Bot Research

Bots are a prominent aspect of nearly all major online social networks and communities. According to a US Securities and Exchange Commission filing by Twitter, about 8.5% of all Twitter users are bots. While some bots may be helpful or informative to certain online social communities, many bots have more malicious purposes, which have become widely known since the 2016 US Presidential election. Malicious bots have become infamous for their impact in influencing public perception, spreading misinformation, and posing a significant risk to freedom of expression and to democracy in general. It is therefore crucial that research be done on effectively and accurately identifying bots.

SentiBot placed first in the DARPA Twitter Bot Challenge against 5 other teams, achieving high accuracy and speed.

Bots in Elections

Researchers now at Northwestern's Security & AI Lab were the first to ever study the use of bots in elections. In a landmark 2014 paper, these researchers developed a dataset consisting of 10 months of Twitter data about the 2014 Indian election, including information on 31 major national candidates and political parties. This was the first time ever that the use of bots in elections had been studied – well before the infamous bot campaigns surrounding the 2016 US Presidential election.

The researchers then developed SentiBot which uses a sentiment-aware architecture to identify influence bots on Twitter. Specifically, SentiBot used a set of contextual variables for each user that is split into four categories, most of which include the user's sentiment.

1. Tweet syntax – average number of hashtags, mentions, links, and special characters
2. Tweet semantics – average topic sentiment, positive sentiment strength, agreement rank, dissonance rank, etc.
3. User behavior – tweet spread, frequency, variance, repeats, etc.
4. Network-centric user properties – in-degree, out-degree, neighborhood contradiction, agreement, and dissonance ranks.

In addition to these generic features, SentiBot proposed a host of novel features that merge sentiment related features with the structure of the follower-followee network on Twitter.

SentiBot then built an ensemble of classifiers using standard machine learning techniques. In this project, SentiBot used a dataset of 7.7 million Indian political tweets by over 500,000 users. Evaluation found that the probability that the classifier will rank a randomly chosen bot as more "bot-like" than a randomly chosen human is 0.73 for the full feature set. Of the 25 most important features in the best classifier, 19 were sentiment related and 14 are topic specific. The five topic independent features were:

1. Sentiment flip-flop
2. Positive sentiment strength
3. Negative sentiment strength
4. Fraction of tweets with sentiment
5. Dissonance Rank

From these results, it is clear that SentiBot is a highly informative and accurate system for detecting bots on Twitter.

In a related effort, the same dataset and a related dataset were used to correctly predict the outcome of the 2014 Indian election as well as the outcome of the 2013 Pakistani election.

The DARPA Twitter Bot Challenge

DARPA, the Defense Advanced Research Projects Agency, is the US Department of Defense's premier research organization whose accomplishments include the invention of the Internet and the creation of real world unmanned vehicles. DARPA was keenly aware of the threat posed by social media well before the 2016 US Presidential election. In 2011, they announced the Social Media in Strategic Communications or SMISC program. A major goal of this $42M program was to develop the techniques required to automatically identify persuasion campaigns and influence operations that leveraged social media.

More than 3 years into the program DARPA announced the Twitter Bot Detection Challenge under the SMISC program. Participants included giant corporations such as IBM, well-funded universities such as Georgia Tech and USC, consortia of universities such as a University of Indiana and University of Michigan team, as well as tiny companies such as Sentimetrix.

The Challenge involved identifying bots that were trying to influence opinion on Twitter to be pro vaccination. Run over 28 day period, the challenge involved over 7,000 accounts. Participating teams could make a guess

at any time. If their guess was correct, a point was added to their score. If their guess was incorrect, ¼ of a point was deducted from their score. In a real world bot detection scenario, time is of the essence as early detection of a bot campaign helps limit its adverse effects. In order to capture this, the DARPA SMISC Twitter Bot Challenge also offered a time bonus: if a team guessed all the bots by day *d*, then one point was added to the team for each day remaining in the challenge, i.e. (28-d) points were added as a speed bonus.

The Challenge captured many aspects of real-world bot detection. First, though lots of tweets and accounts and follower-followee relationships were provided by DARPA, none of the accounts were labeled either as bots or as benign accounts. This made the use of supervised machine learning algorithms very challenging. Second, no information was provided about whether one group was generating the bots, or whether there were multiple such groups.

Working as the leader of the Sentimetrix, Inc. team, Northwestern Professor V.S. Subrahmanian led a team that handily won the competition. His team adapted SentiBot to account for the behaviors and characteristics of the Challenge. The team developed a methodology to identify bot campaigns in future scenarios that would be dynamically adaptable to the situation – in particular, they proposed a 3-phased human-in-the-loop approach in which a first phase looked at both unsupervised clustering of accounts and an assignment of a bot probability to each cluster of accounts based on prior models of bots from other campaigns (in particular the 2014 India election) , a second phase where the findings of the first phase were incorporated into a supervised machine learning model so that features associated with the specific bot campaign being monitored were considered before a combination of supervised ML and anomaly detection were used, and a third phase were more features were added and supervised classification was used directly.

Prof. Subrahmanian's team achieved the highest score of 50.75, having near perfect accuracy in finding the 39 bots (in 40 guesses) and a much higher speed than other teams, finishing 12 days before the end of the competition and a full 6 days before the second ranked competitor.

Overall, SentiBot provides an applicable and accurate framework for applications to identify bots.

Additional Information

References

1. John Dickerson, Vadim Kagan and V. S. Subrahmanian, "Using sentiment to detect bots on Twitter: Are humans more opinionated than bots?." *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*.
2. V. S. Subrahmanian et al., "The DARPA Twitter Bot Challenge," in *Computer*, vol. 49, no. 6, pp. 38-46, June 2016.
3. Kagan, Vadim, Andrew Stevens, and V. S. Subrahmanian. "Using twitter sentiment to forecast the 2013 pakistani election and the 2014 indian election." *IEEE Intelligent Systems* 30, no. 1 (2015): 2-5.

PARTICIPANTS

Lead: V.S. Subrahmanian

John Dickerson, Vadim Kagan, Andrew Stevens.

Northwestern | BUFFETT INSTITUTE FOR GLOBAL AFFAIRS

Northwestern | McCORMICK SCHOOL OF ENGINEERING