Toward Bot Detection in Social Media

Tahora H. Nazer, Fred Morstatter, Liang Wu, Huan liu
Arizona State University
{tahora.nazer, fred.morstatter, wuliang, huanliu}@asu.edu

Motivation

- Bots are automated accounts on social media.
- Bots have become a major nuisance in recent years.
- They can manipulate discussions and manipulate the statistics of the crowd.

Figure 1: Bots in Twitter

Bots are hard to detect as they try to look like normal users.

Question

- How to obtain bots to study?
- How to detect bots in social media?
- How to evaluate the detection method?

Ground Truth

- How to find a set of bots to study?
- Manual annotation is expensive.
- Suspension method is not 100% accurate.

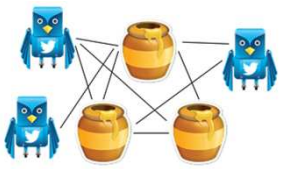

Figure 2: Honeypots in social media

Honeypots

- Honeypots: bots created to lure other bots in the wild.
- Any user that follow them is a bot.
- The assumption is that normal users can easily tell that honeypots are bots.

Honeypot Dataset Statistics	
Number of Honeypots	9
Detected bots	3K
Duration of Crawl	3 Months

Heuristic Based Method

- Decision is made based on one feature
- Features are obtained using observations on recognized bots
- Examples:
  - Number of friends
  - Device used to send posts
  - Timing pattern of posts

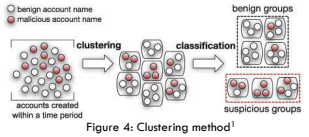

Figure 4: Clustering method[1]

[1] Lee, Sangho, and Jong Kim. "Early filtering of ephemeral malicious accounts on Twitter." Computer Communications 54 (2014): 48-57.

Classification Method

- A classifier is built
- Features are either heuristic or automatically extracted e.g. n-grams

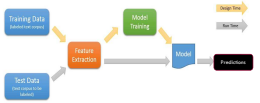

Figure 3: Classification method

Clustering Method

- Cluster users based on a single feature
- Classify clusters based on engineered features

A Novel Method

Challenge

- How to increase the recall without sacrificing the precision?
- How to find more bots?

Idea

- Increase the weight of misclassified bots to have more impact on classification results.

$$\mathbf{D}_{t+1}(i) = \frac{\mathbf{D}_t(i)exp(-\alpha_t y_i h_t(\mathbf{v}_i))}{Z_t}$$

$$Z_t = \sum_{i=1}^{m} \mathbf{D}_t(i)exp(-\alpha_t y_i h_t(\mathbf{v}_i))$$

Method

- Extract LDA topics probability distribution.
- Apply AdaBoost.
- Increase the weight of misclassified bots.
- Reapply the model.

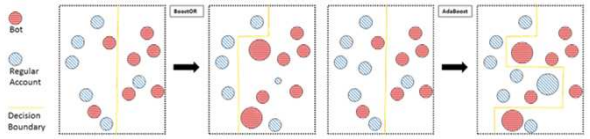

Figure 5. BoostOR vs. AdaBoost

Key Results

Experimental results of BoostOR with varying number of latent dimensions.

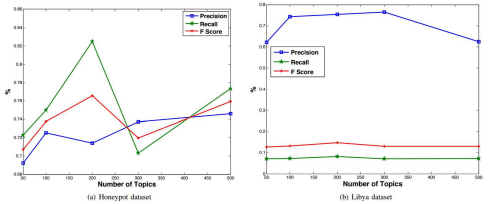

Figure 6. Experimental Results – effect on recall and precision

Table 1. Experimental Results – Comparison with different methods

Method	Precision	Recall	F_1
$Heuristic_{URL}$	49.69%	96.39%	65.58%
$Heuristic_{Retweet\%}$	50.05%	99.33%	66.56%
$Heuristic_{Length}$	50.00%	99.82%	66.63%
$Heuristic_{Time}$	49.99%	99.96%	66.65%
SVM	62.41%	62.52%	62.47%
$AdaBoost$	79.76%	72.41%	75.91%
$BoostOR$	71.42%	82.48%	**76.55%**

Conclusions

- Bots can perform malicious activities in social media.
- The current focus is on achieving high precision which leaves a large number of bots undetected.
- With help of AdaBoost we have achieved high recall without sacrificing precision.

Future Work

- Discovering more features that can help with the bot detection process.
- We will continue to refine the process of acquiring ground truth.
- Application of more classification methods.