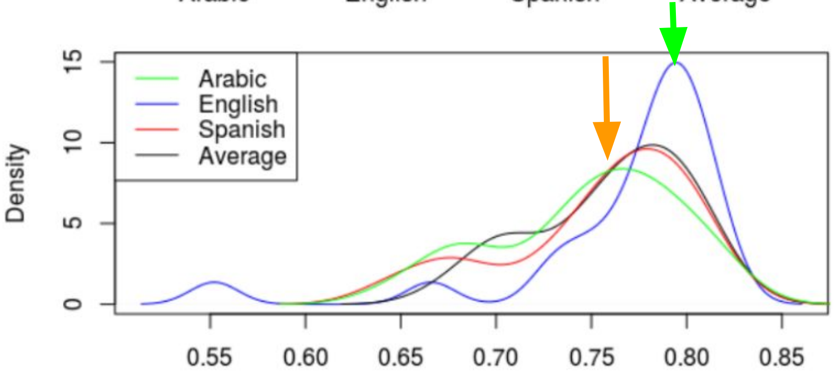
Face detection	Stout *et al.*, Ciccone *et al.*, Veenhoven *et al.*
Objects detection	Ciccone *et al.*
Local binary patterns	Ciccone *et al.*
Hand-crafted features	HaCohen-Kerner *et al.*
Color histogram	Ciccone *et al.*, HaCohen-Kerner *et al.*
Bag of Visual Words	Tellez *et al.*
Image resources and tools (e.g. ImageNet, TorchVision...)	Patra *et al.*, Nieuwenhuis *et al.*, Aragón & López, Schaetti, Takahashi *et al.*

Approaches - Methods

Logistic regression	Sandroni-Dias & Paraboni, HaCohen-Kerner *et al.*, Von Däniken *et al.*, Nieuwenhuis *et al.*
SVM	López-Santillán *et al.*, Aragón & López, Ciccone *et al.*, Patra *et al.*, Tellez *et al.*, Veenhoven *et al.*
Multilayer Perceptron	HaCohen-Kerner *et al.*
Basic feed-forward network	Kosse *et al.*
Distance-based method	Tellez *et al.*, Karlgren *et al.*
IF condition	Gáribo-Orts
RNN	Takahashi *et al.*, Bayot & Gonçalves, Stout *et al.*
CNN	Schaetti
ResNet18	Schaetti
Bi-LSTM	Veenhoven *et al.*

Textual modality

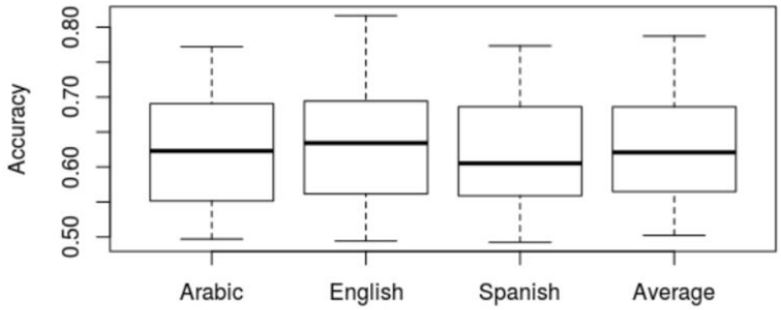


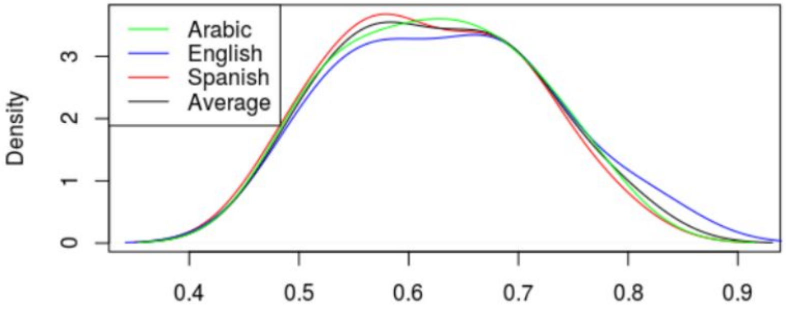


- **AR:** n-grams
- **EN:** n-grams
- **ES:** n-grams

Ranking	Team	Arabic	English	Spanish	Average
1	Daneshvar	0.8090	**0.8221**	**0.8200**	**0.8170**
2	Tellez *et al.*	**0.8170**	0.8121	0.8005	0.8099
3	Nieuwenhuis & Wilkens	0.7830	0.8116	0.8027	0.7991
4	Sierra-Loaiza & González	0.8120	0.8011	0.7827	0.7986
5	Ciccone *et al.*	0.7910	0.8074	0.7959	0.7981
6	Kosse *et al.*	0.7920	0.8074	0.7918	0.7971
7	Takahashi *et al.*	0.7710	0.7968	0.7864	0.7847
8	Veenhoven *et al.*	0.7490	0.7926	0.8036	0.7817
9	Martinc *et al*	0.7760	0.7900	0.7782	0.7814
10	López-Santillán *et al.*	0.7760	0.7847	0.7677	0.7761
11	Hacohen-Kerner *et al.* (B)	0.7590	0.7911	0.7650	0.7717
12	Hacohen-Kerner *et al.* (A)	0.7590	0.7911	0.7650	0.7717
13	Stout *et al.*	0.7600	0.7853	0.7405	0.7619
14	Gopal-Patra *et al.*	0.7430	0.7558	0.7586	0.7525
15	von Däniken *et al.*	0.7320	0.7742	0.7464	0.7509
16	Schaetti	0.7390	0.7711	0.7359	0.7487
	baseline-bow	0.7480	0.7411	0.7255	0.7382
17	Aragon & Lopez	0.6480	0.7963	0.7686	0.7376
18	Bayot & Gonçalves	0.6760	0.7716	0.6873	0.7116
19	Garibo	0.6750	0.7363	0.7164	0.7092
20	Sezerer *et al.*	0.6920	0.7495	0.6655	0.7023
21	Raiyani *et al.*	0.7220	0.7279	0.6436	0.6978
22	Sandroni-Dias & Paraboni	0.6870	0.6658	0.6782	0.6770
	baseline-stats	0.5000	0.5000	0.5000	0.5000
23	Karlgren *et al.*	-	0.5521	-	-

Images modality

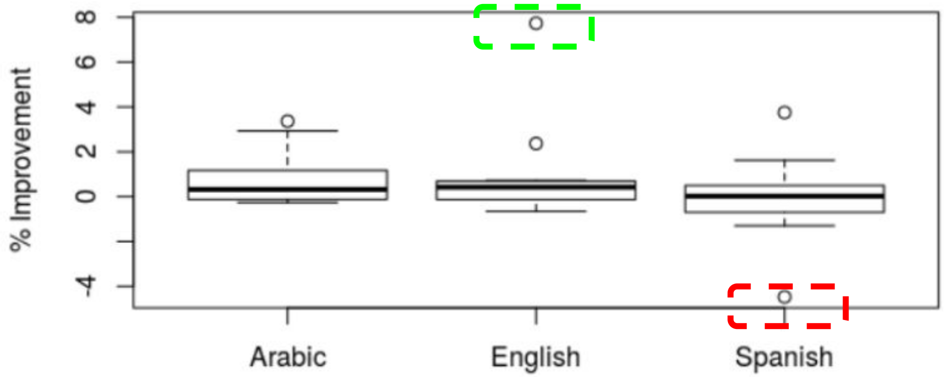




Ranking	Team	Arabic	English	Spanish	Average
1	Takahashi *et al.*	**0.7720**	**0.8163**	**0.7732**	**0.7872**
2	Sierra-Loaiza & González	0.7280	0.7442	0.7100	0.7274
3	Ciccone *et al.*	0.7010	0.6963	0.6805	0.6926
4	Aragon & Lopez	0.6800	0.6921	0.6668	0.6796
5	Gopal-Patra *et al.*	0.6570	0.6747	0.6918	0.6745
6	Stout *et al.*	0.6230	0.6584	0.6232	0.6349
7	Nieuwenhuis & Wilkens	0.6230	0.6100	0.5873	0.6068
8	Tellez *et al.*	0.5900	0.5468	0.5691	0.5686
9	Schaetti	0.5430	0.5763	0.5782	0.5658
10	Martinc *et al.*	0.5600	0.5826	0.5486	0.5637
	baseline-rgb	0.5410	0.5179	0.5191	0.5260
11	Hacohen-Kerner *et al.* (B)	0.5100	0.4942	0.5027	0.5023
12	Hacohen-Kerner *et al.* (A)	0.4970	0.5174	0.4923	0.5022
	baseline-stats	0.5000	0.5000	0.5000	0.5000

- **Best:** Pre-trained CNN w. ImageNet
- **2nd. AR:** VGG16 + ResNet50 from ImageNet
- **2nd. EN:** VGG16 + ResNet50 from ImageNet
- **2nd. ES:** Color histogram + faces + objects + local binary patterns

Improvement with images

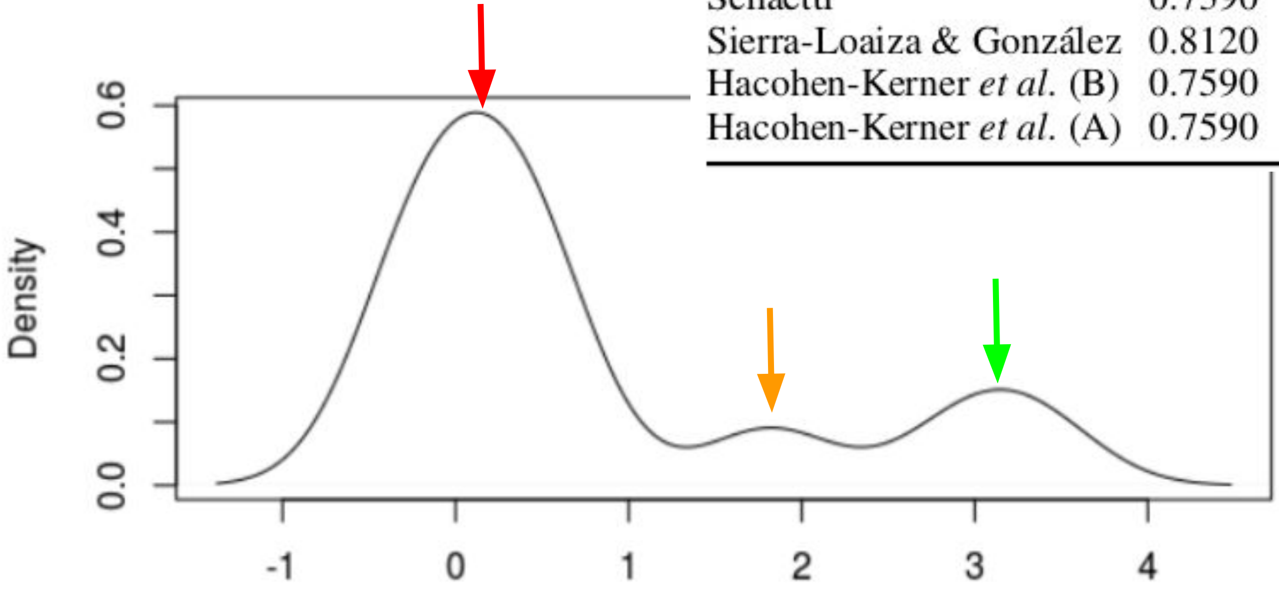


	Arabic	English	Spanish
Min	-0.2635	-0.6526	-4.4717
Q1	-0.0616	-0.0647	-0.6613
Median	0.3185	0.4249	0.0257
Mean	0.7613	1.0102	-0.0609
SDev	1.2513	2.2473	1.9087
Q3	0.8487	0.6788	0.4898
Max	3.3647	7.7309	3.7513
Skewness	1.2095	2.4716	-0.3778
Kurtosis	2.9616	8.0027	4.4883
Normality (p-value)	0.0010	0.0000	0.1316

- In average, there is almost no improvement.
- Some systems obtain high improvements (up to 7.73%)
  - Pre-trained CNN w. ImageNet.

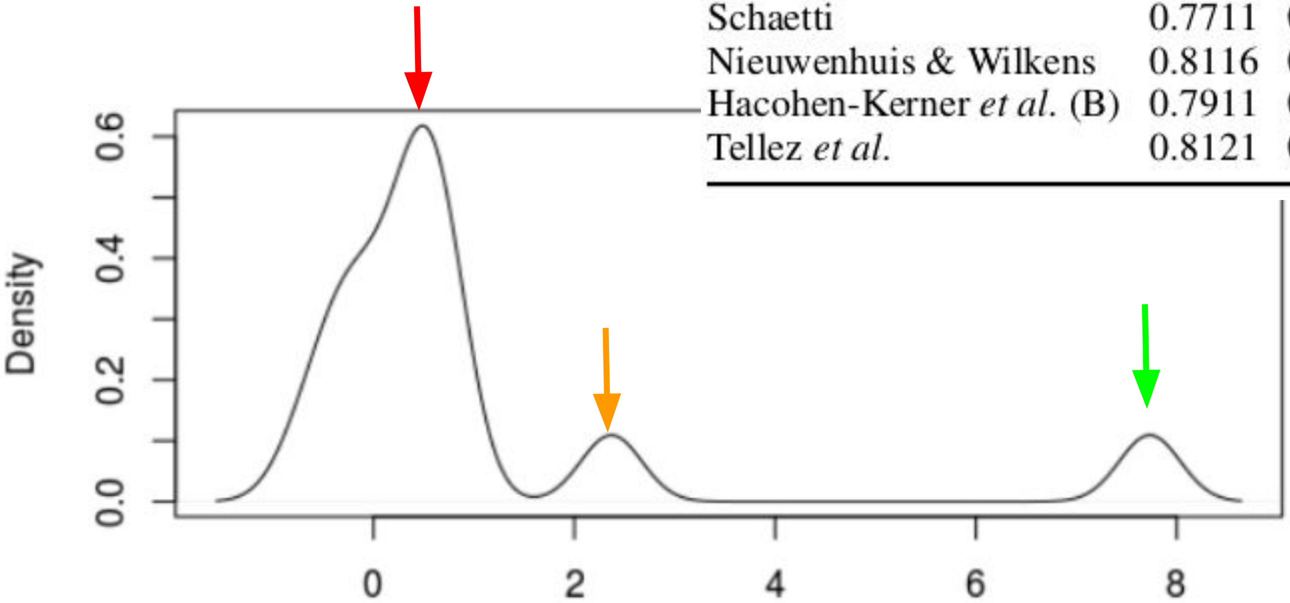
Improvement (AR)

Team	Texts	Images	Combined	Improvement
Gopal-Patra *et al.*	0.7430	0.6570	0.7680	3.3647%
Aragon & Lopez	0.6480	0.6800	0.6670	2.9321%
Takahashi *et al.*	0.7710	0.7720	0.7850	1.8158%
Stout *et al.*	0.7600	0.6230	0.7640	0.5263%
Nieuwenhuis & Wilkens	0.7830	0.6230	0.7870	0.5109%
Ciccone *et al.*	0.7910	0.7010	0.7940	0.3793%
Martinc *et al*	0.7760	0.5600	0.7780	0.2577%
Tellez *et al.*	0.8170	0.5900	0.8180	0.1224%
Schaetti	0.7390	0.5430	0.7390	0.0000%
Sierra-Loaiza & González	0.8120	0.7280	0.8100	-0.2463%
Hacohen-Kerner *et al.* (B)	0.7590	0.5100	0.7570	-0.2635%
Hacohen-Kerner *et al.* (A)	0.7590	0.4970	0.7570	-0.2635%



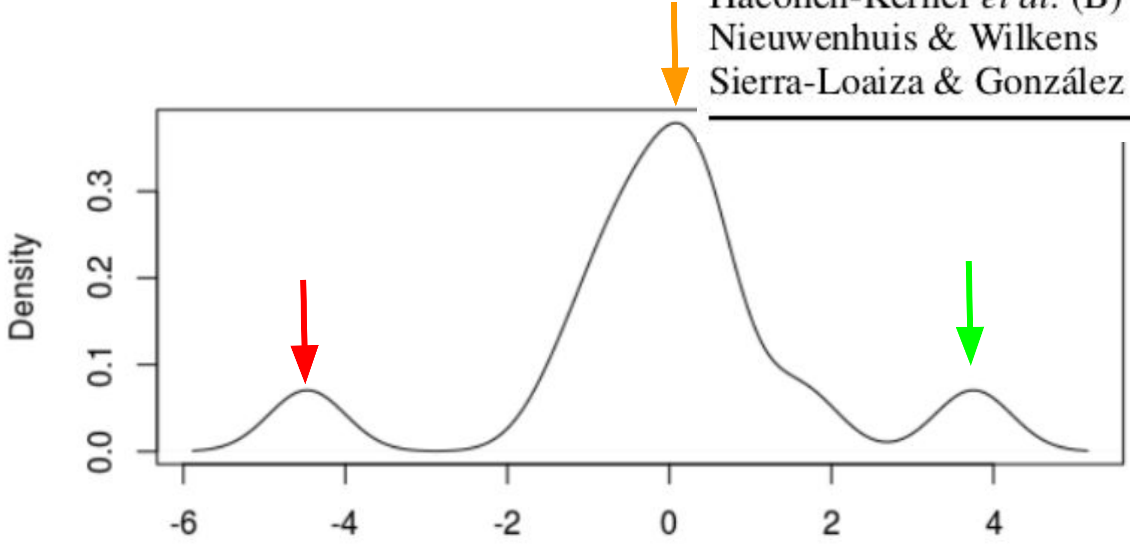
Improvement (EN)

Team	Texts	Images	Combined	Improvement
Takahashi *et al.*	0.7968	0.8163	0.8584	7.7309
Gopal-Patra *et al.*	0.7558	0.6747	0.7737	2.3684
Ciccone *et al.*	0.8074	0.6963	0.8132	0.7184
Aragon & Lopez	0.7963	0.6921	0.8016	0.6656
Sierra-Loaiza & González	0.8011	0.7442	0.8063	0.6491
Hacohen-Kerner *et al.* (A)	0.7911	0.5174	0.7947	0.4551
Stout *et al.*	0.7853	0.6584	0.7884	0.3948
Martinc *et al.*	0.7900	0.5826	0.7926	0.3291
Schaetti	0.7711	0.5763	0.7711	0.0000
Nieuwenhuis & Wilkens	0.8116	0.6100	0.8095	-0.2587
Hacohen-Kerner *et al.* (B)	0.7911	0.4942	0.7889	-0.2781
Tellez *et al.*	0.8121	0.5468	0.8068	-0.6526

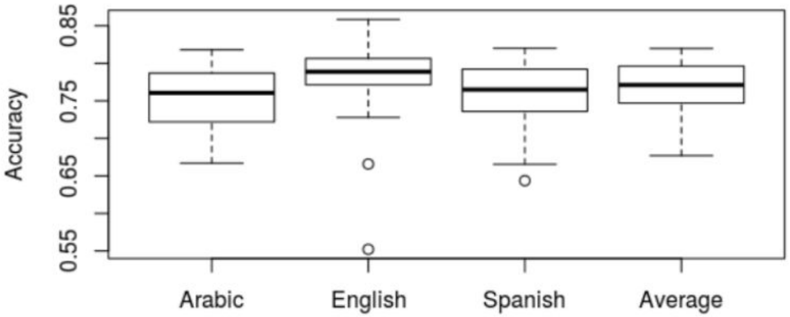


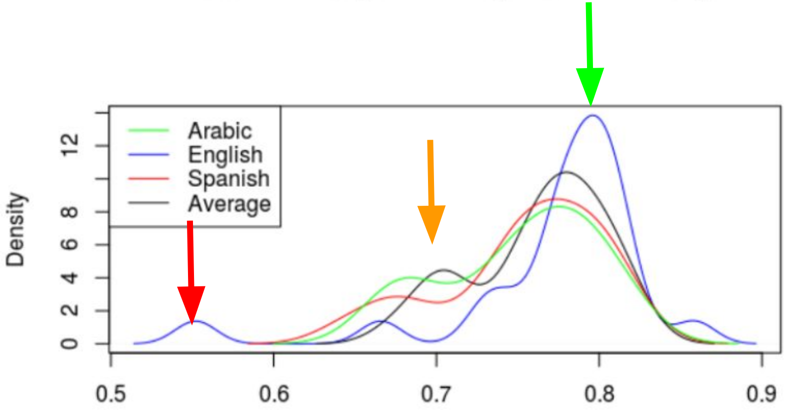
Improvement (ES)

Team	Texts	Images	Combined	Improvement
Takahashi *et al.*	0.7864	0.7732	0.8159	3.7513
Gopal-Patra *et al.*	0.7586	0.6918	0.7709	1.6214
Ciccone *et al.*	0.7959	0.6805	0.8000	0.5151
Aragon & Lopez	0.7686	0.6668	0.7723	0.4814
Stout *et al.*	0.7405	0.6232	0.7432	0.3646
Martinc *et al.*	0.7782	0.5486	0.7786	0.0514
Schaetti	0.7359	0.5782	0.7359	0.0000
Hacohen-Kerner *et al.* (A)	0.7650	0.4923	0.7623	-0.3529
Tellez *et al.*	0.8005	0.5691	0.7955	-0.6246
Hacohen-Kerner *et al.* (B)	0.7650	0.5027	0.7591	-0.7712
Nieuwenhuis & Wilkens	0.8027	0.5873	0.7923	-1.2956
Sierra-Loaiza & González	0.7827	0.7100	0.7477	-4.4717



Final ranking





Ranking	Team	Arabic	English	Spanish	Average
1	Takahashi *et al.*	0.7850	**0.8584**	0.8159	**0.8198**
2	Daneshvar	0.8090	0.8221	**0.8200**	0.8170
3	Tellez *et al.*	**0.8180**	0.8068	0.7955	0.8068
4	Ciccone *et al.*	0.7940	0.8132	0.8000	0.8024
5	Kosse *et al.*	0.7920	0.8074	0.7918	0.7971
6	Nieuwenhuis & Wilkens	0.7870	0.8095	0.7923	0.7963
7	Sierra-Loaiza & González	0.8100	0.8063	0.7477	0.7880
8	Martinc *et al.*	0.7780	0.7926	0.7786	0.7831
9	Veenhoven *et al.*	0.7490	0.7926	0.8036	0.7817
10	López-Santillán *et al.*	0.7760	0.7847	0.7677	0.7761
11	Hacohen-Kerner *et al.* (A)	0.7570	0.7947	0.7623	0.7713
12	Gopal-Patra *et al.*	0.7680	0.7737	0.7709	0.7709
13	Hacohen-Kerner *et al.* (B)	0.7570	0.7889	0.7591	0.7683
14	Stout *et al.*	0.7640	0.7884	0.7432	0.7652
15	von Däniken *et al.*	0.7320	0.7742	0.7464	0.7509
16	Schaetti	0.7390	0.7711	0.7359	0.7487
17	Aragon & Lopez	0.6670	0.8016	0.7723	0.7470
18	Bayot & Gonçalves	0.6760	0.7716	0.6873	0.7116
19	Garibo	0.6750	0.7363	0.7164	0.7092
20	Sezerer *et al.*	0.6920	0.7495	0.6655	0.7023
21	Raiyani *et al.*	0.7220	0.7279	0.6436	0.6978
22	Sandroni-Dias & Paraboni	0.6870	0.6658	0.6782	0.6770
23	Karlgren *et al.*	-	0.5521	-	-

PAN-AP 2018 best results

Language	Textual	Images	Combined
Arabic	0.8170	0.7720	0.8180
English	0.8221	0.8163	0.8584
Spanish	0.8200	0.7732	0.8200

Conclusions

- Several approaches to tackle the task:
  - Deep learning prevailing.
- Textual classification:
  - Best results regarding textual subtask: n-grams + traditional methods (SVM, logistic reg.).
  - The second best result for Spanish: bi-LSTM with word embeddings.
- Images classification approaches based on:
  - Face recognition. <- Failed!
  - Pre-trained models and image processing tools such as ImageNet. <- Best results obtained with semantic features extracted from the images.
  - Hand-crafted features such as color histograms and bag-of-visual-words.
- Texts vs. Images:
  - Textual features discriminate better than images.
  - On average, there is no improvement when images are used.
  - Elaborated representations improves up to 7.73% (English).
- Best results:
  - Over 80% on average (EN 85.84%; ES 82%; AR: 81.80%).
  - English (85.84%): Takahashi et al. with deep learning techniques (RNN for text, ImageNet + CNN for images).
  - Spanish (82%): Daneshvar with SVM and combinations of n-grams (only textual features).
  - Arabic (81.80%): Tellez et al. with SVM + n-grams, and Bag of Visual Words.
- Insight:
  - Traditional approaches still remain competitive, but deep learning is acquiring strength.

Task impact

	PARTICIPANTS	COUNTRIES
PAN-AP 2013	21	16
PAN-AP 2014	10	8
PAN-AP 2015	22	13
PAN-AP 2016	22	15
PAN-AP 2017	22	19
PAN-AP 2018	23	17

Industry at PAN (Author Profiling)

Organisation

autoritas

Participants

FUJI xerox

ANCHORMEN

OTS
OPTICAL TECH & SUPPORT

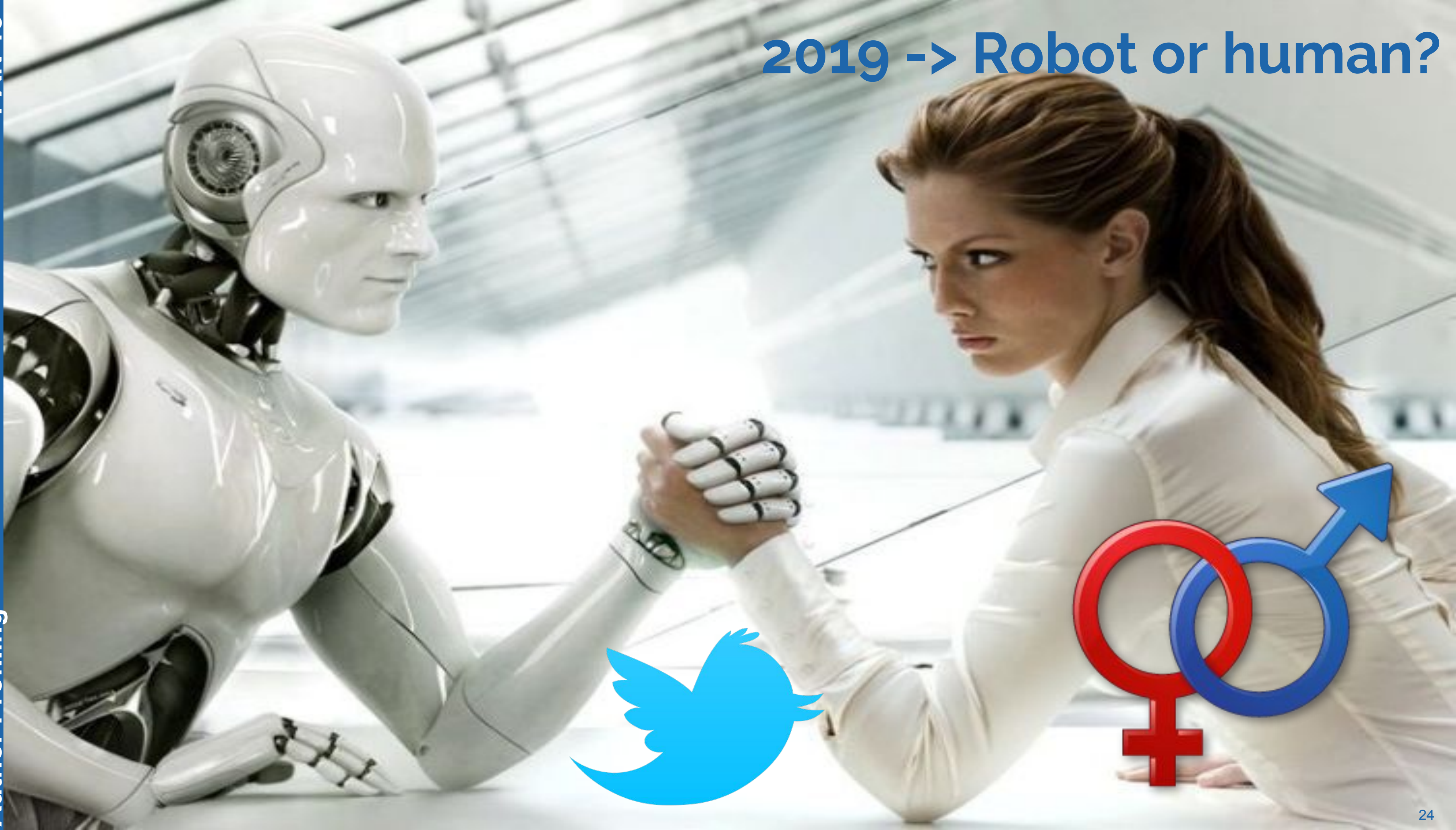
Sponsors

meaning cloud

symanto GROUP

ARABIC AUTHOR PROFILING
FOR CYBER SECURITY
FUNDED BY THE QATAR NATIONAL RESEARCH FUND

2019 -> Robot or human?



On behalf of the author profiling task organisers:

Thank you very much for participating
and hope to see you next year!!