University of Maryland, College Park

Twitter Bot Classification

Intermediate Report

Mohammad Subhan, Sahil Dev, Gabe Margolis, Joshua Goldberg, Jordan Foster

CMSC396H

November 11, 2020

Introduction

Social media has become increasingly prevalent in the past decade. Hosting a wide variety of content ranging from family photos to online communities to discussions to news. There are many different social media platforms which exist on the internet today. Among them are Facebook, Youtube, Instagram, and Whatsapp which sit at the top of the most heavily used networking sites. As of July 2020, Twitter has the 15th most monthly active accounts with approximately 326 million users (Clement). Twitter is a well known social media platform which allows users to post and respond to "Tweets". Tweets can range anywhere from thoughts and questions to videos, GIFs, advertising, anecdotes, and news updates. Tweets are initially shared to the account's followers directly. Followers will then like or retweet content they are interested in to share it with their followers. This chain pushes popular tweets to reach millions of people. The largest percentage of twitter users come from the United States and Japan (Clement). However, in recent years there have been reports of fake bot accounts creating and sharing tweets under the guise of being a real human. This can have serious consequences and in some cases is considered election tampering.

The goal of this research is to be able to accurately identify these bots. By doing so we can create a more democratic process that does not involve the tampering of companies or other countries.

Although this problem has been tackled before by many researchers in the past, our goal is to compare the different solutions out there and figure out which is the most optimal. There is no research currently out there that looks at the problem holistically.

We will measure and evaluate the current solutions out there under a certain set of conditions and compare them. If there are a couple of programs out there that perform stellar the team may consider combining the approaches to develop an even more efficient approach.

By taking a holistic and cumulative approach to solving the problem of twitter bots the team hopes to find the most optimal solution out there. By doing so governments and companies can base their future research on this paper. One thing to note is by doing this research it does make it easier for creators of bots to figure out new approaches that can avoid the current techniques.

Problem Statement

Reports have shown that nearly 48 million accounts on Twitter are bots (Efthimion). Unfortunately, social networking sites do not distinguish between bot accounts and real users. Therefore, bots have the ability to deceive real users. A post from a bot could be difficult to distinguish compared to a real user. Bots have been found posting and responding to political tweets. In 2016, about one-fifth of tweets about the U.S election have come from a group of bot accounts (Efthimion). These bots greatly influenced the content that real Twitter users see on their timeline. They falsely depict the response that the majority of citizens and real users want to show. Many bots can show fake support or anger over the tweet of a politician or content of a news article while other bots spread false information. Twitter Bots which attempt to skew public perception of a person or event cause problems and unrest on social networking sites.

Related Work

In "Detection of Novel Social Botsby Ensembles of Specialized Classifiers" Sayyadiharikandeh et al. seek to differentiate the behaviors of bots and separate them into different categories. Each model is trained to recognize three specific categories of bots, which includes, but is not limited to, traditional spambots, social spambots, and fake followers. Traditional spambots bombard users with content, social spambots support and/or attack users, and fake followers frequently follow users (Sayyadiharikandeh et al., 2020). This approach to detecting bots on Twitter proved to be fairly accurate when compared to other successful bot detection methods. The main metric used for determining the accuracy of the method of bot detection used in the paper, the Ensemble of Specialized Classifiers (ESC), is the area under the ROC curve. This metric, in this case, is how successful a model can determine a human account from a bot account with 1.0 and 0.0 being the highest and lowest scores respectively. ESC produced an AUC of 0.96 which is comparable to the AUC of Botometer which produced one of 0.97 (Sayyadiharikandeh et al., 2020).

The work of Knauth in "Language-Agnostic Twitter Bot Detection" focuses on the fact that there are no human limitations by twitter bots. The data used in this research is from the MIB dataset which contained 8375 twitter accounts. As a result, account features such as number of friends, geo enabled, and protected as a means to determine if a twitter account is a bot not. The features that proved to be most beneficial in bot determination include Levenshtein distance, geo enabled, and statuses count. The work done in this paper was able to predict the account type with an accuracy of 98-99 percent. However, as noted in the paper, bot developers will pick up on research such as this and future researchers must use new data as existing data may not work so well. This is an important thing to note as the paper was written over a year ago.

The paper “Twitter Bot or Not?”, compares the performance of six machine learning models, with 28 account features, 12 features for individual tweets, and 19 features for aggregate data over the past 100 tweets for each account, providing for a somewhat language-aware solution that is only partially capable of interpreting the context of tweet. This differs from previous works mentioned prior, most of which focus on developing new methods of improving on a preexisting one. The six models used along with the cross-validation accuracy for each are as follows. Decision tree, a tree structure where at each step a comparison is made to determine which path to take and the leaves represent a final decision; achieved 89% accuracy. Random forest (with adaptive boosting) is a set of decision trees where the majority predicted class is selected; achieved 92% accuracy. Logistic regression (L1) uses a linear gradient descent model with a sigmoid component for binary classification with lasso regularization; achieved 89% accuracy. Logistic regression (L2) uses ridge regularization to achieve 88% accuracy. Multilayer Perceptron is a multilayer logistic regression model, where each layer feeds as input into the next which achieves 88% accuracy. Voting classifier combines several models and chooses the majority predicted class extending the concept of a random forest which achieves 93% accuracy. In terms of F1 score, the voting classifier performed the best (0.93), followed closely by the random forest with adaptive boosting (0.92) and L1 logistic regression (0.90). Their analysis uses the majority class classifier as a baseline with accuracy of 50.4%.

Challenges

There are many challenges when analyzing Twitter users and Tweets. Some publicly available datasets do not include the Tweet content, but only the Tweet id. By Twitter’s Developer Policy, in many scenarios a dataset can only publicly share the ids of tweets (Littman). While the content of Tweets still on Twitter can be downloaded using the Twitter API, there are additionally rate limitations in the API. Even given the Tweet message, there is also the issue of obtaining context from a short message. How do we represent context? Do we simplify context down to a one dimensional “sentiment” field that just tells us how negative/positive the Tweet is? Doesn't this throw out most of the information? Besides the Tweet content, there are also many other attributes that can be used to determine legitimacy including Tweet time, frequency, hashtags used, etc.

Progress and Planned Contributions

So far, we have found and imported multiple datasets into a Jupyter notebook in our GitHub repository: https://github.com/sdev00/Twitter-Bots. These datasets include user information objects including attributes such as account creation date, number of followers, number of friends, number of statuses, and verified status. We are researching classification algorithms and have applied a Naive Bayes classification and logistic regression to a single dataset to see how it works. We will soon begin to apply multiple past approaches, such as decision trees and random forests, to the datasets to further develop a strong understanding for how they work. We plan to improve methods used by others using machine learning and training techniques. Once seeing the strengths and weaknesses of each approach, we plan to refine an approach and expand on it with at least one novel method. For example, an important improvement over previous techniques would be context sensitivity, or the ability to associate bot activity with a particular motivation. This is trivial for humans but incredibly difficult to implement in a machine learning model. One possible roadblock is that the datasets that we have currently found do not include any tweet data. Attributes such as time of tweet, frequency of tweets and number of hashtags used may be useful in a model. However, getting this information for enough users to train a model may pose difficult due to no Tweets in our datasets and API limitations of Twitter. We will have to decide if we can accurately classify without tweet data or if it will be worth exploring how to get the tweets (or maybe k recent tweets) from a user using the Twitter API if the user is still on Twitter. Additionally, the Twitter API requires a developer account application that we have recently applied for. Furthermore, the free Twitter API has many limitations such as a limit of 250 requests per month.

Organization

We will review past approaches to this problem and discuss the benefits and shortcomings of each. We will then introduce our approach and analyze it in a similar manner.

References

“Bot Repository.” *Botometer.Osome.Iu.Edu*, 2020, botometer.osome.iu.edu/bot-repository/datasets.html. Accessed 29 Sept. 2020.

Clement, J. “Most Used Social Media Platform.” Statista, 21 Aug. 2020, www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

Clement, J. “Twitter: Most Users by Country.” Statista, 24 July 2020, www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/.

Edwards, Chad, et al. “Is That a Bot Running the Social Media Feed? Testing the Differences in Perceptions of Communication Quality for a Human Agent and a Bot Agent on Twitter.” Computers in Human Behavior, vol. 33, Apr. 2014, pp. 372–76. DOI.org (Crossref), doi:10.1016/j.chb.2013.08.013.

Efthimion, Phillip George; Payne, Scott; and Proferes, Nicholas (2018) "Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots," SMU Data Science Review: Vol. 1 : No. 2 , Article 5. https://scholar.smu.edu/datasciencereview/vol1/iss2/5

Knauth, Jürgen. “Language-Agnostic Twitter Bot Detection.” *Proceedings - Natural Language Processing in a Deep Learning World*, 2019, doi:10.26615/978-954-452-056-4_065.

Liberman, Neil. “Decision Trees and Random Forests.” Medium, Towards Data Science, 21 May 2020, towardsdatascience.com/decision-trees-and-random-forests-df0c3123f991.

Littman, Justin. “Where to Get Twitter Data for Academic Research.” *Social Feed Manager*, 14 Sept. 2017, gwu-libraries.github.io/sfm-ui/posts/2017-09-14-twitter-data.

Mattapalli, Revanth, et al. “Twitter Bot or Not?” 2017, github.com/Vignesh6v/Twitter-BotorNot/blob/master/twitter-bot_or_not.pdf.

Nagpal, Anuja. “L1 And L2 Regularization Methods.” *Medium*, Towards Data Science, 14 Oct. 2017, towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c.

Roeder, Oliver. “Why We’re Sharing 3 Million Russian Troll Tweets.” *FiveThirtyEight*, 31 July 2018, fivethirtyeight.com/features/why-were-sharing-3-million-russian-troll-tweets/. Accessed 29 Sept. 2020.

Sayyadiharikandeh, Mohsen, et al. "Detection of Novel Social Bots by Ensembles of Specialized Classifiers." *arXiv preprint arXiv:2006.06867* (2020).

“Social Media: Definition of Social Media by Oxford Dictionary on Lexico.com Also Meaning of Social Media.” Lexico Dictionaries | English, Lexico Dictionaries, www.lexico.com/definition/social_media.