Bot and Gender Detection of Twitter Accounts Using Distortion and LSA

Andrea Bacciu, Massimo La Morgia, Alessandro Mei, Eugenio N. Nemmi, Valerio Neri, Julinda Stefa

SAPIENZA Università di Roma

Abstract

The Author Profiling Task 2019 aims to identify the nature of Twitter accounts, detecting if the writer is a Bot or a Human being and in the last case, the gender of the account owner. Morevor, the task is proposed in two different languages: English and Spanish.
For each instance of the problem and each language, we address the problem differently. We use an ensemble architecture to solve the Bot Detection for accounts that write in English and a single SVM for those who write in Spanish. For the Gender detection we use a single SVM architecture for both the languages, but we pre-process the tweets in a different way. Our final models achieve accuracy over the 90% in the bot detection task, while for the gender detection, of 84.17% and 77.61% respectively for the English and Spanish languages.

Dataset

For each account we have 100 tweets. All the tweets are raw, hence we can determine if it is a tweet or a Retweet. The length of the tweet can vary a lot, with a minimum of only one character and a maximum of more than 900

Bot Detection Features

- **Emojis**
  The average number of emojis used in each tweet.
- **Web link**
  The average number of links shared in each tweet.
- **Hashtag**
  The average number of hashtags used.
- **Len of Tweets**
  The average length of the tweets.
- **Len of ReTweets**
  The average length of the retweets.
- **Semicolons**
  The average number of semicolons
- **Cosine Similarity Score**
  Average of cosine similarity between tweets
- **Sentiment Analysis**
  Average of compund and neutral score of text as returned by Vader
- **Text Distorsion**
  2-8-grams of the tweets after masking the characters with Text Distorsion

Original Text	Distorted Text
RT @BIBLE_Retweet: Great men are not always wise – Job 32:9	** @*****_*******: ***** *** *** *** ****** *** - *** **:*
I don't know. Just making conversation with you, Morty. What do you think, I-I-I… know everything about everything?	* ***'* ****. **** ****** ************ **** ***, *****. **** ** *** *****, *-*-*… **** ********** ***** **********?

Pre-Processing Gender

- **English**
  TweetTokenizer
  SnowballStemmer
- **Spanish**
  TweetTokenizer
  SpaCy Lemmatizer

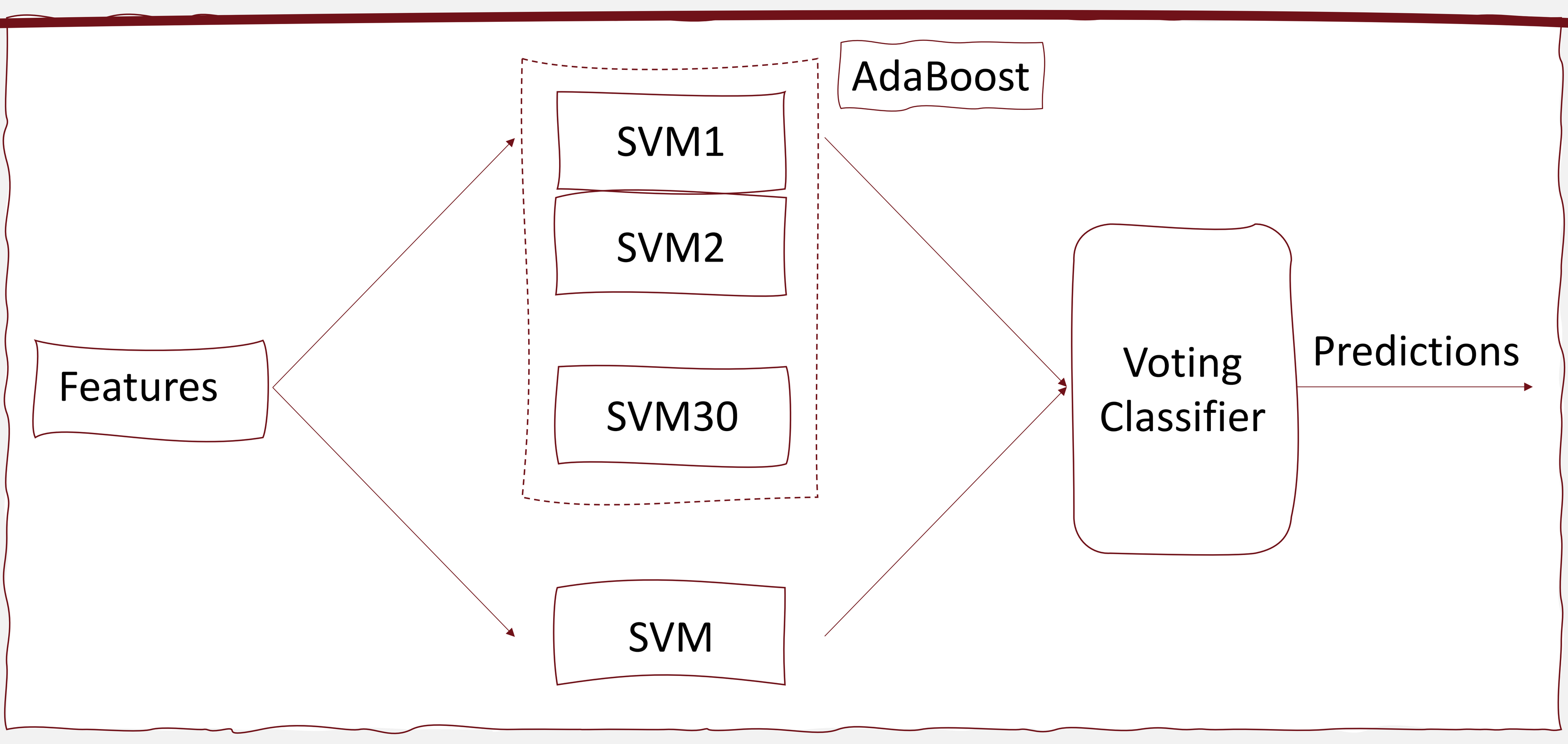
Gender Detection Features

- **English**
  Words1-5 grams
- **Spanish**
  Words1-8 grams

Classifier Description

In the first layer, we have a single SVM and an AdaBoost instance. AdaBoost is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases. In the second layer, we have a Soft-Voting Classifier that ensemble the predictions of the previous layer. We sum up for each class the probability of being the right one as predicted by our classifiers, then we pick as final prediction, the class with the highest value.

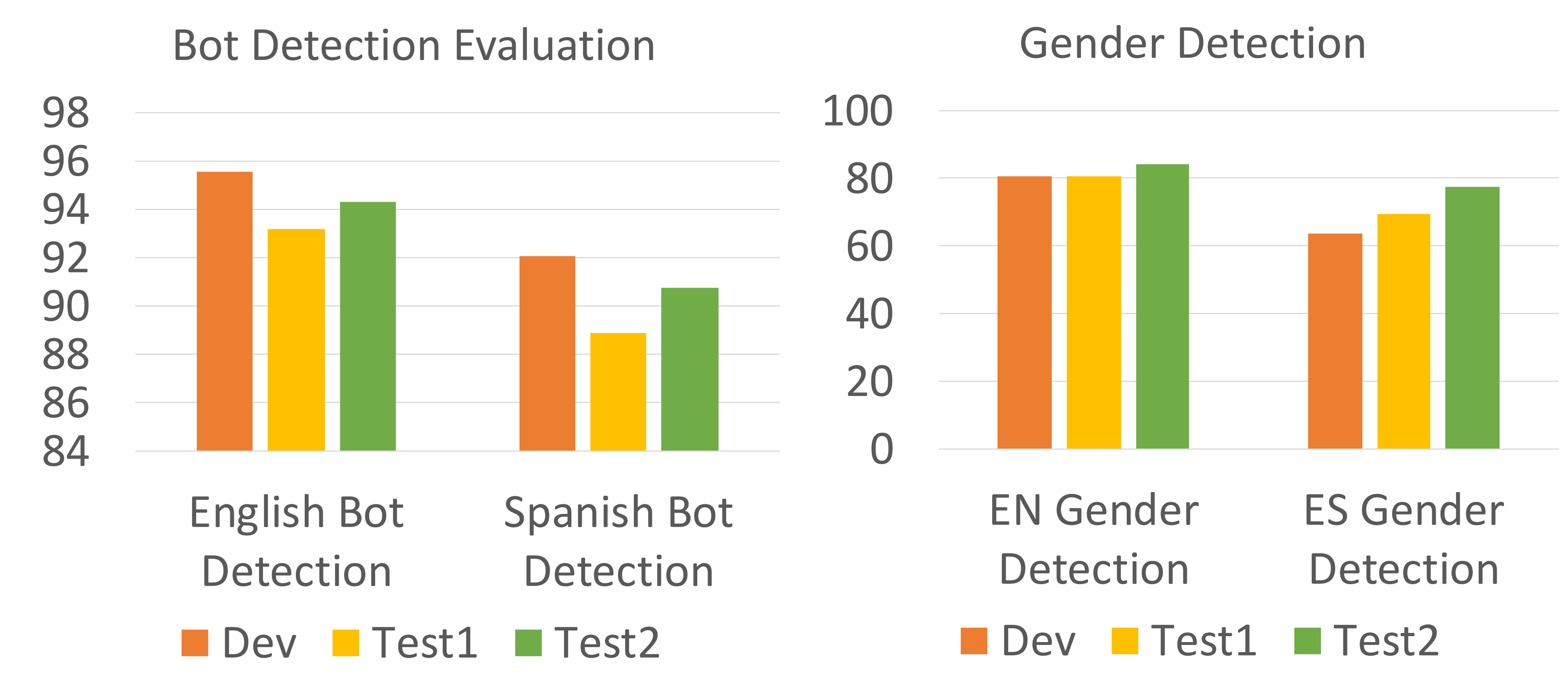
EN Bot Model Architecture

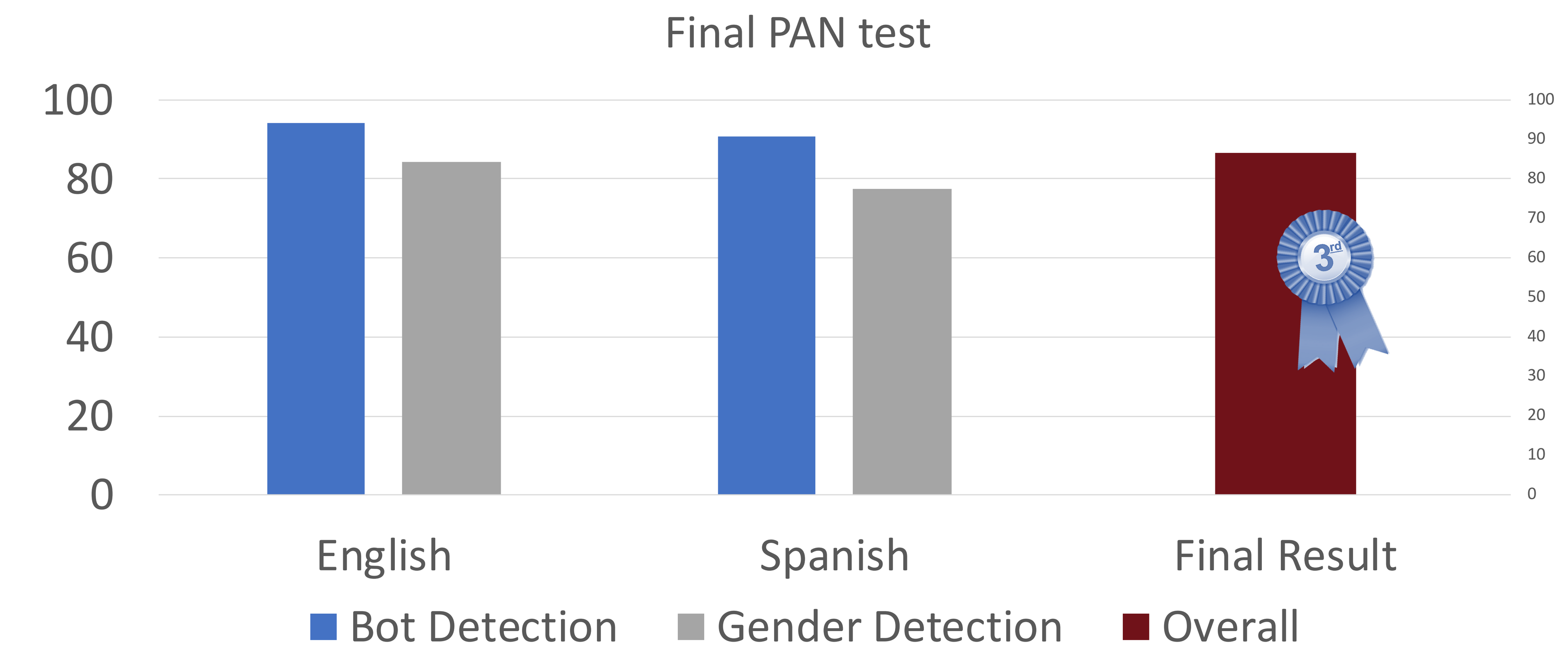


Bot Feature Results

Feature	EN score	ES score
Emojis	71.05	70.65
Web link	79.19	87.93
Hashtag	76.53	-
Len of Tweet	59.59	53.36
Len of ReTweet	86.12	81.41
Semicolons	64.27	55.21
Cosine Smilarity	60.00	64.67
SA: Neutral	61.53	66.19
SA: Compound	51.37	68.36
Text Distortion	90.48	81.73
ALL	77.90	76.73

Results





Conclusion and Future Work

We develop 4 different classifiers, one for each problem subset. For the Bot detection of English written messages, we used an ensemble architecture where the AdaBoost outputs are ensembled by a soft-voting classifier. For each one of the other problems, we use a single fine-tuned SVM. We achieve excellent performances, especially in the bot detection task, where we record a score of about 95% on the English Dev and the English final Test set. Globally our models perform better on the English accounts than the Spanish ones, so we believe that more work is needed to fill this gap.

PAN
CLEF 2019 Conference and Labs of the Evaluation Forum - Information Access Evaluation meets Multilinguality, Multimodality, and Visualization. 9 - 12 September 2019, Lugano