Twitter User Profiling: Bot and Gender Identification

7th Author Profiling Task

PAN 2019 – CLEF Workshop

Dijana Kosmajac
Dr Vlado Keselj
Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia, Canada

Overview

- Introduction
  - Bot Detection on Social Media
- Methodology
  - DNA-inspired User Behaviour Fingerprint
  - Diversity Measures
- Dataset of 7^{th} Author Profiling Task
- Experiments and Results
- Conclusion

Note: for gender detection approach, please refer to the working notes

Bot Detection on Social Media

- Social media - convenient platforms for people to share, communicate, and collaborate.
- Openness of social media is great, but…

  malicious behaviors happen, such as bullying, terrorist attack planning, and fraud information dissemination, etc.
- Important task: detect these abnormal activities as accurately and early as possible to prevent disasters and attacks.
- For this study we approached to a subdomain: **bot detection**

Bot and Gender Detection on Social Media

- DeBot: Twitter Bot Detection via Warped Correlation, Chavoshi et al., 2016
- DNA-Inspired Online Behavioral Modeling and Its Application to Spambot Detection, Cresci et al., 2016

DNA-inspired User Behaviour Fingerprint

- Introduced first time in Cresci et al., 2016

$$A_n = \begin{cases} 0, & \text{plain} \\ 8, & \text{retweet} \\ 16, & \text{reply} \\ 1, & \text{has hastags} \\ 2, & \text{has mentions} \\ 4, & \text{has URLs} \end{cases}$$

User timeline

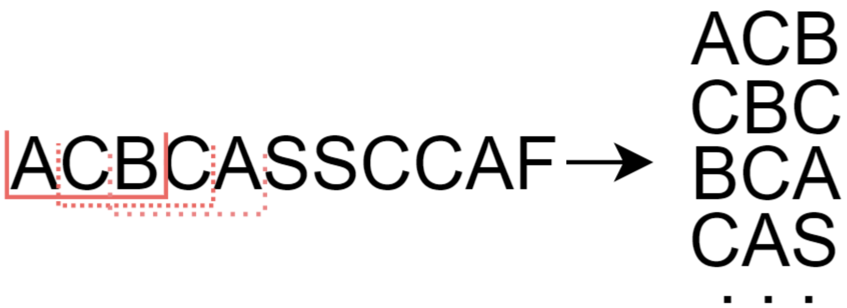
ASCII(65+*code*)

3∗2^3= 24 different labels

ACBCADDCCAF…

DNA-inspired User Behaviour Fingerprint

- We used 1-, 2-, 3- and 4-grams
  - 3-gram example:



Diversity Measures

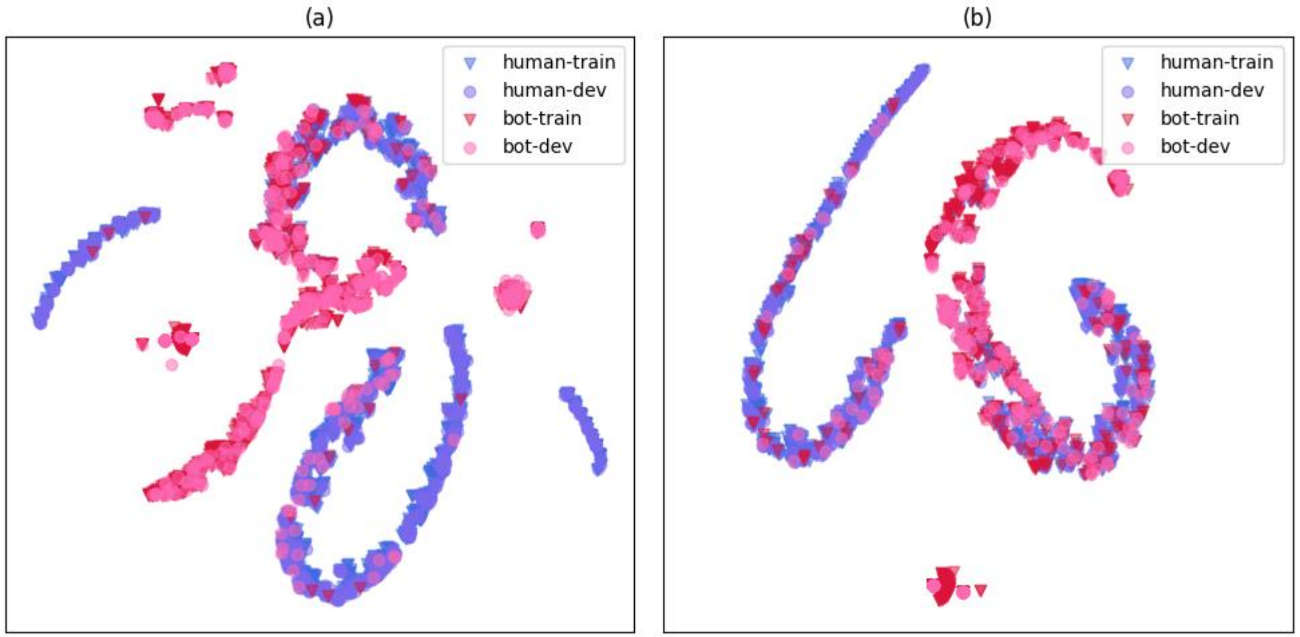
- Yule's $K = C\left[-\frac{1}{N} + \sum_{m=1}^{m_{max}} V(m,N)\left(\frac{m}{N}\right)^2\right]$
- Shannon's $H = -\sum_{i=1}^{V(N)} p_i \ln(p_i)$
- Simpson's $D = \frac{1}{\sum_{i=1}^{V(N)} p_i^2}$
- Honore's $R = 100\,\frac{\log(N)}{1-\frac{V(1,N)}{V(N)}}$
- Sichel's $S = \frac{V(2,N)}{N}$

Dataset

- Bot t-SNE visualization. (a) English, (b) Spanish

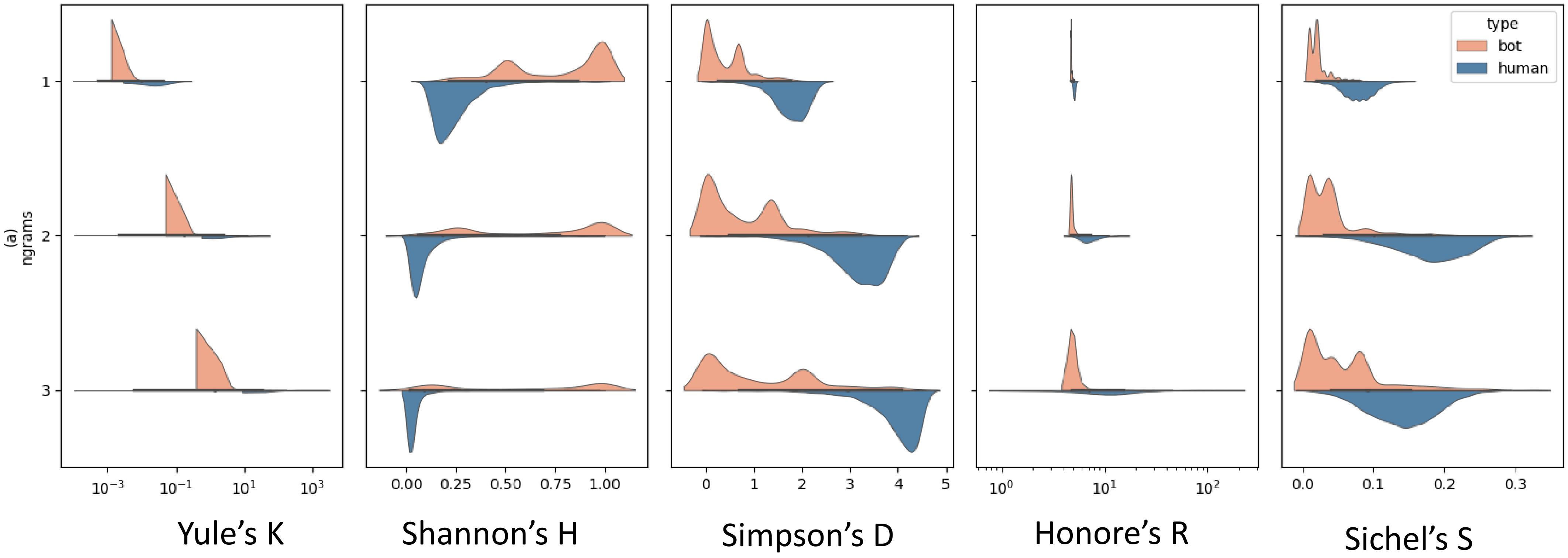
- English:
  - 2,880 train and 1,240 dev

- Spanish:
  - 2,080 train and 920 dev



Dataset

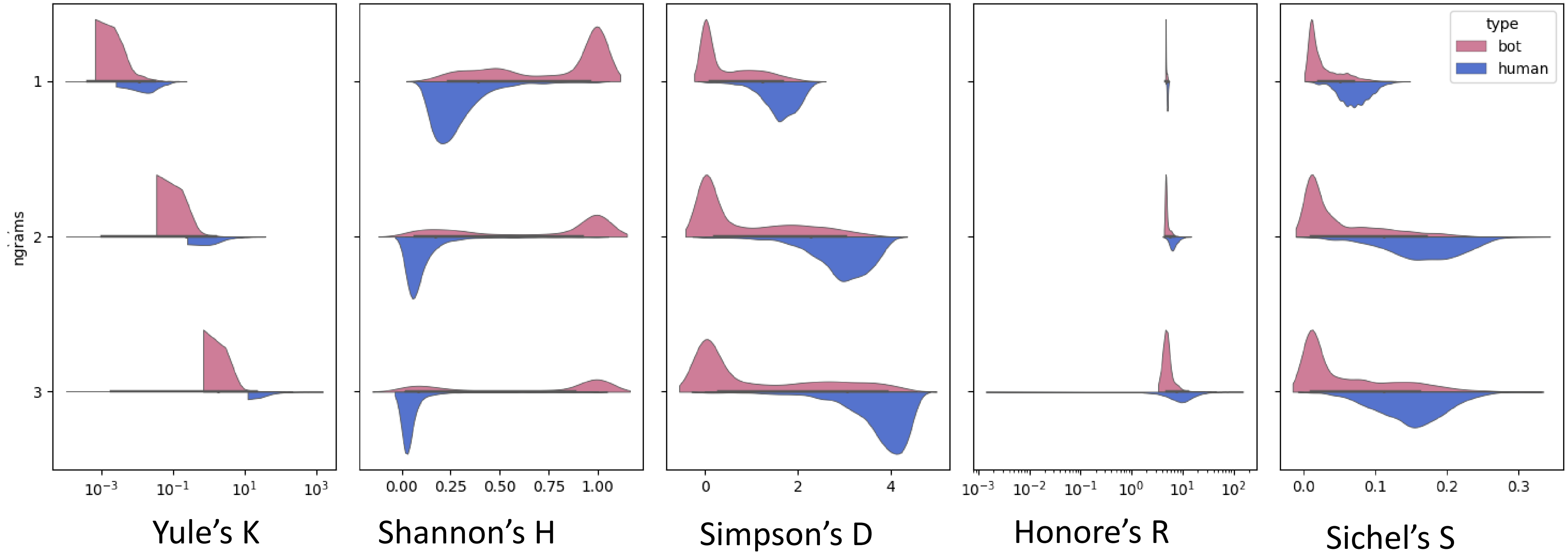
- Diversity measures visualization for English



Introduction | Methodology | Dataset | Experiments | Conclusion

Dataset

- Diversity measures visualization for Spanish



Experiments with language-specific training

- Experiment 1: character n-grams range 2-4, w/o diversity measures.
- Experiment 2: character n-grams 1-3, w/ diversity measures

		E1			E2		
Dataset	Classifier	Precision	Recall	F1	Precision	Recall	F1
English	GB	0.9197	0.9153	0.9151	0.9263	0.9234	0.9233
	SVM	0.9174	0.9161	0.9161	0.9253	0.9242	0.9241
	LR	0.8840	0.8750	0.8743	0.9261	0.9242	0.9241
	KNN	—*	—*	—*	0.9284	0.9258	0.9257
	RF	0.9284	0.9218	**0.9215**	0.9293	0.9266	**0.9265**
Spanish	GB	0.8666	0.8663	0.8663	0.8429	0.8391	0.8387
	SVM	0.8602	0.8598	0.8597	0.8164	0.8163	0.8163
	LR	0.8663	0.8663	0.8663	0.8510	0.8478	0.8475
	KNN	—*	—*	—*	0.8617	0.8587	0.8584
	RF	0.9115	0.9033	**0.9028**	0.8503	0.8489	**0.8488**

Table 1. Bot classification. Results tested on development dataset. Per language training dataset.
∗ not available due to memory restrictions.

Experiments with combined training

- Experiment 3: same as E1, only combined training set
- Experiment 4: same as E2, only combined training set

		E3			E4		
Dataset	Classifier	Precision	Recall	F1	Precision	Recall	F1
English	GB†	0.9252	0.9242	**0.9241**	0.9330	0.9306	**0.9305**
	SVM	0.9094	0.9081	0.9080	0.9199	0.9177	0.9176
	LR	0.9121	0.9113	0.9112	0.9214	0.9202	0.9201
	KNN	—*	—*	—*	0.9256	0.9242	0.9241
	RF	0.9189	0.9153	0.9151	0.9256	0.9242	0.9241
Spanish	GB†	0.8896	0.8880	**0.8879**	0.8512	0.8424	0.8414
	SVM	0.8588	0.8587	0.8587	0.8490	0.8435	0.8429
	LR	0.8478	0.8478	0.8478	0.8473	0.8446	0.8443
	KNN	—*	—*	—*	0.8586	0.8543	**0.8539**
	RF	0.8764	0.8696	0.8690	0.8498	0.8435	0.8428

Table 2. Bot classification. Results tested on development dataset. Combined training dataset. † used as final classifier (E4 for official ranking). ∗ not available due to memory restrictions.

Official results

- 13th place in total, better than all baselines.

Dataset	Bot	Gender
English	0.9216	0.7928
Spanish	0.8956	0.7494
Average	0.9086	0.7711

Table 4. Final results on test dataset. Averaged per language.

Conclusion and Future Work

- A novel, yet simple method for bot detection on social media.
- Language independent, since it does not use the language-specific features.
- Disadvantage – doesn't consider language-specific features which may be more fine-grained.

- Explore the effect of the length of the user fingerprint on ability to differentiate bot and genuine users.
- Explore the effect of the timespan the fingerprint is collected.
- Explore the effect of using variable length fingerprint.
- Explore possibility of unsupervised bot detection using diversity measures and clustering.