Bot and Gender Identification from Twitter

Notebook for PAN at CLEF 2019

Radarapu Rakesh, Yogesh Vishwakarma,
Akkajosyula Surya Sai Gopal, and Anand Kumar M

Department of Information Technology, NITK, Surathkal
{rakeshradarapu77, surya.akkajosyula, thisisyogeshvishwakarma}@gmail.com,
m_anandkumar@nitk.edu.in

Abstract The popularity of social media raises a concern about the quality of content over its platforms. The quality of data is important, especially for fair and considerable predictive analysis. If the quality of data is less, it may result in the prediction of wrong circumstances of an event. This causes misleading trending problems and more importantly, the sensitive stock price may fluctuate. The contents available on social media can be corrupted and overflowed by bots. There are a variety of bots available such as Spam Bots, Influence Bots, etc. Our target is to identify such bots on Twitter. Twitter data is mostly used by data analysts for applications related to scientific predictions or opinion analysis. This working note is capitalized on earlier approaches and Machine Learning (ML) approaches used to classify between a bot and human and find the gender further for interesting studies in crime detection etc. By sharing many attributes for user profiles, we have identified the pattern to find out that the given user is a bot or human based on the tweets posted.

1 Introduction

Before we dive into the proposed methodology used for solving the Bot and Gender identification task, let us have a look at why Author Profiling is needed and its effects on society. We will go through the background of Twitter and see how bots and fake accounts have an adverse impact.

1.1 Background of Twitter

Twitter is a social networking service which was originated in America and serves as an online news portal. It provides users to post messages (tweets) and also to interact privately with another user. A tweet was limited to only 140 characters until November 7, 2017. This limit was doubled for almost all popular languages on Twitter from then. Twitter provides features like post, like, and retweet a tweet for any registered user,

while those unregistered can only read. Twitter can be accessed through the web and an application is provided for the convenience of smartphone users. Twitter, Inc. has its headquarters in San Francisco, California, and spreads across the world with more than 25 offices [4]. Twitter took birth in March 2006 and Jack Dorsey, Noah Glass, Biz Stone, Evan Williams were the ones who came up with this revolutionary platform [5]. It gained popularity within no time since its launch in July of that year. By 2012, around 1.6 billion search queries per day were handled by Twitter with more than 100 million users posting around 340 million tweets a day. As of 2016, the monthly active users count raised to more than 319 million. With debates and news covering politics of the United States, it has been a hot-bed ever since 2015. Twitter was the largest source of breaking news on the day of the U.S. presidential election. Around 40 million tweets were sent by 10:00 p.m. (Eastern Time) that day regarding the elections. It provided the information on Brett Kavanaugh's Supreme Court nomination [1] [9] and the 2018 midterm elections of the United States [2].

1.2 Impact of Bots on Twitter

Nowadays the presence of bots over Twitter has increased rapidly. Even though Twitter has anti-spam policies, the bots can bypass them easily. Bots are easy to implement by anyone who has basic knowledge on that domain. Some bots react to a particular keyword which is more likely related to a trending topic. Some bots randomly tweet predefined statements or phrases. In some other cases, the bots are designed to highlight a popular entity like an anime character, historic icon, celebrity characters, etc based on the popular phrases related to the fields. In addition to them, some bots are not fully automated (in other words we can say fake accounts). According to studies [18], around 7 of the followers of a Twitter account are fake. Some studies [19], though less scientific put that number as high as 35 percent. To solve all these problems Twitter has sued suppliers of fake accounts to shut down them.

Recently some politicians have been accused of buying influence on social media. These are a few popular Japanese bots used for authorized purposes [3]:

- Snoopy (@SNOOPYbot)
- Nagato FakeBot (@NagatoFakeBot)
- Moomin (@moominvalley)
- Peter Drucker (@DruckerBOT)
- Random phrases (@kotobabot)

2 Literature Survey

Now, let us take a look at the work done earlier and the studies related to our proposed Author Profiling task. Twitter bots are becoming ubiquitous. They are increasingly growing and becoming more creative.

- Studies show that 16 percent of user accounts have the possibility of being automated [18].

- A study [10] reveals that 10.5 percentage of accounts are bots, and around 36.2 percentage categorized as "cyborgs" ("Human Assisted Bots").
- Studies say that a majority of bots lead to "follow-spam" (making a profile as spam by its bot followers) for Twitter users [18].
- It may serve as a follow back request, as mentioned by Mowbray in his 'The Twittering Machine' [10].
- According to Lotan [6], the bots may be deployed to have a big number of followers for a user who pays a good amount for these fake followers just to raise popularity.
- Many bots declare themselves as "automated" and manipulate news feed updates, blog updates, images and video updates with their benign tweets (Ex: Chu, Gianvecchio, Wang, and Jajodia, 2012, p. 812) [18].
- Twitter is also thinking of using automated bots for various purposes. For example, the account @MagicResend analyzes its followers' Twitter activities and sends the analyzed data which can help in identifying the fake accounts [6].
- 'BotOrNot' is a project which determines the probability of an account being a bot based on machine learning techniques [7].
- Other methods of analyzing accounts that were automated have found lower evidence of bot given the verified accounts and for accounts with a high number of followers [17].
- Recently a Twitter bot challenge was conducted by DARPA [7] that aims to identify 'Influence Bots', which are promoting discussions regarding terror activities by ISIS, etc. [11].
- The paper [19] proposed gender identification in Russian Tweets with different classical ML classifiers, namely Logistic regression (LR), Support Vector Machine (SVM).
- PAN 2017 Author Profiling Task [15] discusses an overview of various approaches based on a deep neural network (DNN) and machine learning. However, deep learning could not outperform traditional machine learning models.
- PAN 2018 Author Profiling Task [14] describes an overview of gender identification from a multimodal perspective, where texts but also images are given as data set.

3 Problem Statement

Bots on social media create influence on the users for commercial, political or ideological purposes by posing as humans. For example, consider when we are shopping we usually check the rating of the product and the number of ratings, taking this as an advantage the bots will rate the product and change the view of the product (positively or negatively). The bots on social media can directly tilt the fate of politics which makes them an even greater threat (for reference see Brexit referendum [8] or US Presidential elections [2]). This influence made the German political parties reject the use of bots in their election campaign. In other ways, bots are well known for spreading rumours or fake news. With these major threats, the identification of the bots on social media using Author Profiling raised its importance from various domains like marketing, forensics, and security. The aim is to find whether the author is human or bot and if the author is human we have to further classify gender (male or female) based on the tweets given.

4 Proposed Methodology

In this section, we will see the data set description and methodology used to build our proposed model and its parameters.

4.1 Data set Description

The training data set of the English author profiling task at PAN 2019 consists of XML files each corresponding to a unique author (Twitter user) and there are a total of 4,119 files and a validation set contains 1240 XML files.

- An XML file for each author (Twitter user) with a total of 100 tweets. The name of the XML file is a unique id that corresponds to the author.
- A truth.txt file that contains a list of these author-ids and the truth about gender.
- The truth file gives the author-id, truth for the human/bot and bot/male/female as well.

4.2 Pre-processing of Data set

For pre-processing of text, we are using the NLTK library to create a TweetTokenizer, TreeBankTokenizer, etc to perform tokenization and some other procedure using regex and identified the emoticons and renamed them with negative and positive emojis to generalize the emojis used. We are using the following pre-processing steps as shown in Figure 1.

4.3 System Architecture

The task is to find the gender from a given set of tweets - is it a bot or a human, further a male or a female, in case of a human. For this, we have used the sequential neural network classifier which is fed with 'max-num' frequent words, this sequence is then analyzed to predict the gender. We have also used Linear SVM as a learning algorithm that is used for solving multi-class classification problems from large data sets. Figure 2 gives an idea of our proposed methodology.

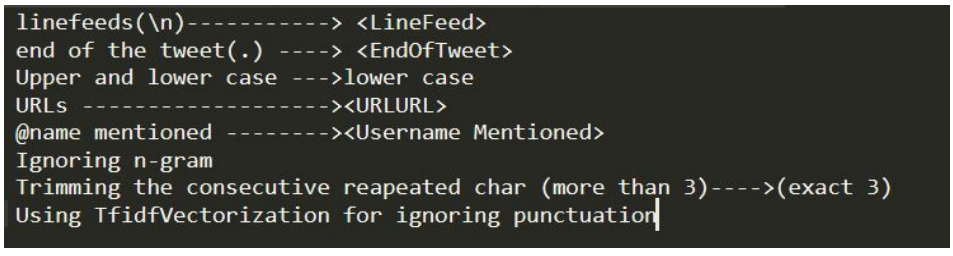

Figure 1. Pre-processing of Data set

We are using vectorize pipeline i.e., we obtained vectors as a pipeline of word and character n-grams. We performed (with using two methods: manually and the scikit-learns grid search function) several hyper-parameter tuning experiments for best parameter values. Finally, we combined the training and validation data and tuned the parameters manually for performing cross-validation. We considered only the English data set for the shared task, but the model can be extended to or implemented on other language data sets as well. We experimented with different features and the final model contains the following n-grams features in it:

- Word unigrams, bigrams, and trigrams.
- Character (3,5)-grams.
- Term frequency-inverse document frequency (tf-idf) weighting.
- Use of LSA to perform dimensional reduction on n-grams.

In Latent Semantic Analysis (LSA), Linear dimensionality reduction was done using truncated Singular-Value Decomposition (SVD), using the TruncatedSVD function from the scikit-learn library.

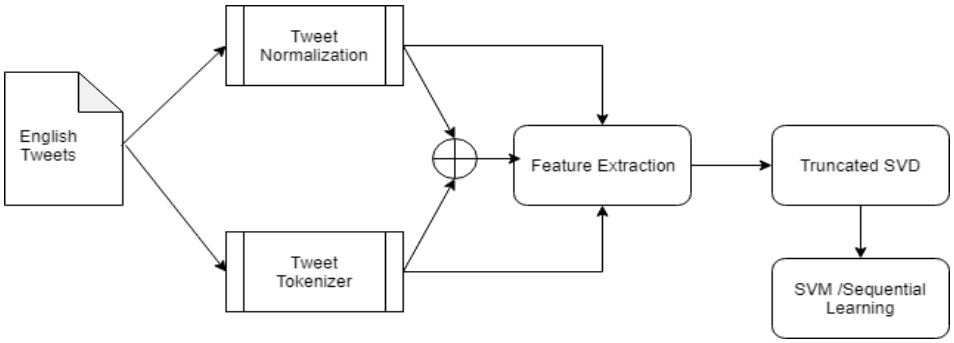

Figure 2. Concept Diagram

Let us look at the detailed steps and parameters we have tried in the two approaches. One was with neural networks using the 'keras' library. We used three 'dense' layers with 1024, 512, 3 neurons at the corresponding levels, with the activation functions like 'relu', 'sigmoid' and 'softmax'. We used the 'Tokenizer' from keras to prepare the input, which is an index array. The other approach is using 'LinearSVC' itself, but with a different pre-processing. We used the regular expression matching to find typical character sequences that are used as emojis, user-mentions, URLs, etc. and gave them a positive/negative score based on the relevance with the tweet. These two approaches also gave us significant cross-validation results, as shown in Table 1. So, we combined these pre-processing, added a few more as mentioned previously in section 4.2 and made our final model.

5 Results and Analysis

The performance of the proposed author profiling system is calculated by accuracy. First, we need to calculate the accuracy for identifying bots vs humans. Then in the case of humans, we will calculate the accuracy for identifying males vs. females. But, we tried to identify the male or female or bot at first and then from the result, we identified whether it is a human or a bot, just to simplify the approach. The cross-validation results show considerable accuracy compared with the test results provided by organizers. The software submission and evaluation are done using TIRA [12]. We hope to increase the accuracy by adding more bot specific and relevant features in pre-processing in the future.

The profiling data is distributed into train and validation sets. So, to avoid leaving aside a significant portion of the training set for validation and to prevent over-fitting on the training set, we combined the data set and evaluated the accuracy. The proposed model is evaluated using stratified 3-fold, 5-fold, and 10-fold cross-validation during our experiments. Here, Table 1 shows the cross-validation result which helps to choose the best model. We have submitted the TweetNorm+TweetTok+SVM model to the organizers.

Table 1. Cross-Validation results

Methods	3-fold	5-fold	10-fold
TweetTok+Neural Network	80.21	80.65	81.87
TweetNorm+SVM	82.11	81.71	83.15
TweetNorm+TweetTok+SVM	83.29	83.52	84.49

Our model got a test accuracy of 0.7242 for bots vs. humans and 0.4951 for gender, which for the English data set alone [16]. Low Dimensionality Statistical Embedding (LDSE) [13] baseline accuracy given by the organizers using Naive Bayes for bots 0.9054 and gender 0.7800.

6 Conclusion and Future Work

We have developed an approach to classify the bots and humans on Twitter with better accuracy. We have also provided some hints on the existence of spurious accounts on social media and its future possibilities. We hope the results could have been better if we had done the bots vs. human part first and later the gender classification. However, this task of Author Profiling is still a subject of interest as various methodologies can be applied to increase the performance. So, we hope that our approach would spark some ideas in others for further improvements. Further, unsupervised tweet specific embedding, Twitter POS tagging will improve the results on larger data sets. We could further extend to other social media also like Facebook, Quora, Instagram, etc.

References

1. Donald j. trump on twitter: "justice ginsburg of the u.s. supreme court has embarrassed all by making very dumb political statements about me. her mind is shot - resign!". https://bit.ly/2KJWwSa. (Accessed on 06/28/2019).
2. For election day influence, twitter ruled social media - the new york times. https://nyti.ms/2zKTXtp. (Accessed on 06/28/2019).
3. They've got a twitter bot for that | the japan times. https://bit.ly/2FzSFTC. (Accessed on 06/28/2019).
4. Twitter - company. https://about.twitter.com/en_us/company.html. (Accessed on 06/28/2019).
5. Twitter turns six. https://bit.ly/2X3arEE. (Accessed on 06/28/2019).
6. Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P Gummadi. Measuring user influence in twitter: The million follower fallacy. In *fourth international AAAI conference on weblogs and social media*, 2010.
7. Zafar Gilani, Liang Wang, Jon Crowcroft, Mario Almeida, and Reza Farahbakhsh. Stweeler: A framework for twitter bot analysis. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 37–38. International World Wide Web Conferences Steering Committee, 2016.
8. Miha Grčar, Darko Cherepnalkoski, Igor Mozetič, and Petra Kralj Novak. Stance and influence of twitter users regarding the brexit referendum. *Computational social networks*, 4(1):6, 2017.
9. Kyle Griffin. Twitter post: "cspan has posted the entire nearly 8-minute exchange between kamala harris and kavanaugh on the mueller probe. it is worth your time.". http://tiny.cc/8anx8y, septemeber 2018. (Accessed on 06/28/2019).
10. Stefanie Haustein, Timothy D Bowman, Kim Holmberg, Andrew Tsou, Cassidy R Sugimoto, and Vincent Larivière. Tweets as impact indicators: Examining the implications of automated bot accounts on twitter. *Journal of the Association for Information Science and Technology*, 67(1):232–238, 2016.
11. Sara PérezSoler, Esther Guerra, Juan de Lara, and Francisco Jurado. The rise of the (modelling) bots: towards assisted modelling via social networks. In *Proceedings of the 32nd IEEE/ACM International Conference on Automated Software Engineering*, pages 723–728. IEEE Press, 2017.
12. Martin Potthast, Tim Gollub, Matti Wiegmann, and Benno Stein. TIRA Integrated Research Architecture. In Nicola Ferro and Carol Peters, editors, *Information Retrieval Evaluation in a Changing World - Lessons Learned from 20 Years of CLEF*. Springer, 2019.
13. Francisco Rangel, Marc Franco-Salvador, and Paolo Rosso. A low dimensionality representation for language variety identification. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 156–169. Springer, 2016.
14. Francisco Rangel, Paolo Rosso, Manuel Montes-y Gómez, Martin Potthast, and Benno Stein. Overview of the 6th author profiling task at pan 2018: multimodal gender identification in twitter. *Working Notes Papers of the CLEF*, 2018.
15. Francisco Rangel, Paolo Rosso, Martin Potthast, and Benno Stein. Overview of the 5th author profiling task at pan 2017: Gender and language variety identification in twitter. *Working Notes Papers of the CLEF*, 2017.
16. Rosso Paolo Rangel, Francisco. Overview of the 7th author profiling task at pan 2019: Bots and gender profiling. In *Cappellato L., Ferro N., MÃijller H, Losada D. (Eds.) CLEF 2019 Labs and Workshops*. CEUR Workshop Proceedings. CEUR-WS.org, 2019.
17. V Satuluri. Stay in the know [blog post], 2013.

18. T Velayutham and Pradeep Kumar Tiwari. Bot identification: Helping analysts for right data in twitter. In *2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA)(Fall)*, pages 1–5. IEEE, 2017.
19. Vivek Vinayan, JR Naveen, NB Harikrishnan, M Anand Kumar, and KP Soman. Amritanlp@ pan-rusprofiling: Author profiling using machine learning techniques. In *FIRE (Working Notes)*, pages 8–12, 2017.