Title: Analyzing Twitter Bot Activity on Academic Articles

Select one:

MACROBUTTON HTMLDirect ▢***Poster*** ***MACROBUTTON HTMLDirect*** ▢***Work-in-Progress Paper***

ABSTRACT

Given its ascendancy as a way to make connections worldwide, social media is affecting all areas of people's lives. This paper focuses on analyzing how Twitter bots interact with scholarly articles. The growing number and increasing complexity of Twitter bots make it hard to identify who is actually tweeting about scholarly articles. The purpose of this paper is to provide metrics to determine—based on an analysis of the relationship between Twitter bots and several research factors—whether a given scholarly paper has been disseminated via a bot. We developed and tested several supervised machine learning classification models that address this problem in relation to both numerical and categorical features, based on which the best results achieved was F1-score of 63%.

Background

Twitter is one of the world's most used social media platforms with 126 million daily active users (Shaban, 2019). Although Twitter is a useful medium for disseminating scholarly information (Mohammadi et al., 2018), the extent to which scientific papers are shared via Twitter bots rather than by individual users has yet to be determined. The operation of Twitter bot accounts is problematic because they can easily render communication noisy by generating huge amounts of traffic that may, in turn, have a negative influence on people's opinions (Efthimion et al., 2018).

Objective

To address this gap, this paper focuses on understanding Twitter bots' activity related to scientific papers. We built machine learning models to predict whether a research article was posted by Twitter bots.

Methods

We used two sources to obtain data for this study: Altmetric.com, and the Twitter API[1]. Using Altmetric dataset, we identified research articles mentioned on Twitter and the Twitter user handles that tweeted the articles. We used the Twitter API to gather more information about these Twitter accounts including followers, retweets and location.

Botometer (Davis et al., 2016) is one of the most popular tools to classify Twitter accounts as bots or not and has been used in different settings including health, political and shooting related events to identify non human Twitter accounts (Badawy et al., 2018; Broniatowski et al., 2018; Kitzie et al., 2018). In this study, we used Botometer to obtain a bot score for each Twitter account. Bot score is calculated based on several factors such as Twitter account activity, friends' network, and content.

From the Altmetric dataset, which contains more than 10 million entries and takes up to 13GB memory space, we used Twitter API to collect information about 182,277 Twitter accounts and the Botometer API to get their bot-scores.

We merged the Altmetric data with the data from Twitter and scores obtained from Botometer. Before feeding this data to our machine learning algorithms, we had to do some data preprocessing. Our Botometer dataset had eight categories of bot scores, each ranging from 0 to 5 as shown in Table 1. For each paper, we aggregated all its bot scores for all eight categories into a new feature called "overall score". These scores range from 0 to 40, with the higher numbers indicating a higher likelihood of a Twitter account being a bot. We used a higher threshold than the average of all the eight features. Additionally, we noticed that non-English Twitter accounts generate unusually high bot scores. Botometer is more likely to consider a non-English Twitter account as a bot. Thus, we cleaned our data by removing all tweeters not tweeting in English to avoid the need to extract non-English textual features and reduce misclassification.

***Table 1**: Distribution of Botometer Features*

Feature	Feature description
Content	User's tweet content score

[1] https://developer.twitter.com/

English	Whether the content is in English
Friend	How many friends/followers the user has
Network	Networks of friends of the user
Sentiment	Sentiment of the user’s tweets
Temporal	User’s behavior over time
Universal	An overall score for the user
User	User’s profile information

We also removed all the correlated features and the information we did not use in our models (e.g., mentions of articles on Reddit, Wikipedia, GooglePlus). Additionally, we noticed that the Altmetric score (a weighted count of all of the online attention an article has received) is not properly scaled and few scores go as high as 8000. We created a new feature by scaling the score between 0 and 1. After all this cleanup we prepared a new dataset with the features shown in Table 2.

Table 2*: List of features in the dataset*

Feature	Feature description
Altmetric ID	ID of the article in Altmetric
Scopus	The research area of the article
Altmetric score scaled	Altmetric score scaled between 0 and 1 based on social activity
User loc	Location of the user
Overall score	Overall bot-score for the Twitter user
isBot	Binary classifier generated from overall score based on a threshold of 20

Results

The mean overall score from Botometer was 11.6 while the median was 9.1. The highest bot score was 38.5 and the minimum score was zero. As most of the users had a bot score below 16, we took 20 as the bot threshold, which is 50% of the score and had proven efficient in previous studies (Shao et al., 2018). In Figure 1, we show the median bot score for each country. From the figure we can see that non-English speaking countries have higher median scores than the English speaking countries. This could be a bias towards the English language on Botometer score.

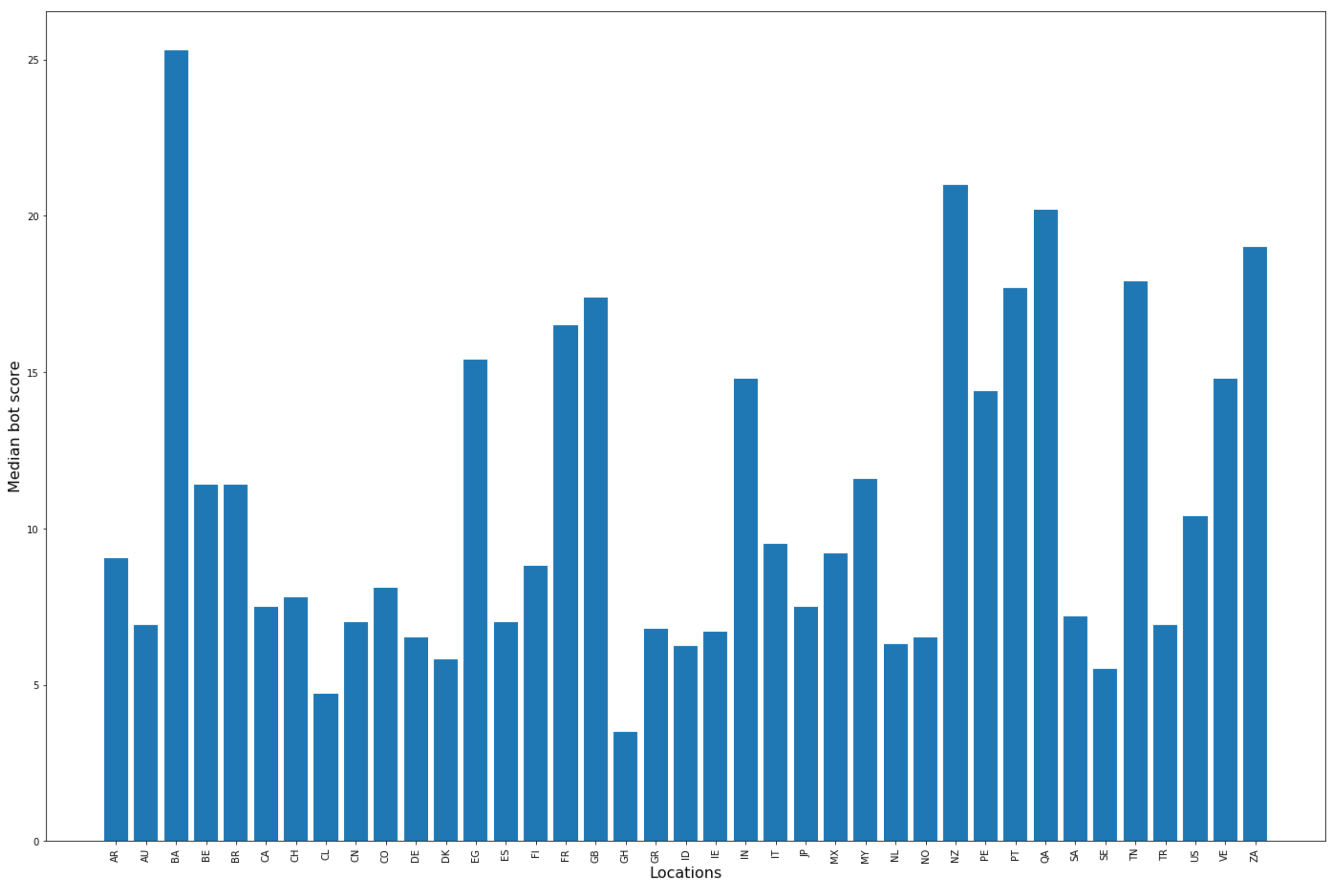

Figure 1*: Median bot scores based on tweeter location.*

We also noticed that some disciplines of research have more bot activity than others. Figure 2 below shows median bot scores for different research areas.

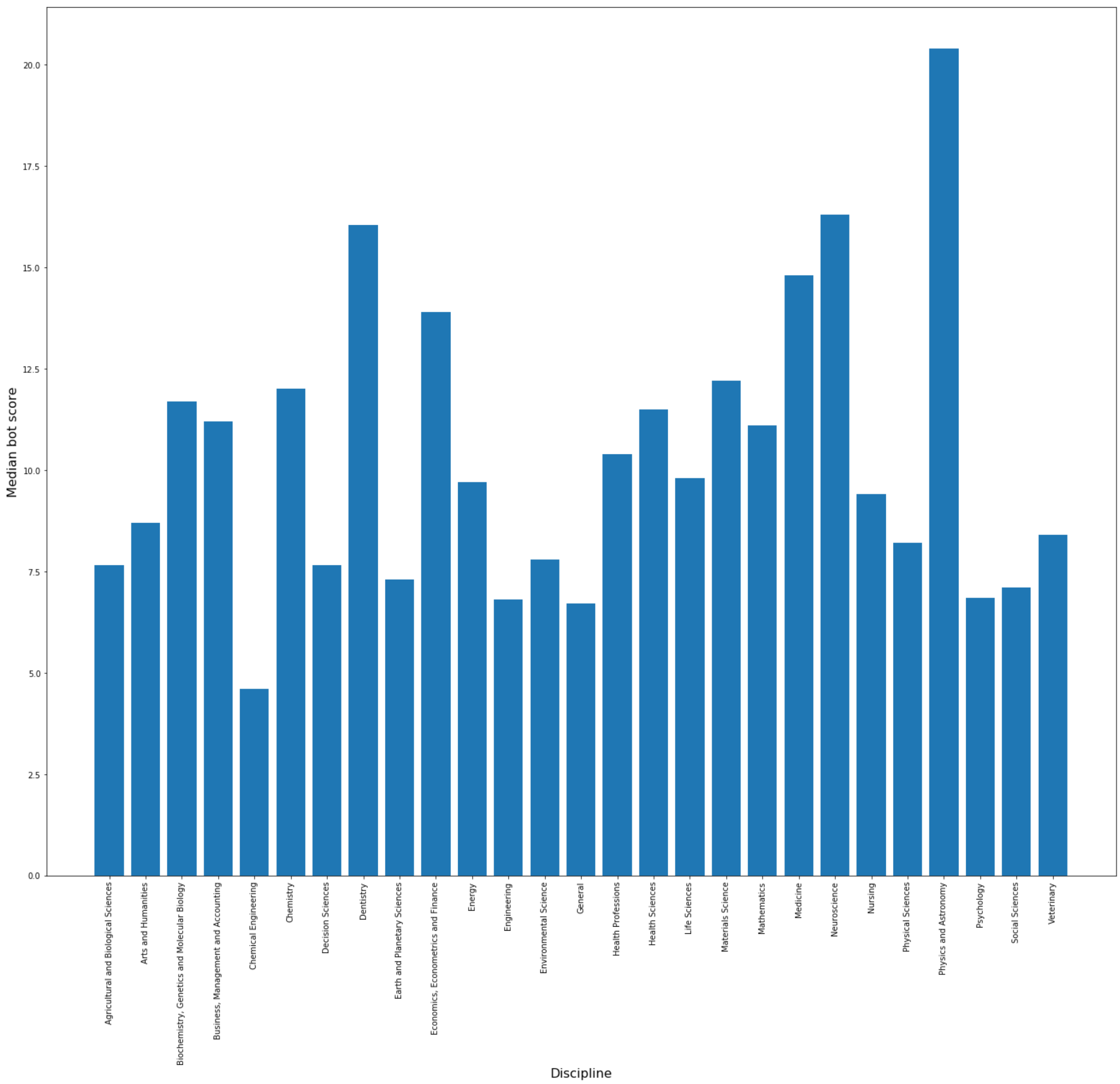

Figure 2*: Median bot scores based on research discipline.*

We created a new binary feature *isBot*, whereby each paper was classified as posted by bot or not posted by bot, based on its overall score in relation to a predetermined threshold (20). We created this new feature based on the Botometer score to create a clear distinction between bots' postings vs. non-bots' postings classifications, which helped us better train the models.

We developed our machine learning algorithms to classify Twitter accounts that posted research articles as bots or not. Setting the threshold at 20 for our models created an uneven distribution of data such that 25% of the articles were classified as posted by bots and around 75% were classified as not posted by bots.

We fed this data to linear regression and the k-nearest neighbors (KNN) classifiers, which reported accuracy of .86 and .85, respectively. From the classification report of the logistic regression model, we noticed that many of our tuples had been classified as false negatives.

In order to overcome this problem, we undersampled papers that were not posted by bots to produce an even distribution in our dataset. We created a new dataset with all the papers where the *isBot* is true and the same number of papers where *isBot* is false to make it an even distribution. While doing that, we also made sure to choose the papers not posted by bots in a random order so that the data is not skewed towards a certain category of research. With this new dataset, we ran the logistic regression classifier which returned an F-1 accuracy of 0.54, and a support vector machine (SVM) classifier which returned a score of 0.56 as shown in Table 3.

We then used the KNN algorithm with the number of neighbors set to five and a cross-fold validation with five folds as well, which yielded a score of 0.63.

Table 3: *Results from different classification models*

Classifier	F1-score	Comment
Logistic regression	0.54	Accuracy on evenly distributed dataset
SVM	0.56	Accuracy on evenly distributed dataset
KNN	**0.63**	Better accuracy on evenly distributed dataset

Future Work

In the future, we plan to collect more data, build more textual features, and run different models such as Random Forest and Neural Networks to improve the model accuracy. We also plan to cluster the research papers and identify the level of spamming based on different research subjects and areas. We also plan to build a model to predict the possibility and scale of spamming on a certain research article and research area. We will also consider building a model that takes into consideration non-English tweets.

References

Badawy, A., Ferrara, E., & Lerman, K. (2018). Analyzing the Digital Traces of Political

Manipulation: The 2016 Russian Interference Twitter Campaign. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 258–265.

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health*, *108*(10), 1378–1384.

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). Botornot: A system to evaluate social bots. *Proceedings of the 25th International Conference Companion on World Wide Web*, 273–274.

Efthimion, P. G., Payne, S., & Proferes, N. (2018). Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots. *SMU Data Science Review*, *1*(2), 5.

Kitzie, V. L., Mohammadi, E., & Karami, A. (2018). “Life never matters in the DEMOCRATS MIND”: Examining strategies of retweeted social bots during a mass shooting event. *Proceedings of the Association for Information Science and Technology*, *55*(1), 254–263.

Mohammadi, E., Thelwall, M., Kwasny, M., & Holmes, K. L. (2018). Academic information on Twitter: A user survey. *PloS One*, *13*(5), e0197265.

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*(1), 4787.

Hamza Shaban (2019). Twitter reveals its daily active user numbers for the first time. (2019, February 7). *The Washington Post*. https://www.washingtonpost.com/technology/2019/02/07/twitter-reveals-its-daily-active-user-numbers-first-time/