CSE 472: Social Media Mining
Project II- Evolution of the Bots
Project Report

Jay Shah
jgshah1@asu.edu
Arizona State University

Kunal Vinay Kumar Suthar
ksuthar1@asu.edu
Arizona State University

1^{st} December 2019

Abstract

With the advent of internet there has been substantial increase in content polluters and fake bots on various popular social media platforms like Twitter, Facebook and many others. Out of these, political bots majorly involve in intentionally automating interactions related to political issues and elections. Hence it becomes critical for such a platform to detect these bots and eliminate them when found. And with increased knowledge they have improved over the time to refrain from getting detected using traditional methods. Through this study we have (1) collected a subset of data-set comprising of accounts who were active on topics related to 2016 US Presidential elections (2) used cluster analysis to identify different types of bots and their patterns (3) compared these results to that of Lee et al.[2] to understand intentions of these bots (4) presented an analysis on how these bots have evolved over time with knowledge.

Keywords fake bot detection, evolution of bots, cluster analysis, US election twitter data, social media mining, influence of twitter bots on 2016 US election

1 Introduction

1.1 Problem Statement

The fundamental purpose of any social media platform is to allow interactions amongst individuals, entities and groups of individuals [1]. There has been a great increase in the amount of consumption of many such platforms that govern various types of interactions. Starting from updates in the business worlds, technology sector to recent happenings in sports and political issues and latest works in academia and research. And people's trust on such platforms have been highest than ever before[3]. Hence it becomes the utmost responsibility of these platforms to monitor any misuse of such an offering that can potentially influence the crowd in a biased manner. Not all bots are harmful, some are benign too. For instance, some of them accumulate content from various domains serving as a simple news feed or others provide service as customer care to growing demands for companies.

Recent political scenarios such as the 2016 US presidential elections witnessed a lot of *fake bots* and *fake news* on platforms like Twitter, whose agenda had been to propagate biased and deceptive news within the majority and automate user interactions in the process.

The real challenge here is not to find anomalous behaviors within social media accounts, because most of the fake bots nowadays try to create an impression that certain content is highly endorsed by many other and hence it creates an unintentional influence of such a bot exploiting the vulnerability of the platform [5]. They try and create a profile with credible followers so that it becomes hard for any algorithm or human to detect. And with increasing dependence on social media these bots pose the risk of increasing the visibility of deceiving content benefiting a particular entity in a biased manner.

Hence through this study we aim to find and understand the behaviors of such bots and how it have evolved comparing it to Lee et. al[2] results. Their results indicated key characteristics about the content polluters found in 2011 US election data set and provided insights to build robust classifiers. We plan to map our findings to theirs and provide an analysis how they have evolved over the time using similar metrics and clustering techniques. In the sections below we describe how we scraped and processed the data, performed cluster analysis and understanding the intentions based on the extracted features.

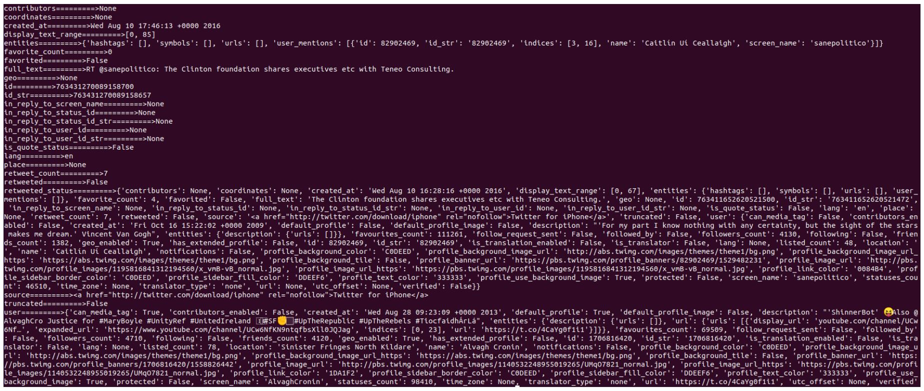
2 Implementation and Results

2.1 Task 1: Data Collection and Pre-Processing

In order to extract relevant Twitter data for 2016 US elections, we obtained the tweet ids of the tweets that were related to political topics during the election timeline from Harvard Data-verse[4]. Using the tweet ids from this data, we were able to extract the tweet-specific data for each one of them using Hydrator tool [10] that includes: [*'contributors', 'coordinates', 'created at', 'display text range', 'entities', 'favorite count', 'favorited', 'full text', 'geo', 'id', 'id str', 'in reply to screen name', 'in reply to status id', 'in reply to status', 'id str', 'in reply to user id', 'in reply to user id str', 'is quote status', 'lang', 'place', 'retweet count', 'retweeted', 'retweeted status', 'source', 'truncated', 'user'*] (Sample Data Structure shown in Fig. 1).

From this data, containing both human-user accounts and bot accounts, we have used Botometer API[6] to detect whether a particular account is a bot or not. We use a threshold score of 2.5 to judge whether an account is a bot or not. Account with scores > 2.5 are judged as bot and $<= 2.5$ are judged as not bots. From a total amount of 3,00,000 tweet ids we were able to scrape 1,85,564 tweets, because these involved tweets that were deleted by Twitter due to some reasons. And from those tweets, we found 57,196 twitter accounts. These 57,196 accounts include both humans and bots potentially. In order to filter out humans and find the bot accounts we use the Botometer API[6]. We found 9016 bot accounts out of 57,196 twitter handles. We have made our code public to access[1].

[1] https://github.com/jaygshah/Evolution-of-Fake-Bots-on-Twitter



```
contributors=========>None
coordinates=========>None
created_at=========>Wed Aug 10 17:46:13 +0000 2016
display_text_range=========>[0, 85]
entities=========>{'hashtags': [], 'symbols': [], 'urls': [], 'user_mentions': [{'id': 82902469, 'id_str': '82902469', 'indices': [3, 16], 'name': 'Caitlín Uí Ceallaigh', 'screen_name': 'sanepolitico'}]}
favorite_count=========>0
favorited=========>False
full_text=========>RT @sanepolitico: The Clinton foundation shares executives etc with Teneo Consulting.
geo=========>None
id=========>763431270089158700
id_str=========>763431270089158657
in_reply_to_screen_name=========>None
in_reply_to_status_id=========>None
in_reply_to_status_id_str=========>None
in_reply_to_user_id=========>None
in_reply_to_user_id_str=========>None
is_quote_status=========>False
lang=========>en
place=========>None
retweet_count=========>7
retweeted=========>False
retweeted_status=========>{'contributors': None, 'coordinates': None, 'created_at': 'Wed Aug 10 16:28:16 +0000 2016', 'display_text_range': [0, 67], 'entities': {'hashtags': [], 'symbols': [], 'urls': [], 'user_
mentions': []}, 'favorite_count': 4, 'favorited': False, 'full_text': 'The Clinton foundation shares executives etc with Teneo Consulting.', 'geo': None, 'id': 763411652620521500, 'id_str': '763411652620521472',
 'in_reply_to_screen_name': None, 'in_reply_to_status_id': None, 'in_reply_to_status_id_str': None, 'in_reply_to_user_id': None, 'in_reply_to_user_id_str': None, 'is_quote_status': False, 'lang': 'en', 'place':
None, 'retweet_count': 7, 'retweeted': False, 'source': '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>', 'truncated': False, 'user': {'can_media_tag': False, 'contributors_en
abled': False, 'created_at': 'Fri Oct 16 15:22:02 +0000 2009', 'default_profile': False, 'default_profile_image': False, 'description': "'For my part I know nothing with any certainty, but the sight of the stars
 makes me dream.' Vincent Van Gogh", 'entities': {'description': {'urls': []}}, 'favourites_count': 111261, 'follow_request_sent': False, 'followed_by': False, 'followers_count': 4130, 'following': False, 'frien
ds_count': 1382, 'geo_enabled': True, 'has_extended_profile': False, 'id': 82902469, 'id_str': '82902469', 'is_translation_enabled': False, 'is_translator': False, 'lang': None, 'listed_count': 48, 'location': '
', 'name': 'Caitlín Uí Ceallaigh', 'notifications': False, 'profile_background_color': 'C0DEED', 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png', 'profile_background_image_url_
https': 'https://abs.twimg.com/images/themes/theme1/bg.png', 'profile_background_tile': False, 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/82902469/1529482231', 'profile_image_url': 'http://pbs.
twimg.com/profile_images/1195816841312194560/x_vmB-vB_normal.jpg', 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1195816841312194560/x_vmB-vB_normal.jpg', 'profile_link_color': '0084B4', 'prof
ile_sidebar_border_color': 'C0DEED', 'profile_sidebar_fill_color': 'DDEEF6', 'profile_text_color': '333333', 'profile_use_background_image': True, 'protected': False, 'screen_name': 'sanepolitico', 'statuses_cou
nt': 46510, 'time_zone': None, 'translator_type': 'none', 'url': None, 'utc_offset': None, 'verified': False}}
source=========><a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
truncated=========>False
user=========>{'can_media_tag': True, 'contributors_enabled': False, 'created_at': 'Wed Aug 28 09:23:09 +0000 2013', 'default_profile': True, 'default_profile_image': False, 'description': "'ShinnerBot' 😝Also @
AlvaghCro Justice for #MaryBoyle #UnityRef #UnitedIreland □#SF✊□#UpTheRepublic #UpTheRebels #TiocfaidhÁrLá", 'entities': {'description': {'urls': []}, 'url': {'urls': [{'display_url': 'youtube.com/channel/UCw
6Nf…', 'expanded_url': 'https://www.youtube.com/channel/UCw6NfKN9ntqfbsXll0JQJag', 'indices': [0, 23], 'url': 'https://t.co/4CaYg0f1i1'}]}}, 'favourites_count': 69509, 'follow_request_sent': False, 'followed_by'
: False, 'followers_count': 4710, 'following': False, 'friends_count': 4120, 'geo_enabled': True, 'has_extended_profile': False, 'id': 1706816420, 'id_str': '1706816420', 'is_translation_enabled': False, 'is_tra
nslator': False, 'lang': None, 'listed_count': 78, 'location': 'Sinister Fringes North Kildare', 'name': 'Alvagh Cronin', 'notifications': False, 'profile_background_color': 'C0DEED', 'profile_background_image_u
rl': 'http://abs.twimg.com/images/themes/theme1/bg.png', 'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png', 'profile_background_tile': False, 'profile_banner_url': 'https:
//pbs.twimg.com/profile_banners/1706816420/1558826442', 'profile_image_url': 'http://pbs.twimg.com/profile_images/1140532248955019265/UMqO7821_normal.jpg', 'profile_image_url_https': 'https://pbs.twimg.com/profi
le_images/1140532248955019265/UMqO7821_normal.jpg', 'profile_link_color': '1DA1F2', 'profile_sidebar_border_color': 'C0DEED', 'profile_sidebar_fill_color': 'DDEEF6', 'profile_text_color': '333333', 'profile_use_
background_image': True, 'protected': False, 'screen_name': 'AlvaghCronin', 'statuses_count': 98410, 'time_zone': None, 'translator_type': 'none', 'url': 'https://t.co/4CaYg0f1i1', 'utc_offset': None, 'verified'
```

Figure 1: Data Structure of the extracted tweet

2.2 Task 2: Cluster Analysis

2.2.1 Features

The quality of a classifier is dependent on the discriminative power of the features. From the data structure in Figure 1 above, we extract total of 10 features which forms a 10 dimensional feature vector : *[length of screen-name, longevity, total friends, total followers, friend-follower ratio, total tweets, Average number of tweets per day, total retweets received, total favorited, total favorites received]* adherent to the features extracted in [2]. In the paper [2], there are four types of user features: (1) User Demographics(UD) (2) User Friendship Networks(UFN) (3) User Content(UC) (4)User History (UH). We were not able to incorporate all features present in the above types in the paper[2], but a subset of those features due to limited data and restricted twitter API calls. The features selected in the above types are:

- **UD**: (1) length of screen name (2) longevity of the account
- **UFN**: (1) the number of following (2) the number of followers (3) the ratio of the number of following and followers
- **UC**: (1) the number of posted tweets (2) the average number of posted tweets per day (3) total retweets received (4) total favorited (5) total favorites received

2.2.2 Clustering

We experiment clustering on a subset of these 10 features in order to find clusters with highest discrimination power. For this data we created a feature matrix consisting of [*length of screen name, longevity of account, friend-follower ratio, total tweets, Average number of tweets per day, total retweets received, total favorites received]* for each account instance.

Here we chose these seven features for the following reasons:

- **length of screen name** and **longevity of account** gives us the descriptive information about a user and it's particular account.
- **friend-follower ratio** gives us the information about the account's friendship networks.

- **number of tweets** and **number of tweets per day** tell us about the account activity, as many bots tend to be active in short spans of time [7], [8].
- **retweets received** and **favorites received** tells us more about the confidence whether the account is bot or not. [9] hypothesized that bots cannot generate their own content hence this feature measures should as low as possible in the ideal scenarios of bot accounts.

We perform cluster analysis using two Unsupervised approaches: Probabilistic and Vectorized Clustering methods. First of all, we need to estimate the number of clusters in the data set. We use *Density-Based Spatial Clustering of Applications with Noise(DBSCAN)* to estimate the inherent cluster dimensionality of the data. After determining the number of clusters, we perform our two approaches.

For the probabilistic approach, we use the iterative method of Expectation Maximization(EM) algorithm used in the paper by Lee et al.[2] to perform clustering. We model the number of clusters as *Gaussian Mixture Models(GMM)* i.e. model the clusters as Gaussians and the whole data space as the sum of Gaussians, and estimate the optimal set of parameters using *Expectation Maximization(EM)*. After the end of the algorithm, we now have all the data points with a label i.e. each one belongs to a cluster. For the vectorized approach, we use the K-means algorithm to perform clustering. It randomly allocates cluster centroids and iteratively computes new centroids till the centroids converge.

For visualizing them in two dimensions, we use the feature dimensionality reduction techniques of PCA and TSNE. Principal Component Analysis(PCA) is a dimensionality reduction technique which preserves variance and discrimination power. T-distributed Stochastic Neighbor Embedding(TSNE) is a dimensionality reduction algorithm which is used to visualize high dimensional datasets. Using this algorithm, we get well spread clusters in 2 dimensional space. We reduce dimensions from 9016 x 10 to 9016 x 2, in order to visualize the clusters in 2D. The clusters for the GMM and KMeans on PCA and TSNE are shown in the following figures 2, 3, 4 and 5 respectively.

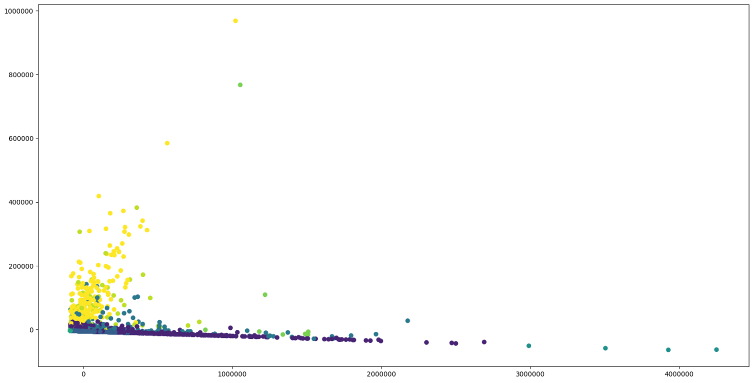


Figure 2: Visualization on GMM and PCA

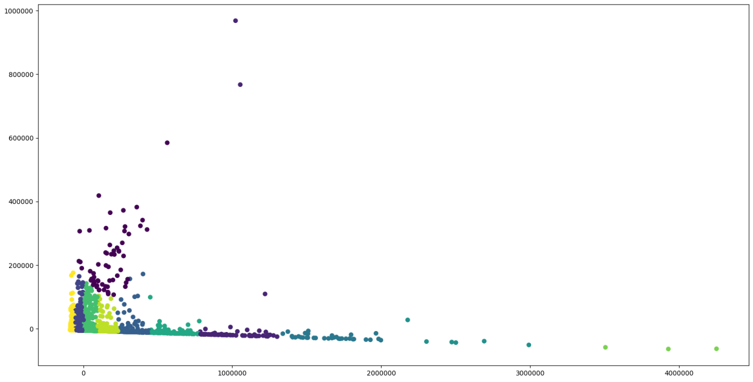


Figure 3: Visualization on K-Means and PCA

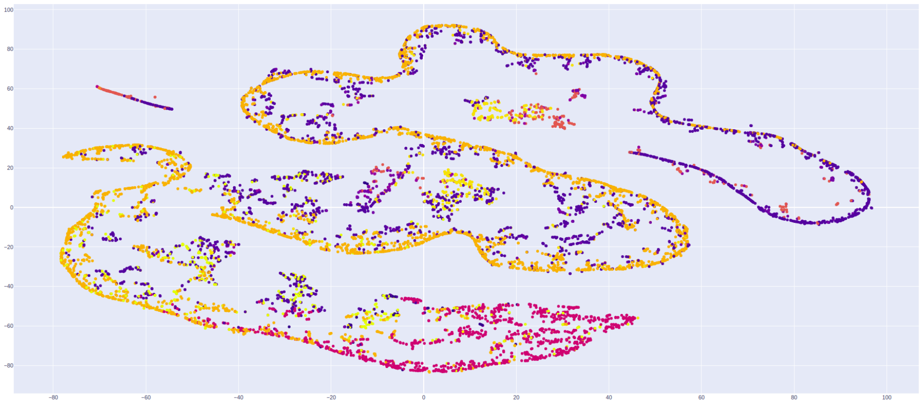


Figure 4: Visualization on GMM and TSNE

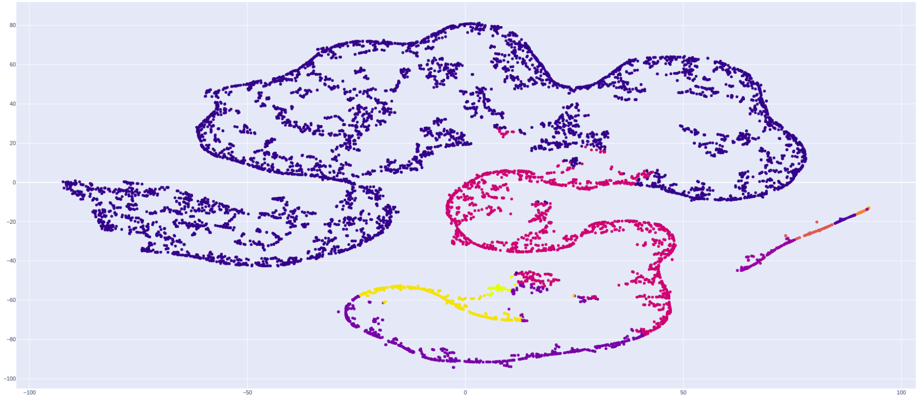


Figure 5: Visualization on K-Means and TSNE

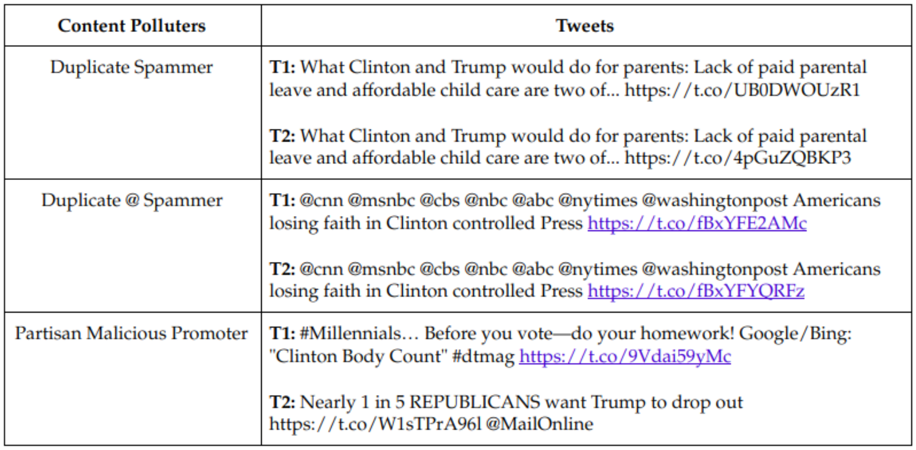
2.3 Task 3: Understanding the intention of each detected type

In the experiments performed by Lee et al. in [2], they generated 9 clusters in the data set using Expectation-Maximization(EM). These clusters were further grouped into four major categories of content polluters:

- **Duplicate Spammers**: These content polluters post almost identical tweets with or without links.
- **Duplicate @ Spammers**: These content polluters are similar to the Duplicate Spammers, in that they post tweets with a nearly identical content payload, but they also abuse Twitter's @username mechanism by randomly inserting a legitimate user's @username. In this way, a content polluter's tweet will be delivered to a legitimate user, even though the legitimate user does not follow the content polluter.

- **Malicious Promoters**: These content polluters post tweets about online business, marketing, finance and so on. They have a lot of following and followers. Their posting approach is more sophisticated than other content polluters because they post legitimate tweets (e.g., greetings or expressing appreciation) between promoting tweets.
- **Friend Infiltrators**: Their profiles and tweets are seemingly legitimate, but they abuse the reciprocity in following relationships on Twitter.

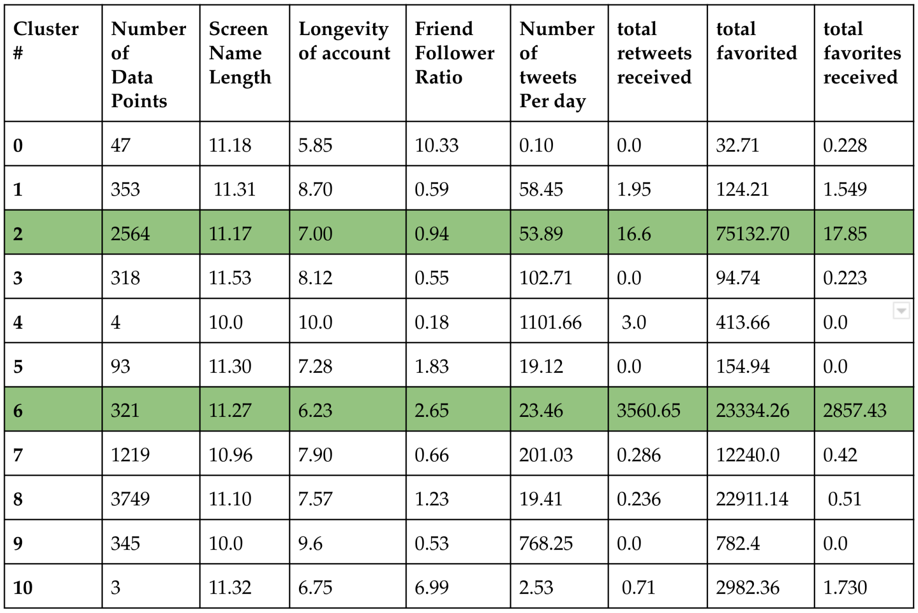
In our cluster analysis, we find 11 clusters using DBSCAN on 9016 bots. We perform clustering using Expectation Maximization(EM) algorithm and we also find similar types/groups of content polluters: Duplicate Spammer, Duplicate @ Spammer, and Partisan Malicious Promoters(Illustrated in Figure 6 below).



Content Polluters	**Tweets**
Duplicate Spammer	**T1:** What Clinton and Trump would do for parents: Lack of paid parental leave and affordable child care are two of... https://t.co/UB0DWOUzR1 **T2:** What Clinton and Trump would do for parents: Lack of paid parental leave and affordable child care are two of... https://t.co/4pGuZQBKP3
Duplicate @ Spammer	**T1:** @cnn @msnbc @cbs @nbc @abc @nytimes @washingtonpost Americans losing faith in Clinton controlled Press https://t.co/fBxYFE2AMc **T2:** @cnn @msnbc @cbs @nbc @abc @nytimes @washingtonpost Americans losing faith in Clinton controlled Press https://t.co/fBxYFYQRFz
Partisan Malicious Promoter	**T1:** #Millennials… Before you vote—do your homework! Google/Bing: "Clinton Body Count" #dtmag https://t.co/9Vdai59yMc **T2:** Nearly 1 in 5 REPUBLICANS want Trump to drop out https://t.co/W1sTPrA96l @MailOnline

Figure 6: Content Polluters

The average features for our 11 clusters are shown in Figure 7 below:



Cluster #	Number of Data Points	Screen Name Length	Longevity of account	Friend Follower Ratio	Number of tweets Per day	total retweets received	total favorited	total favorites received
0	47	11.18	5.85	10.33	0.10	0.0	32.71	0.228
1	353	11.31	8.70	0.59	58.45	1.95	124.21	1.549
2	2564	11.17	7.00	0.94	53.89	16.6	75132.70	17.85
3	318	11.53	8.12	0.55	102.71	0.0	94.74	0.223
4	4	10.0	10.0	0.18	1101.66	3.0	413.66	0.0
5	93	11.30	7.28	1.83	19.12	0.0	154.94	0.0
6	321	11.27	6.23	2.65	23.46	3560.65	23334.26	2857.43
7	1219	10.96	7.90	0.66	201.03	0.286	12240.0	0.42
8	3749	11.10	7.57	1.23	19.41	0.236	22911.14	0.51
9	345	10.0	9.6	0.53	768.25	0.0	782.4	0.0
10	3	11.32	6.75	6.99	2.53	0.71	2982.36	1.730

Figure 7: Average feature vectors for every cluster

2.4 Task 4: Analysis-How bots have changed/appeared/disappeared throughout the years?

Although Lee et.al[2] collected data over a period of time to analyse user behaviors, we have collected data of accounts who were active and posting related to political discussions during the election timeline. Based on the data they had, they extracted features related to user demographics, friendship networks, tweet content and the temporal aspects to analyse their automation fashion that challenged conventional bot detection methods. We use similar features excluding the temporal aspects of an account to understand their behavior.

As shown from figure-7, we are able to detect Duplicate and Duplicate @ Spammers [clusters: 0, 5 and 7]. Whereas for the sophisticated bots from the subset of data we performed analysis on, we found two types of major sophisticated bots:

- **Sophisticated-Active bots**: We call bots sophisticated-active as these bots have a long account history, decent amount of followers but they were active only during the election time with content posting i.e. number of tweets per day and favorited tweets with little feedback from the community i.e. retweets received and favorites received. For instance, here, [clusters: 1, 3, 4, 7, 8 and 9]
- **Sophisticated-Influential**: These are bots who have also been active during the election timeline but their content has been curated in a more sophisticated manner. They received a relatively more amount of endorsements from the community and their activity on the platform has not been random enough compared to Duplicate or Sophisticated-Active bots. Here, they are clusters: 2 and 6.

3 System requirements and Execution instructions

Operating System: Ubuntu Linux/Mac

Step 1: Setup a python 3.7 virtual environment by typing the following commands:

- *sudo apt install python3.7 python3-venv python3.7-venv*
- *python3.7 -m venv smmp2*
- *source smmp2/bin/activate*

Step 2: Install the following dependencies using pip and activate the virtual environment or run the *setup.sh* bash script:

- *pip install botometer==1.4*
- *pip install certifi==2019.9.11*
- *pip install chardet==3.0.4*
- *pip install cycler==0.10.0*
- *pip install idna==2.8*
- *pip install joblib==0.14.0*
- *pip install kiwisolver==1.1.0*
- *pip install matplotlib==3.1.2*
- *pip install numpy==1.17.4*
- *pip install oauthlib==3.1.0*
- *pip install pkg-resources==0.0.0*
- *pip install plotly==4.3.0*
- *pip install pyparsing==2.4.5*
- *pip install PySocks==1.7.1*
- *pip install python-dateutil==2.8.1*
- *pip install requests==2.22.0*
- *pip install requests-oauthlib==1.3.0*
- *pip install retrying==1.3.3*
- *pip install scikit-learn==0.21.3*
- *pip install scipy==1.3.3*
- *pip install six==1.13.0*
- *pip install sklearn==0.0*
- *pip install tweepy==3.8.0*
- *pip install urllib3==1.25.7*

Step 3: Download *data-4.json* file from the following link and copy it in the code folder.

- *https://drive.google.com/file/d/1jimmQULyMMu9Qj77DJFfroHxfVsb1iDP/view*

Step 4: Within the virtualenv run the following command to generate the 57196 x 10 feature vectors from the JSON data structure:

- *python generate_feature_vectors.py*

Step 5: Run the following command to classify the twitter users as bot or human:[The user_labels.pickle file which is created by this command is already provided, so this step can be skipped]

- *python detect_bot.py*

Step 6: Run the following command to perform clustering and display the clusters:

- *python generate_display_clusters.py*

4 Conclusion

Through this project, we hypothesize and study that with the increase in technology and analysis the sophistication of bots on social media platforms has increased in reference to political discourse. We observe the bots categories from Lee's study and are able to spot similar and more sophisticated bots in our data set, which have evolved over time. Our analysis leads us to believe that traditional machine learning and data analysis methods may require a number of features and may not be apt at correctly identifying bots. We were able to cluster 9,016 bot accounts into distinguishable 11 clusters and further into 3 major categories.

5 Limitations and Future Work

For this project, we used only 3,00,000 tweet ids out of the 50,00,000 tweet ids available on the Harvard data-verse[4] owing to the limitations of CPU and memory utilization of our personal machines. Also another factor that influenced our decision for smaller data size was because of the Botometer API[6]. It took us approximately 120 hours to label the accounts as a bot or not. Hence it would have not been feasible to include more data maintaining the deliverable time-frame in mind. We plan to overcome this limitation and extend this approach including more data for better insights.

As of now we use only the tweet mentions and @ tags within clustering approaches. One of the things we can extend this is to use content based clustering with some domain level knowledge of making communities. For instance, here, using content to cluster bots based on their inclinations to certain party specific to use of keywords and using other language processing tools.

Also, we have included only the tweets labeled as relevant to political discussions by Harvard Data-verse[4]. We plan to study and analyse results of tweets that are not related to election topics in order to better classify these bots. Limited amount of data specific to certain might result in biased conclusions. We plan to compare bots' interactions, event-based activities and temporal information for both election related and non-related.

6 Acknowledgements

We would like to thank Professor Huan Liu for teaching the course CSE 472 at Arizona State University[11]. Insights from his lectures and his book[1] proved very helpful to us in deciding the methodologies to be used and deriving conclusions. Also, we would like to thank Tahora H. Nazer for her experienced guidance on this project topic.

7 References

[1] R.Zafarani, M. Ali Abbasi, H.Liu, ”Social Media Mining, An Introduction”
[2] Kyumin Lee, Brian David Eoff, and James Caverlee. Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter. In ICWSM. AAAI, 2011.
[3] Soo Hui Lee, ”Does trust really matter? A quantitative study of college students' trust and use of news media”
[4]US 2016 Elections Dataset from https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/PDI7IN

[5] Aiello, L.M., Deplano, M., Schifanella, R. and Ruffo, G. People are strange when you're a stranger: Impact and influence of bots on social networks. In Proceedings of the 6th AAAI International Conference on Weblogs and Social Media (2012). AAAI, 10–17

[6] Botometer API access from https://rapidapi.com/OSoMe/api/botometer/details

[7] Sangho Lee and Jong Kim. Early filtering of ephemeral malicious accounts on Twitter. Computer Communications, 54:48–57, 2014.

[8] Yinglian Xie, Fang Yu, Kannan Achan, Rina Panigrahy, Geoff Hulten, and Ivan Osipkov. Spamming botnets: signatures and characteristics. ACM SIGCOMM Computer Communication Review, 38(4):171–182, 2008.

[9] Jacob Ratkiewicz, Michael Conover, Mark Meiss, Bruno Gonalves, Alessandro Flammini, , and Filippo Menczer. Detecting and Tracking Political Abuse in Social Media. In ICWSM, pages 297–304. AAAI, 2011.

[10] Tweet Content extraction from https://github.com/DocNow/hydrator

[11] http://www.public.asu.edu/ huanliu/SMM18F/cse472.htm