Detecting and Characterizing Bot-Like Behavior on Twitter

SiHua Qi(✉), Lulwah AlKulaib, and David A. Broniatowski

George Washington University, 2121 I St NW, Washington, DC 20052, USA
{qisihu,lalkulaib,broniatowski}@gwu.edu

Abstract. Social media is becoming a platform of choice for people to voice their opinion on topics of discussion. To evaluate these opinions, it is important to have an accurate assessment of who is saying what. Unfortunately, social media are also the home of bots which makes the assessment difficult. Bots are computer programs designed to mimic human behavior online in social networks. They are used to pursue a variety of goals, including, but not limited to, spreading information, and influencing targets.

In this paper, we describe a machine learning framework that uses content-based features extracted from Twitter to detect bot-like behavior on the platform. Unlike other machine-learning approaches to bot detection, we seek to generate explanations of why specific accounts are categorized as bots; thus allow us to modify these criteria as bots' behaviors evolve. We have therefore developed the criteria mentioned in an article published in Medium [1] to detect bot-like behavior in our dataset then evaluate the results. We then explain the different types of bots that used as our datasets and compare the significant features for each type of bots in a logistic regression method.

Keywords: Bots · Bot-like behavior · Logistic regression

1 Introduction

Online social network sites are becoming more popular each day. According to a report by Dream Grow, Twitter is considered among the top 15 most popular social networking sites. It has 330 M active users monthly, which puts it in 4th place on that list. [2] Unfortunately, the number of bot accounts is surprisingly large. A new paper from University of Southern California and Indiana University suggests that up to 15% of twitter accounts are in fact bots rather than people [3].

Bots on Twitter are accounts controlled by a software, automatically producing content, and interacting with other users. Some of these bots use Twitter as a tool to announce news headlines, others utilize the platform for marketing, such bots are considered useful bots. However, there is a growing record of misuse of bot accounts. These accounts would be designed to mimic human behavior, then sold to users aiming to boost their popularity with fake followers, [4] used to promote terrorist propaganda, [5] or used by some organizations to influence public opinion [6].

R. Thomson et al. (Eds.): SBP-BRiMS 2018, LNCS 10899, pp. 228–232, 2018.
https://doi.org/10.1007/978-3-319-93372-6_26

Twitter allows bots on the platform that adhere to their rules, automated likes are not allowed, and automated retweets are only allowed for entertainment, informational, or novelty purposes. [twitter-automation] These rules, and other issues have been addressed after the DARPA challenge which was a Twitter bot detection challenge to study malicious activities carried by bot accounts. [7] While working on this project we noticed that published papers talk about bot detection. Meanwhile, users that try to hide that their accounts constantly violate Twitter rules, are not always fully automated. Bots can be turned on and off as needed. When the program is not running the account, a human would be posting, which makes these accounts harder to detect.

In this paper, we defined our criteria that determines what a bot account acts like as bot-like behavior. We used criteria from an article published by Nimmo [1], and some others that we added as the study evolved. Unlike other approaches that try to predict whether an account is a bot or not based on holdout data [3], we use a statistical approach that aims to provide explanatory insight into why our assignment is made. The rest of this paper is organized as followed. In Sect. 2, related work is discussed. In Sect. 3, We propose criteria used for bot-like-behavior detection. We use the criteria to train our model which detects accounts that satisfy any of the criteria and generates a report for each dataset with the results. In Sect. 4, we explain the results generated by the study.

2 Related Work

Twitter has been widely used since 2006, and the open structure of twitter lead people to question who is tweeting early on. Chu et al. [8], classified twitter users into human, bot, and cyborg accounts using 4 components each of which checks a specific criterion, then compute a score that enables classification. Davis et al. [9], created a service that evaluates the extent to which a Twitter account exhibits the similarity to the known characteristics of social bots. Their platform fetches a given account's recent activity, then computes and returns a bot-likelihood score. The DARPA Twitter bot challenge [7] also addressed four different features they assigned to different teams to work on in the bot detection challenge. The detection systems created in the challenge were all semi-supervised and all teams used human judgement to augment automated bot identification processes.

3 Bot-Like Behavior Detection

3.1 Data Collection

Twitter has a set of API functions [10] that supports user information collection. Our data was collected using the Twitter API, where we crawled the most recent 200 posts by users from a known bot list [11]. We used a dataset consisting of 4 types of manually verified twitter bots: Fake Followers, Traditional Spam Bots, Social Spam Bots and Content Polluters. We also pulled a list of verified legitimate users from the same source.

3.2 Methods

We designed a study to describe a list of user behaviors for each twitter account. Using the article by Nimmo [1], we created a program that would detect the features that indicate bot-like behavior. After data collection, we ran the script that generated results for each user against our criteria. Using those results, we applied a stepwise logistic regression model based on Akaike Information Criteria (AIC) values to determine which of the 19 features were relevant when detecting bot-like behavior. Features are explained in Table 1.

Table 1. Features used for bot-like-behavior detection.

Feature name	Explanation
digit_screen_name	screen_name consists of digits only
scramble_name	screen_name consists of alpha numeric scrambles
default_profile_image	using default profile image
default_background_image	using default background image
url_shortner	using url shorteners in tweet content
low_post_high_result	retweet count or like count is more than number of followers for given account
multi_language	more than 2 languages appeared in tweets crawled
tweet_frequency	average daily tweet number
time_range	average days between two consecutive tweets
rt_number	number of retweets/total tweets crawled
#of_mentions	average number of mentions in original tweets crawled for this account
#of_hyperlinks	average number of hyperlinks in original tweets crawled for this account
#of_friends	number of friends
#of_followers	number of followers
status_num	number of tweets
#of_favorites	number of favorited tweets
most_recent_time	most recent tweet timestamp
tweet_avg_word_number	average number of words in each original tweet
tweet_lexical_diversity	number of unique words used in all crawled original tweets

4 Analysis

We tested our script on all four bot categories data that we collected. Examining the results, we used the cut off value $|z| = 2$ as a threshold to extract features which are more relevant to the model. The $|z| = 2$ cut off value corresponds to two-sided hypothesis test with a significance level of = 0.05. A big magnitude of z-score indicates that the corresponding true regression coefficient is not 0 and that the variable matters. Based on the features we got, we described a series of bot-like behaviors.

For all four types of bots, there were 2 features in common: “most_recent_time” with a negative z-score and “status_num” with a positive z value which indicated if a user is more active recently, the user is most-likely not a bot and the user with a high number of tweets are more likely be a bot.

Based on the reports generated (Table 2), we noticed that each bot type has features more significant to that type. Fake follower bots are “simple accounts that inflate the number of followers of another account” [9]. Our results showed that fake follower accounts do not tweet frequently, but they have a significant number of friends consistent with their purpose: making other accounts popular. On the other hand, content polluters bots are designed to generate spam while masquerading as humans [12]. According to our analysis, content polluters have a high average number of tweets per day, and a significant number of friends. This corresponds to the idea of spam accounts in general, where accounts are trying to increase their outreach. In contrast, traditional spam bots are “a group of automated accounts spamming job offers” [9], which are easily identifiable as automated. In our dataset, the average time between two posts by traditional spam bots is short. Traditional spambots also rarely post retweeted content. Such behavior is consistent with accounts that are designed to posting job advertisements. Finally, social spam bots are “spammers of products on sale at Amazon.com” or “spammers of paid apps for mobile devices” [9]. The report shows that these accounts post several of hyperlinks, and do not engage in conversations (twitter mentions). Social spam bot behavior that we found here is consistent with their content suggested by the source.

Table 2. Features relevant to bo types.

Feature name	Fake followers		Content polluters		Traditional spam		Social spam	
	z	P > |z|	z	P > |z|	z	P > |z|	z	P > |z|
status_num	4.64	0	13.555	0	3.367	0.001	2.71	0.007
tweet_frequency	4.541	0	13.598	0	−4.378	0	−4.986	0
#of_friends	3.409	0.001	5.242	0	4.239	0	–	–
avg_word_number	−2.239	0.025	–	–	–	–	–	–
multi_language	−2.732	0.006	–	–	–	–	–	–
most_recent_time	−3.177	0.001	−5.005	0	−5.316	0	−8.441	0
scramble_name	–	–	−2.994	0.003	4.03	0	−3.55	0
rt_number	–	–	−4.51	0	−3.74	0	–	–
#of_favorites	–	–	3.648	0	–	–	–	–
url_shortner	–	–	–	–	−2.379	0.017	–	–
avg_time_btw_status	–	–	–	–	−3.263	0.001	–	–
#of_hyperlinks	–	–	–	–	−4.098	0	2.93	0.003
#of_followers	–	–	–	–	–	–	4.383	0
default_background_ image	–	–	–	–	–	–	−2.147	0.032
low_post_high_result	–	–	–	–	–	–	−2.579	0.01
#of_mentions	–	–	–	–	–	–	−2.797	0.005

5 Conclusion

Our results demonstrate that bot-like behavior differs significantly with bot design. Specifically, one may be able to infer the functional purpose for which the bot was created by exploring the specific features along which that particular bot type differs from human users. Future work in the area of bot detection could benefit from combining explanatory approaches grounded in traditional statistical analyses, in addition to the machine-learning approaches that are already in widespread usage.

References

1. Nimmo, B.: #BotSpot: Twelve ways to spot a bot, 28 August 2017. https://medium.com/dfrlab/botspot-twelve-ways-to-spot-a-bot-aedc7d9c110c
2. Kallas, P.: Top 15 most popular social networking sites and apps, February 2018. https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/
3. Varol, O., Ferrara, E., Davis, C.A., Menczer, F., Flammini, A.: Online human-bot interactions: detection, estimation, and characterization (2017). http://arxiv.org/abs/1703.03107
4. Confessore, N., Dance, G., Harris, R., Hansen, M.: The follower factory. New York Times, 27 January 2018. https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html
5. The ISIS Twitter census: defining and describing the population of ISIS supporters on Twitter. States News Service, 13 March 2015
6. Ferrara, E., Wang, W., Varol, O., Flammini, A., Galstyan, A.: Predicting online extremism, content adopters, and interaction reciprocity (2016). https://doi.org/10.1007/978-3-319-47874-6_3. http://arxiv.org/abs/1605.00659
7. Subrahmanian, V.S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Zhu, L., Ferrara, E., Flammini, A., Menczer, F.: The DARPA Twitter bot challenge. Computer, **49**(6), 38–46 (2016). https://doi.org/10.1109/mc.2016.183. http://ieeexplore.ieee.org/document/7490315
8. Chu, Z., Gianvecchio, S., Wang, H., Jajodia, S.: Who is tweeting on Twitter. Paper presented at the 21–30, 6 December 2010. https://doi.org/10.1145/1920261.1920265. http://dl.acm.org/citation.cfm?id=1920265
9. Davis, C.A., Varol, O., Ferrara, E., Flammini, A., Menczer, F.: BotOrNot: a system to evaluate social bots (2016). https://doi.org/10.1145/2872518.2889302. http://arxiv.org/abs/1602.00975
10. Rules and policies. https://help.twitter.com/en/rules-and-policies/twitter-automation
11. Bot repository (2017). https://botometer.iuni.iu.edu/bot-repository/datasets.html
12. Lee, K., Eoff, B., Caverlee, J.: Seven months with the devils: a long-term study of content polluters on Twitter