[2]

Bot Identification: Helping Analysts for Right Data in Twitter

Pradeep Kumar Tiwari
Central Research Laboratory
Bharat Electronics Limited
Bengaluru, India
pradeepkumart@bel.co.in

T Velayutham
Central Research Laboratory
Bharat Electronics Limited
Bengaluru, India
tvelayutham@bel.co.in

***Abstract*— As social networks are becoming popular; it raises concerns among data analyzers for the quality of content over social media platforms. For better and fair predictive analysis, the quality of data is important. Low quality content may result into prediction of improper cause of an event, misleading trending issues and more importantly the sensitive stock price may fluctuate. The content over social media may be flooded or corrupted by various bots such as Influence Bots, Spam Bots. We are targeting twitter for the identification of such bots, as it is mostly used by data scientists for applications related to scientific prediction and sentiment analysis. In this paper, we capitalize on earlier approaches and used a machine learning based approach for the classification between a bot profile and human profile. We have identified 10 attributes of user profile and tweet pattern for an account and calculated a score called *botScore* for each profile to model the behavior as bot or as human. We have extended the list of features in distinguishing between bot and human to more fine-grained label. The method proposed was found to be more accurate than traditional Baye's classification technique.**

Keywords*— right data; bot Identification; social networks; TweetPattern; BotScore; Sentiment Analysis;

I. Introduction

Today it is well-known that bots are hovering over Twitter. Although Twitter has rather made stringent anti-spam policies. Anyone with little knowledge of programming can create bots. There are few bots which acts only when they see a particular keyword, usually the trending topic related keywords. Other bots may randomly tweet preset phrases such as proverbs or defined statements. Or in few cases, if the bot is designed to emulate a popular entity (Media Channel, celebrity, historic icon, anime character etc.) their popular phrases or exclusive comments will be tweeted. However, not all the bots are fully automated, and the term "bot" usually refers to twitter accounts that are simply "fake" accounts. [7].

According to studies, nearly 7 percent of the followers of average user are "fake", not humans. Some studies [10], though less scientific put that number as high as 35 percent. Twitter has sued some of the biggest suppliers of these fake accounts in federal court to shut down them. However, until this is resolved bots will keep flourishing into the twitter environment.

Twitter bots [1] include:

- *Spambots* spreads spam on various topics.
- *Paybots* illicitly make money. Some paybots copy tweet content from respected sources like @BBC and paste in the form of tiny-URLs that direct users to sites that pay the bot creator for redirecting traffic to their site.
- *Influencebots* try to influence Twitter conversations on a specific topic. Recently some politicians have been accused of buying influence on social media.

Snoopy (@SNOOPYbot), Nagato FakeBot (@NagatoFakeBot), Moomin (@moomin_valley), Peter Drucker (@DruckerBOT), Random phrases (@kotoba_bot) are few popular Japanese bots used for authorized purposes.

This paper makes the following contributions:

- Semi-Automated but scalable approach to identify twitter bots based on critical features of the user profile.
- Methodology to detect and test the proposed approach over the collected dataset of 138774 users and over a bot sample of 5000 accounts.
- Comparison with Baye's classifier on prediction accuracy.

II. Background

As of July 2016, Twitter reported 500 million tweets per day and 255 million monthly active users[1] with 10% of the India. population purportedly on Twitter[2].

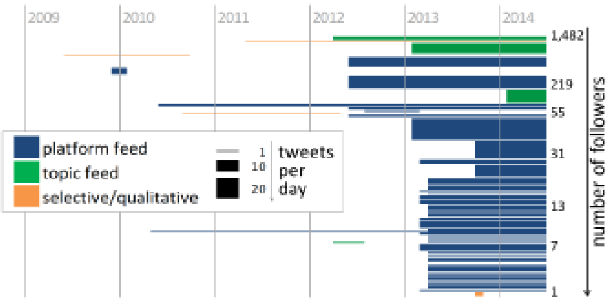


Fig1: Timeline of average tweeting activity from first to last tweet for 51 Twitter accounts. [1]

Studies [15, 16, 17] showed that there is a pattern in tweeting activity which varies from selective tweets to platform feeds

[1] https://about.twitter.com/company/
[2] Calculated based on data from: http://www.census.gov/population/international/data/

978-15090-6403-8/17/$31.00 © 2017 IEEE

and the quanta of followers. Based on tweeting pattern, it is likely to predict what a user is going to tweet.

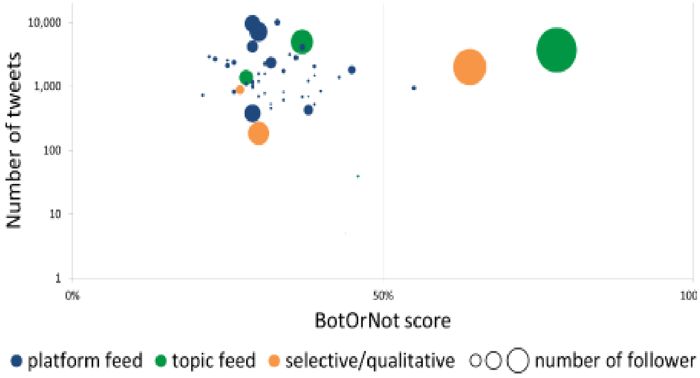


Fig2: Number of tweets and followers, BotOrNot score, and type of account for 51 Twitter accounts [3].

As the number of tweets increases and follow a similar pattern of tweets, it becomes more likely to be having higher bot score. Indicated by

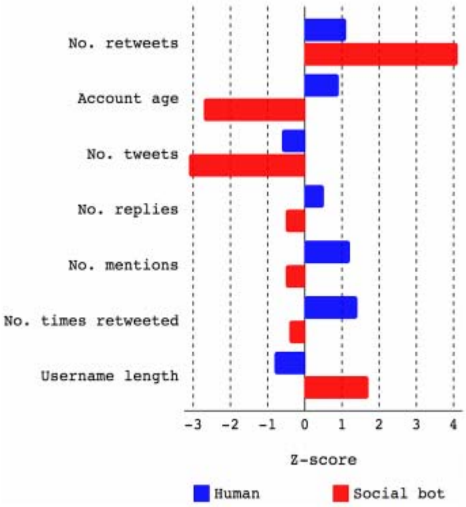


Fig3: Classification based on above parameters [4]

Figure 3 shows that the User behavior is the best way to discriminate social bots from humans. Social bots retweet more often than humans and have longer and arbitrary user names, whereas genuine accounts have some similarity with the name of the user. While bots produce frequent tweets, replies and mentions, they are retweeted less than humans. Also, the bot accounts tend to be more recent.

III. Related Work

Twitter bots are becoming ubiquitous and at the same time they are growing increasingly creative. Studies reveal that 16% of Twitter accounts "Highest possibility of being automated" (Zhang & Paxson, 2011, p. 102) [6], and a work by Chu, Gianvecchio, Wang, and Jajodia (2012) [3] [12] reports that 10.5% of Twitter accounts are bots, with an additional 36.2% categorized as "cyborgs" (defined as a "Human Assisted Bot or vice versa").

Studies show that majority of Twitter bots follow other tweeters, however this may be due to "follow-spam," which makes a profile as spam by expressing it as bot followers (Mowbray, 2010) [7]. It also can be a mean to influence the follower to make him follow back, as mentioned by Mowbray in his 'The Twittering Machine' [7]. Conversely, according to Lotan (2014) [8], the bots may be deployed to achieve the same purpose and are available at the rate of $5 for 4,000 followers, which makes it easier and inexpensive to have large 'followings'.

Many bots declare themselves as automated and contribute twitter with benign tweets of news feed updates, blog updates, images and video updates (Chu, Gianvecchio, Wang, & Jajodia, 2012, p. 812) [6]. In point of fact, Twitter is also exploring the automated bots for various purposes. For example, the account @MagicRecs analyzes its followers twitter activities and sends suggestions and recommendations based on the analyzed data (Satuluri, 2013) [9].

The Truthy project's BotOrNot[3] determines the probability of a twitter account being a bot especially social bot based on machine and statistical learning (Ferrara, Varol, Davis, Menczer & Flammini, 2014) [4]. Most other automated methods of analyzing Twitter accounts have found lower evidences of bot given the account is twitter verified and for those accounts which has high number of followers.

Recently DARPA conducted a twitter bot challenge [11] to identify Influence bots, which are illicitly shaping discussions regarding terror activities by ISIS, efficiently. Since the intended challenge focused on identifying influence bots that are trying to diffuse a sentiment *s* on a topic *t*, participants had to:

- Segregate influence bots from other bots.
- Segregate influence bots about topic *t* from that of other topics.
- Segregate influence bots about topic *t* that sought to spread sentiment *s* from that of neutral or opposite of it

Though, this competition was specific to a particular challenge, it exposed many attributes which contribute to a profile contributing to a bot.

IV. Proposed Methodology

A. Attributes of Socialbots

We selected following attributes from the twitter accounts:

User Profile:

- Reputation (#f followers) / (#f followers + #f friends): This is a standard metric for estimating the popularity of users in Twitter [5].
- Profile Completeness: Amount and accuracy of information provided in the profile.
- Length of Usernames: Usually related to the name of user, however it varies from 4 to 15 characters.
- Age of the Account: Depending on two factors, one date of birth and other date of creation of account weights are assigned to it.

Tweet Patterns:

[3] http://truthy.indiana.edu/botornot/

- Tweeting Device pattern: Usage statistics for a period of time over the web from a mobile device or from a Computer or through the APIs (bots use API and humans use Web or mobile).
- Tweeting interval (periodic / regular): Frequency of tweets over a period of time, time of tweet, periodicity of tweet, regularity of tweet.
- Content of tweet: What is posted as part of tweet, its content length variation, url_mentions pattern, url_ratio per tweet (much higher for bots).
- Similarity between tweets: Comparing tweets to get a similarity score. We are calculating similarity score from length of tweet, no of tag_mentions, no of url_mentions?

The objective is to build a classifier to identify accounts likely belonging to bots, and we took a supervised learning approach. "Supervised" means we need labeled data, i.e. we need to know at the outset which accounts belong to bots and which belongs to humans.

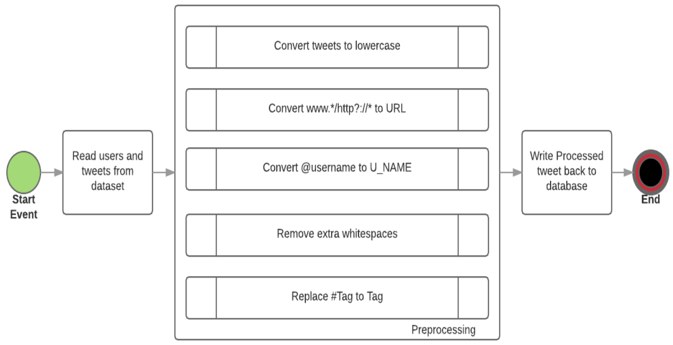

Start Event
Read users and tweets from dataset
Convert tweets to lowercase
Convert www.*/http?://* to URL
Convert @username to U_NAME
Remove extra whitespaces
Replace #Tag to Tag
Preprocessing
Write Processed tweet back to database
End

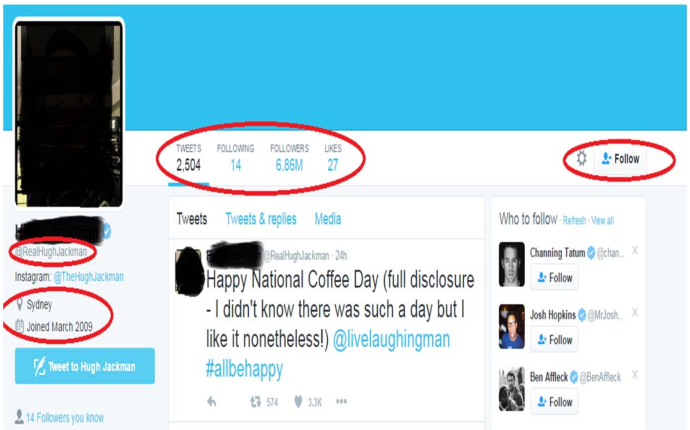
Fig4: Preprocessing the Collected Tweets

Though not being a follower/ friend of a particular profile, we still can get the required information about a user: Name, UserName/length, AgeOfAccount, Tweets, Followers and Following and their Count. It's important and makes our dataset to create feature vector.

B. Step by Step Training Process

Step 1: To automatically classify a user/tweet, first the classifier needs to be trained. To do that, a list of manually classified users/tweets is required.

Step 2: A feature vector needs to be created. The feature vector is the most crucial item in employing a classifier. A good feature vector can foresee how successful the results of the classifier will be.


[redacted]
[redacted]

[redacted] Happy National Coffee Day (full disclosure - I didn't know there was such a day but I like it nonetheless!) @livelaughingman #allbehappy

Fig5: Profile Attributes Collected of a NonFriend/Follower [14]

Step 3: Profile Score is the first member of the feature list array and TweetPattern score is the second member of feature list array. The feature list is used to train classifiers. We are using a custom classifier, named BotClassifier.

Step 4: In a sample tweet such as "*Happy National Coffee Day (full disclosure - I didn't know there was such a day but I like it nonetheless!) @livelaughingman #allbehappy*", the features to be extracted are "@livelaughingman, "#allbehappy," "urls, if any" and "length of text of tweet" .

Creating Labels

Jajodia et al. [12] manually inspected accounts and applied a Twitter version of the Turing test - if it looks like a bot, and tweets like a bot, then it's a bot. We have increased our number of followers from one of the publicly available service for research purpose, fiverr (a website offering dubious services) from 54 to 5000, so We came to know that remaining 4946 followers are bots. We will build our model based on our following bot's behavior.

C. Creating Features

With the help of Twitter REST API's richness, we created feature set without violating most of the twitter's term of service and with the help of python APIs we were able to collect user profile information in a JSON blob. The JSON blob can be converted to required data structure by preprocessing it. We were able to get friends, followers and following counts. However, twitter limits the number of resources one can ask through APIs, that is 180.

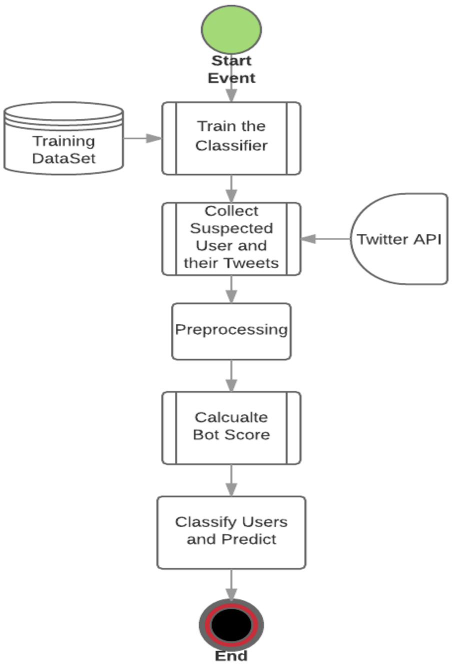
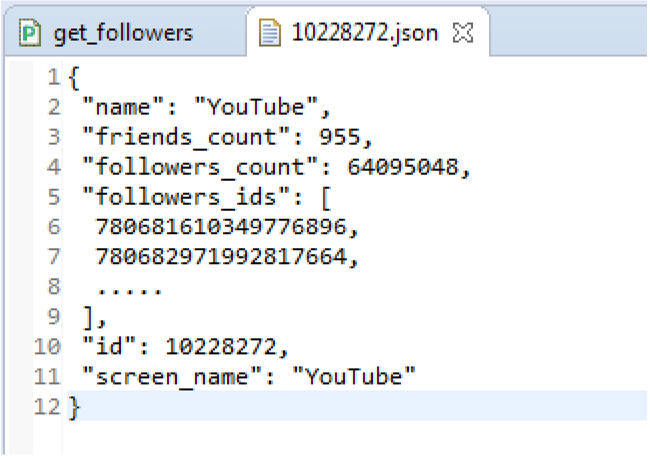


Fig6: Flow chart of the Classification

We achieved this challenge by receiving the data in chunks, usually 15-minute window is provided, and later merging it. A sample outset looks like following:



```
{
"name": "YouTube",
"friends_count": 955,
"followers_count": 64095048,
"followers_ids": [
 780681610349776896,
 780682971992817664,
 .....
],
"id": 10228272,
"screen_name": "YouTube"
}
```

Fig7: Sample Outset from twitter-APIs

D. Creating of Feature Vector

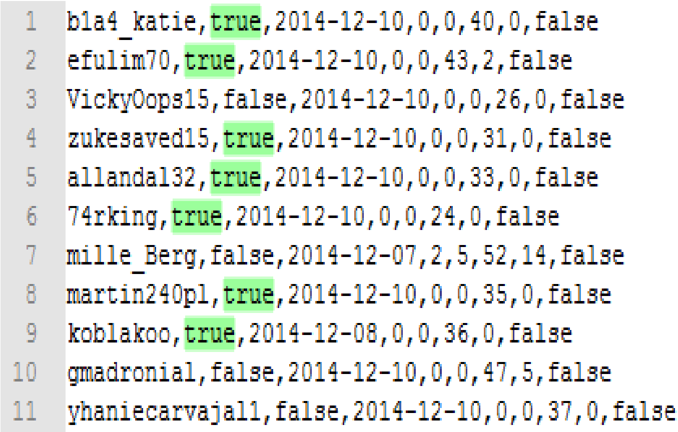
Feature vector consist of three dimensions: *profileScore* (∑ Wi f(value)), *tweetPatternScore* (∑ Wi f(value)) and third dimensions as result human or bot.

For *prfileScore*, we are using weighted score, which is calculated based on

- Followers and followings – reputation (followers divided by followers plus followings)
- Screen_name – length (int)
- Profile_image – default/Boolean
- Created_At – date (will give age)

For *tweetPatternScore*, a random sample of tweet with a fixed window of 100 tweets was selected and following parameters were taken into consideration:

- Frequency of tweets (e.g. 10 tweets/day/hour)
- Similarity between tweets: length of tweet, no of user mentions (hash tags), no of tag mentions (@ tags), url_mentions (hyperlinks)
- Periodicity: yes/no, if Yes, hourly (2), daily (1), weekly (0.5)



```
b1a4_katie,true,2014-12-10,0,0,40,0,false
efulim70,true,2014-12-10,0,0,43,2,false
VickyOops15,false,2014-12-10,0,0,26,0,false
zukesaved15,true,2014-12-10,0,0,31,0,false
allandal32,true,2014-12-10,0,0,33,0,false
74rking,true,2014-12-10,0,0,24,0,false
mille_Berg,false,2014-12-07,2,5,52,14,false
martin240pl,true,2014-12-10,0,0,35,0,false
koblakoo,true,2014-12-08,0,0,36,0,false
gmadronial,false,2014-12-10,0,0,47,5,false
yhaniecarvajal1,false,2014-12-10,0,0,37,0,false
```

Fig8: Sample Collection of Data with 8 Attributes

We have implemented the above approach using the following pseudo code:

Pseudocode 1: Determining Bot/Human status

```
1) CollectionOfTweets ← Using Python-twitter APIs
2) Users ← Using Python-twitter APIs and extracting
          from CollectionOfTweets
3) Create dataset of bot profiles from fiverr[5]
4) processedTweetsDatabase ← preprocess(tweet)
5) for every profile
   a) profileMap[] = select (screen_name, profile_img,
        created_on, followers_count, friends_count)
            from processedTweetsDatabase;
   b) reputation = followers_count / (followers_count +
            friends_count)
   c) age = Normalized (created_on)
   d) profileScore =$W_1$(1/created_on)+$W_2$(1/reputation)
        +$W_3$(length(screen_name))+$W_4$(profile_img Value)
   e) TweetSimilarityScore =
        frequency(occuranceOfTweets)*(avgLengthofTwee
        t+avgNoOfUserMentions+avgNoOfTags+avgNoOf
        URLMentions)/4 * periodicity
   f) featureVector=[profileScore,TweetSimilarityScore
            ] : Bot/Human
   g)  if (TweetSimilarityScore >= 0.75 &&
```

[4] https://github.com/mirkeet/TwitterBot/

[5] https://www.fiverr.com/

```
                profileScore >= 0.80)
                profile = bot;
        else
                profile = human;
        end if
    end for
6)  return profile
```

V. Results and Discussions

We compared our technique of classification with the Naiive Bayes classification and found that its results are encouraging.

TABLE I. Characteristics of data sets

Data Set	No of Examples	Input Attributes	Output Classes
mirKeet[4] Tweeter Bot	138774	8	2

TABLE II. Number of Instances in the training and test data set

Data Set	No Of Training Data	No of Test Data	Total
mirKeet[4] Tweeter Bot (and Fiverr)	138774	500	139274

TABLE III. Performance Metrices

Classifier	Phase	TP rate	FP Rate	Precision
BotClassifier	Train	0.896	0.191	0.830
	Test	0.840	0.203	0.802
Naiive Bayes	Train	0.783	0.260	0.783
	Test	0.797	0.253	0.799

VI. Conclusion and future scope

We developed an approach to identify twitter bots with an accuracy of almost 83%. We hinted on the possibility of existence of spurious accounts on social media. The approach, though not mature will evolve with the bigger data set and advanced machine learning algorithms. Also, the approach could further be extended to other social networks like facebook, quora, Instagram etc.

Acknowledgment

We would like to acknowledge the contribution and support given from our organization, Bharat Electronics Ltd, Bangalore in order to carry out this work efficiently. We would also like to acknowledge our friends who equally helped us in doing the same.

References

[1] Haustein, Stefanie, et al. "Tweets as impact indicators: Examining the implications of automated “bot” accounts on Twitter." *Journal of the Association for Information Science and Technology* 67.1 (2016): 232-238.

[2] Gilani, Zafar, et al. "Stweeler: A Framework for Twitter Bot Analysis."*Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016.

[3] Ferrara, Emilio, et al. "The rise of social bots." *arXiv preprint arXiv:1407.5225* (2014).

[4] Cha, Meeyoung, et al. "Measuring User Influence in Twitter: The Million Follower Fallacy." *ICWSM* 10.10-17 (2010): 30.

[5] Zhang, Chao Michael, and Vern Paxson. "Detecting and analyzing automated activity on twitter." International Conference on Passive and Active Network Measurement. Springer Berlin Heidelberg, 2011.

[6] Mowbray, Miranda. "The Twittering Machine." In WEBIST (2), pp. 299-304. 2010.

[7] Satuluri, V. (2015, September 24). Stay in the know [blog post]. Retrieved from https://blog.twitter.com/2013/stay-in-the-know

[8] http://www.forbes.com/sites/tristanlouis/2013/04/07/twitters-growing-spam-problem/#30abfe5b4a59, accessed on 20 October, 2016

[9] Communication to: V.S. Subrahmanian, Dept. of Computer Science & UMIACS, University of Maryland, College Park, MD 20742. vs@cs.umd.edu

[10] Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011, December). The socialbot network: when bots socialize for fame and money. In Proceedings of the 27th Annual Computer Security Applications Conference (pp. 93-102). ACM.

[11] Available at wileyonlinelibrary.com [Accessed 10 Feb 2017]

[12] https://twitter.com/RealHughJackman [Accessed 10 Feb 2017]

[13] Huang, Yuki, and Jie Shan. "Modeling and Visualizing Regular Tweeting Pattern of Purdue Campus." (2015).

[14] Bruns, Axel, and Stefan Stieglitz. "Towards more systematic Twitter analysis: metrics for tweeting activities." International Journal of Social Research Methodology 16.2 (2013): 9