

NAIVE-BAYESIAN CLASSIFICATION FOR BOT DETECTION IN TWITTER

PABLO GAMALLO, CiTIUS, Univ. of Santiago de Compostela
SATTAM ALMATARNEH COLE group, UVigo



Centro Singular de Investigación
en Tecnoloxías da
Información

1. OBJECTIVES

- To participate in the Bots and Gender Profiling shared task at PAN 2019 (English and Spanish).
- To design a basic bayesian strategy to detect bots from human tweets based on **linguistic features**, including textual content of tweets and automatically built lexicons.

3. LEXICAL FEATURES

They were derived from four different weighted lexicons that were automatically built from annotated corpora for each language:

Generic human-bot lexicon: lemmas mainly used by bots and lemmas mainly used by humans.

Generic female-male lexicon: lemmas mainly used by women and lemmas mainly used by men.

Sentiment human-bot lexicon: lemmas with positive or negative polarity and classified as being used by bots or humans.

Sentiment female-male lexicon: lemmas with polarity and classified as being used by women or men.

4. TWEET SIMILARITY

Similarity between sequential pairs of tweets, t_1 and t_2 , is defined as follows:

$$Sim(t_1, t_2) = \frac{\| L_{t_1} \cap L_{t_2} \|}{\| L_{t_1} \cup L_{t_2} \|} \quad (1)$$

where L_{t_1} and L_{t_2} are the lexical lemmas (nouns, adjectives, verbs, and adverbs) of tweets t_1 and t_2 , respectively, and $t_1 \prec t_2$.

5. EXPERIMENTS (TRAINING AND DEVELOPMENT SPANISH DATASETS)

features	bot/human accuracy	male/female accuracy
<i>bow</i>	0.85	0.67
<i>lex</i>	0.65	0.57
<i>text</i>	0.71	0.50
<i>bow+text</i>	0.83	0.68
<i>lex+text</i>	0.90	0.67
<i>bow+lex+text</i>	0.80	0.63

2. TEXT-BASED FEATURES

- hashtags
- user references
- url links
- retweets
- textual emoticons
- emojis
- onomatopoeia
- language abbreviations
- alliterations
- size of tweets
- identical tweets
- token/lemma ratio
- **tweet similarity**

6. DISCUSSION

- Lexical and textual features (lex+tex) outperforms bag-of-words (bow) in the bot/human classification, but accuracy goes down when combining all features (bow+lex+text).
- The Spanish accuracy for bot/human classification outperforms two of the baselines (including embeddings) and reached the 16th position out of 44 submitted systems: **0.88 accuracy** in the Spanish test dataset.