Terminator 6: Humans vs Bots (on Twitter)

Munaf Arshad Qazi
New York University
maq249@nyu.edu

Maksim Temnogorod
New York University
mt1697@nyu.edu

***Abstract*—The dynamics of social media have become increasingly complex. One topic of particular relevance is the existence of automated bots, whose activity has proven to have significant societal consequences. Twitter, used by over 300 million users, is home to around 20 million such bots. The task of differentiating between these accounts and genuine human users is made all the more challenging by their variety. We are interested in studying the behavior of different bots, in part as an effort to promote Twitter as an open and unbiased platform.**

***Keywords*—twitter, machine learning, logistic regression, random forests, NLP.**

I. Introduction

Social media is one of the most powerful platforms for people to voice and propagate their ideas. However, it is often difficult to know exactly who (or what) is listening, forwarding, or even generating a user's words. The existence of bots on social media platforms has gained significant attention in recent years with respect to inflated follower counts for political candidates and other public figures on Twitter [1]. While some work in the detection of bot accounts continues to be conducted with the explicit perspective of minimizing spam, the broader Twitter bot landscape is likely more complex, and the variety of purposes for which bots are built should be taken into account.

We have observed that among the different bots inhabiting twitter, the least interesting (though still potentially problematic) are those that at first glance resemble an ordinary person but are in fact fake, existing purely as follower-list fodder. They are easily identified by a low follower/following ratio [3] and little activity otherwise. Another class of bots are novelty accounts such as those that compose linguistic experiments or else parody some cultural trope, often to humorous effect. These are distinguishing by high follower/following ratios (perhaps comparable to celebrities), frequent and temporally regular tweets, but little like/retweet activity. Also, they are often honest about their bot status. We also recognize a third class, identifiable by high follower and friend counts and more interaction in the form of likes and retweets. This includes special interest/news aggregates as well as advertising and phishing accounts. These bots may be either helpful or malicious, deceptive or overt, and are probably most difficult to distinguish from human users based on these features alone. We intend to factor these different classes of bots into our approach.

II. Motivation

In addition to being behind the popularity inflation of political figures, the use of Twitter bots have been known to promote misinformation and even dramatically influence the stock market [2]. Given the vast popularity of social media and the anonymity of online interactions, it is unsurprising that bots are capable of such subversive activity. While the potential to challenge democracy is on the more critical end of the bot-danger spectrum, on the other end there is spam and phishing. It is important to continuously improve bot detection methods, in part by paying attention to the diversity of bots on Twitter.

III. Related Work

[1] J. Dickerson, V. Kagan and V. Subrahmanian, "Using Sentiment to Detect Bots on Twitter: Are Humans more Opinionated than Bots?", in International Conference on ASONAM, 2014.
[2] E. Ferrara, O. Varol, C. Davis, F. Menczer and A. Flammini, "The rise of social bots", Communications of the ACM, vol. 59, no. 7, pp. 96-104, 2016.
[3] E. Shellman, "Bot or Not: an end-to-end data analysis in Python", 2015.

IV. Data

Our dataset consists of 100 English-language Twitter accounts (50 labeled as bots) with relevant features pulled from the Twitter API. The accounts identified as bots were collected by searching among user accounts with the query "bot." The remaining accounts were largely drawn from among our friends.

V. Algorithms Used

We propose to train separate models using random forest and logistic regression algorithms, and compare their performance before applying AdaBoost to improve the classifiers. Additionally, we intend to incorporate an SVM content analysis model to provide additional training features.