Stacked Gender Prediction from Tweet Texts and Image

Giovanni Ciccone, Arthur Sultan*,**, Léa Laporte*, Előd Egyed-Zsigmond*, Alaa Alhamzeh*,** and Michael Granitzer***

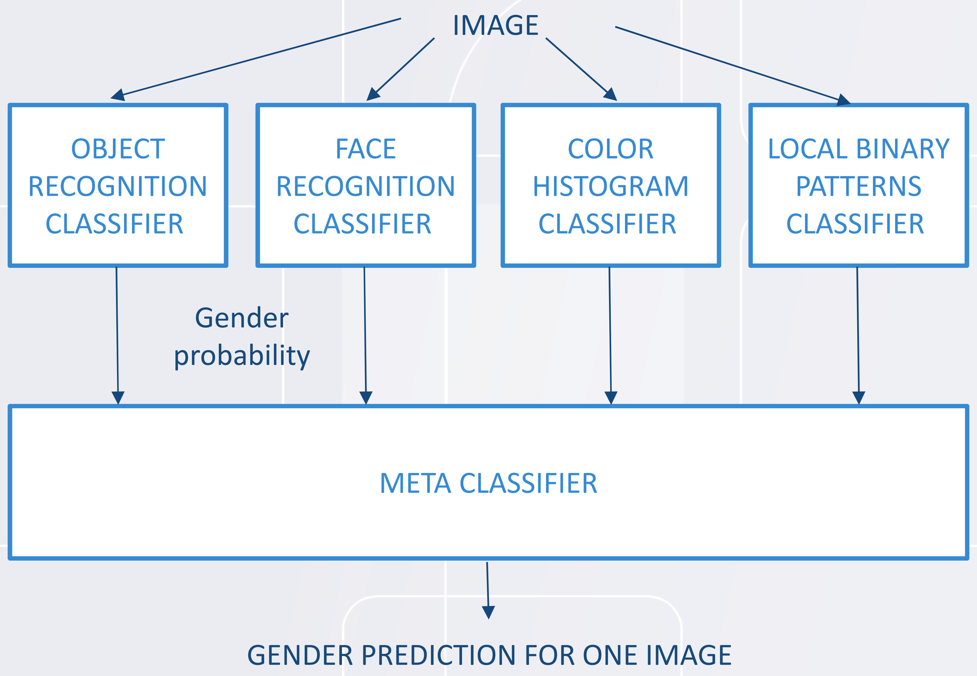
** Université de Lyon - INSA Lyon - LIRIS UMR5205 : {firstname.lastname}@insa-lyon.fr*

*** Universität Passau : michael.granitzer@uni-passau.de*

Laboratoire d'InfoRmatique en Image et Systèmes d'information

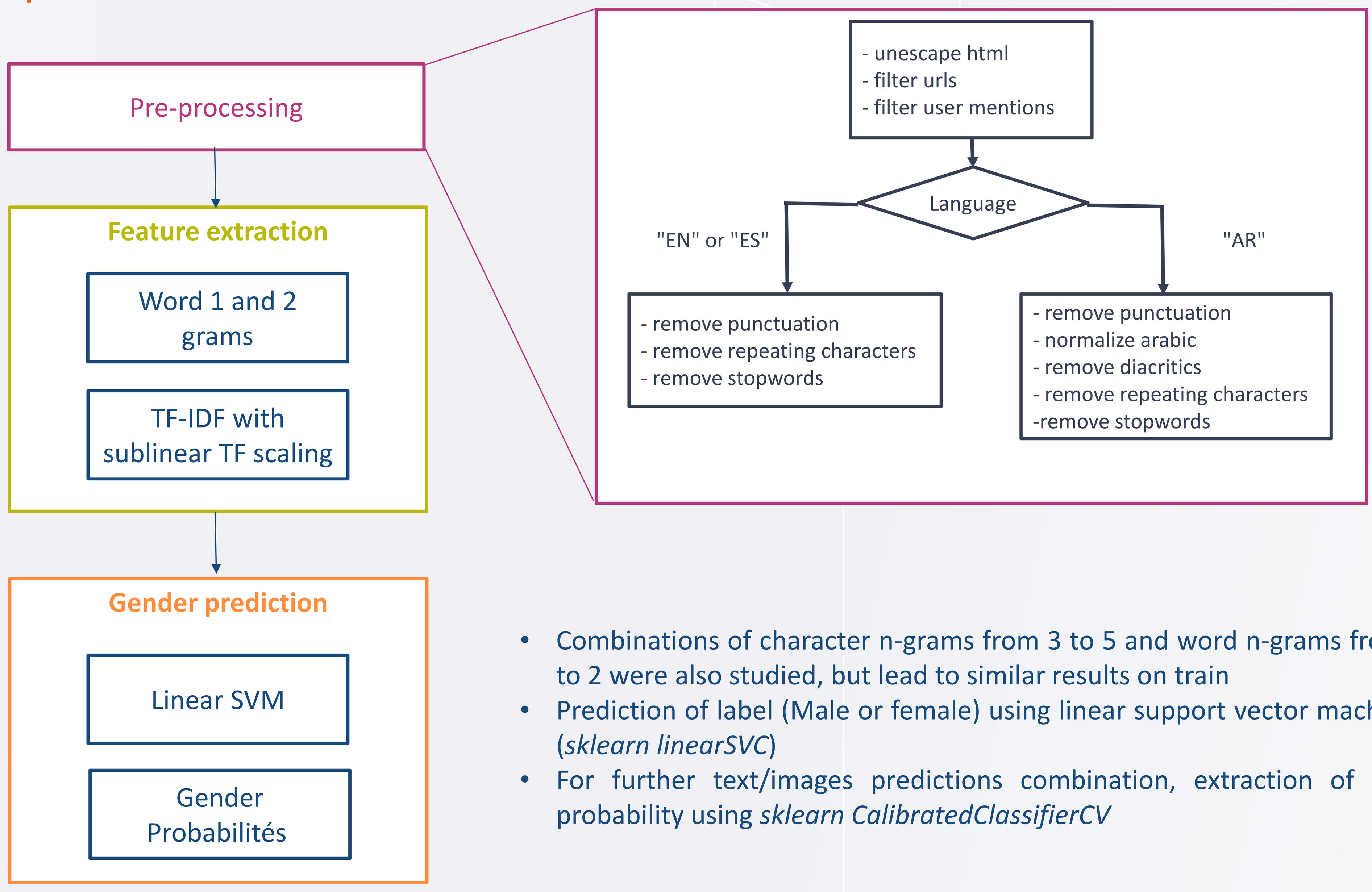
LIRIS UMR 5205 CNRS / INSA de Lyon / Université Claude Bernard Lyon 1 / Université Lumière Lyon 2 / Ecole Centrale de Lyon

Gender classification for each Tweet image



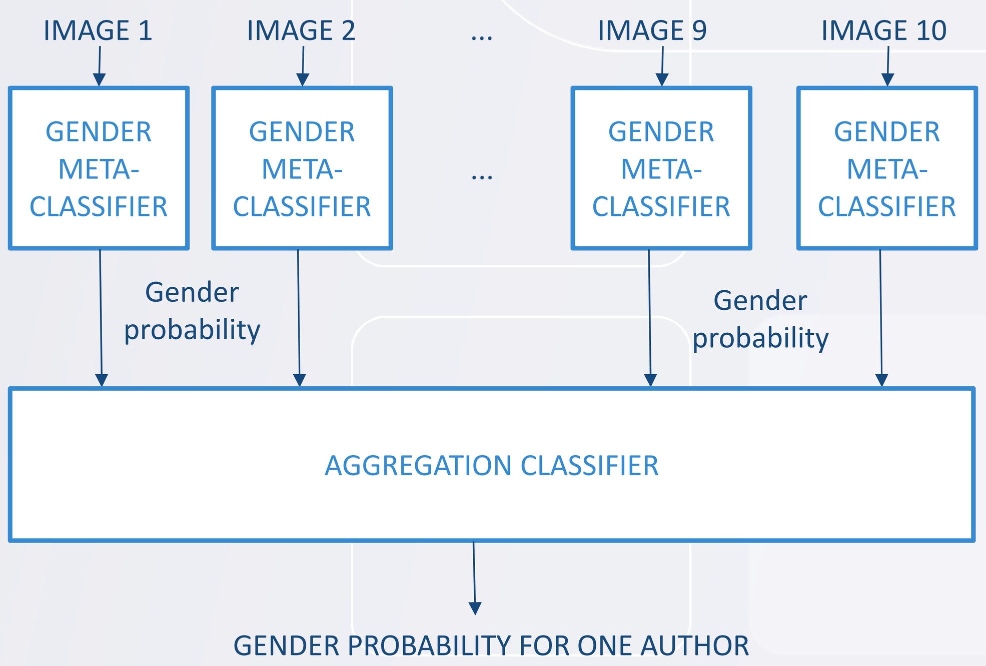
- Images in test and train sets are represented using the 4 tweet image representations.
- One classifier is trained for each type of images representations, using 56% of the training dataset, to predict a gender probability
- A meta classifier is fed with the gender probability p redicted by each classifier and a metal model is learned to predict the gender.
- For each image in the test set, the probability of an image to belong to a given gender is thus predicted.

Gender prediction from Tweet texts



- Combinations of character n-grams from 3 to 5 and word n-grams from 1 to 2 were also studied, but lead to similar results on train
- Prediction of label (Male or female) using linear support vector machines (*sklearn linearSVC*)
- For further text/images predictions combination, extraction of label probability using *sklearn CalibratedClassifierCV*

Stacked gender prediction from Tweet images



- Gender probability is predicted for each image using the four classifier and the meta classifier
- An "aggregation" classifier used the gender probabilities of the 10 images from each author to predict the gender probability of the author
- The aggregation classifier was trained on 8% of the training dataset (120 images from Arabic set, 240 images from English set, 240 images from Spanish set)

Tweet image representations

Object recognition — **Images are represented by the objects they contain, detected by an image detection algorithm, such as :**

$$V_{object} = \{O_1 : I_1, O_2 : I_2, \ldots, O_i : I_i\}$$

where O_i an object identified in the image and Ii the importance weight of the object, computed as the sum of recognition confidence scores provided by the image detection algorithm for that object. We used YOLO[1] with a confidence threshold of 0.2 as the recognition algorithm.

Face recognition — **Images are represented by a vector of two features, respectively the number of men and women detected in the image. We used a pre-trained network[2] that detects both the faces and the gender for each faces in an image.**

Color histogram — **Images are represented by a standard color histogram of size 768.**

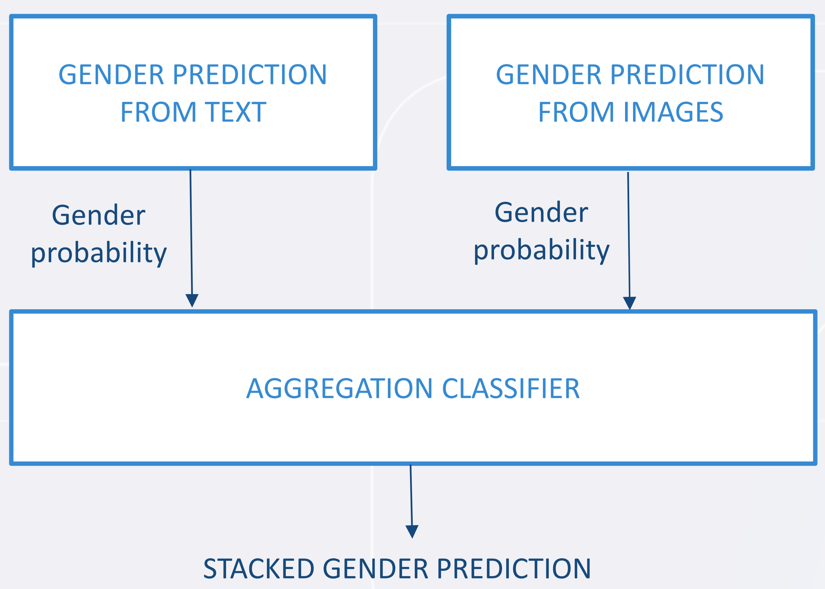
Local binary patterns — **Images are represented by a vector of local binary patterns, for 24 points and a radius of 8, of size 26. Local binary pattern is a visual descriptor widely used for classification in computer vision that allow to analyze textures.**

1. https://pjreddie.com/darknet/yolo/ (2018)
2. Won, D.: face-classification. https://github.com/wondonghyeon/face-classification (2018)

Conclusion

- Text based classification gives better results than image based classification
- The pre-processing phase (tokenizing, cleaning) is very important. We improved the standard Arabic tokenizer
- Combining the text and image based classification can be further improved

Stacked image and text classification and results



- Gender probabilities predicted from images and texts are used as inputs for a final classifier to predict the gender of the author
- Final classifier is trained on 20% of the training dataset (300 Arabic authors, 600 English authors and 600 Spanish authors)

	ACCURACY ON TEXT ONLY	ACCURACY ON IMAGES ONLY	ACCURACY ON TEXT AND IMAGES
ARABIC	0.7910	0.7010	0.7940
ENGLISH	0.8074	0.6963	0.8132
SPANISH	0.7959	0.6805	0.8000

LIRIS · CNRS · INSA · UCB Lyon 1 · UNIVERSITÉ LUMIÈRE LYON 2 · ÉCOLE CENTRALE LYON