Twitter Bot Behavior: How Twitter Bots Interact With People

Alic Szecsei
University of Iowa
alic-szecsei@uiowa.edu

Willem DeJong
University of Iowa
willem-dejong@uiowa.edu

ABSTRACT

Twitter bots are often cited as affecting the political process by manipulating the trending topics data; similar behavior is also cited on other social platforms, such as Facebook. We present our use of unsupervised machine learning, combined with Indiana University's *BotOrNot* service, to classify Twitter users as bots based on statistical analysis of their accounts, and then examine the ways in which they interact with other users. Determining how these bots interact with human users can help to focus bot-detection algorithms to target those bots that interact with human users in malicious ways.

1. INTRODUCTION

1.1 Background & Motivation

Social bots, also known as *sybil accounts*, are programs that automate interaction on social platforms. While some may simply be humorous or helpful accounts that don't attempt to hide their status as bots, others have more manipulative goals; they may flood a social network with spam, or attempt to more subtly influence the thoughts and behavior of the humans it interacts with. While social networks are extremely effective at causing social change and improving the quality of life of their users, they are also at risk of automated manipulation by bots.

Aral and Walker (2011) showed that social networks are highly effective at manipulating the public[1], and the automation of such behavior only increases this efficiency. In addition, Ratkiewicz (2011) showed that political bots actively manipulated the 2010 U.S. midterm elections[6].

1.2 Problem Statement

While there have been multiple approaches to bot detection[7][8][3], these have been restrained to simple detection. Very few have attempted to examine the ways that these fake accounts interact with real users. Our goal is to find a number of bot accounts and determine how they use social media to affect their target users. Determining which bots are attempting to manipulate social networks and which are providing services to human users is an import aspect of bot detection, and one we believe can be improved by examining how malicious bots interact with human users.

DOI: 10.1145/1235

1.3 Proposed Approach

In this paper, we use data from a bot-detection service run by Indiana University to determine whether or not users are bots. We then pull their latest tweets, as well as user data, and use the collected data in an unsupervised machine learning algorithm to cluster the users into 50 groups. We then take the data for each cluster and analyze common behavioral patterns.

1.4 Key Results

We found a general inverse trend between the *BotOrNot* score for a cluster and the number of retweets made by the cluster. In addition, a similar inverse trend exists for the number of links tweeted by users, and the number of mentions made by users.

Lorem ipsum lorem ipsum.

2. RELATED WORK

Davis (2016)[4] and Dickerson (2014)[5]. [TODO: Talk about cited papers, what their results were, how those results were relevant to our data] Lorem ipsum lorem ipsum.

Lorem ipsum lorem ipsum

Lorem ipsum lorem ipsum.

Lorem ipsum lorem ipsum.

Lorem ipsum lorem ipsum.

3. PROPOSED APPROACH

To retrieve a list of human and bot Twitter accounts, we compiled a list of 113 users with an approximately even split between humans and bots. We then retrieved data for their followers, and their followers' followers, leaving us with 9,025 Twitter users, which we then classified.

After discovering a number of inconsistencies with the overall BotOrNot score, we determined that more information was required for automated analysis of Twitter accounts. Using unsupervised machine learning to cluster accounts let us successfully organize a large number of accounts into separate categories, which we manually classified and verified. As described by Bessi and Ferrera[2], we retrieved the most important descriptors of bots: whether they're using the default appearance, their retweet-to-tweet ratio, and others, in addition to the *BotOrNot* score and category scores. However, our best results were found when simply clustering based on the *BotOrNot* category scores.

To classify each cluster, we set up a basic Python script using Selenium that displayed a sample set of Twitter feeds, and allowed a user to submit a category for the user[TODO: What categories?]. Based on the categories reported for the cluster, we could determine what type of Twitter account a user was likely to be. We then examined the average *BotOrNot* scores for these categories.

4. RESULTS & DISCUSSION

5. CONCLUSION

While some spambots use hashtags to make their tweets more visible, the vast majority of bots don't directly interact with human users. Instead, they're more likely to simply follow a large number of users, and either pass as human or hope that the user follows them back.

5.1 Further Work

Although we were able to manually identify advertising links, when we retrieved data on individual Tweets we did not expand Twitter's shortened URL format. This made media, such as photos, appear identical to other links, since Twitter represents media as URLs. In addition, the sample size for manual classification was small by necessity; setting up a web service such as Mechanical Turk to crowdsource this account classification would improve analysis and clustering.

6. REFERENCES

[1] S. Aral and D. Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9):1623–1639, 2011.

[2] A. Bessi and E. Ferrara. Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday*, 21(11), 2016.

[3] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *IEEE Trans. Dependable Secur. Comput.*, 9(6):811–824, Nov. 2012.

[4] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer. Botornot: A system to evaluate social bots. *CoRR*, abs/1602.00975, 2016.

[5] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian. Using sentiment to detect bots on twitter: Are humans more opinionated than bots? In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pages 620–627, Aug 2014.

[6] J. Ratkiewicz, M. Conover, M. Meiss, B. Goncalves, A. Flammini, and F. Menczer. Detecting and tracking political abuse in social media, 2011.

[7] G. Stringhini, C. Kruegel, and G. Vigna. Detecting spammers on social networks. In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, pages 1–9, New York, NY, USA, 2010. ACM.

[8] C. Xiao, D. M. Freeman, and T. Hwa. Detecting clusters of fake accounts in online social networks. In *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security*, AISec '15, pages 91–101, New York, NY, USA, 2015. ACM.

APPENDIX

A. CONTRIBUTIONS

Alic Szecsei provided data retrieval methods for Twitter accounts, programmed the machine learning clustering, and wrote the data analysis.

Willem DeJong programmed BotOrNot score retrieval methods, retrieved data for Twitter accounts to store in SQL databases, and created many of the graphs and charts.