International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Detection of Spam Bots on Twitter using Machine Learning

Anirudh Sankaranarayanan, Kanshiram U., Gokuladharshan T.P., Suganya T.

Abstract*: Twitter is a popularly used microblogging website that is used to share views, opinions, and updates. However, in recent times, an epidemic of spammer accounts have spread across the website causing disorder and chaos among the normal users. These spammers either aim to promote some commercial agenda or disturb the peace in the online environment. Our project aims to analyze the tweets made by users and predict if they might be spammers so that appropriate action can be taken on them. This is done using machine learning. The random forest algorithm has been modified by giving weighted importance to certain variables assigned using domain knowledge that has been obtained from exploratory analysis of various twitter data sets and knowledge from scientific research papers. A bag of words has also been added to the algorithm, in order to quickly identify the key phrases used by spam bots. By identifying the spammers we can systematically report them and create a more peaceful online environment.

Keywords: twitter; spam; machine learning; classification; random forest; bots; features; social network

I. INTRODUCTION

Users of different age groups are actively engaging on Twitter, a microblogging site. With over 321 million active users, Twitter is regarded as a popular online social network (OSN). OSNs have made the world smaller and reachable. With the help of OSNs, one can easily stay in touch with their family members, friends, and relatives. People can effortlessly join these social networks by providing their details like username, age, gender, phone number, etc. There are a countless number of OSNs available on the internet among which Facebook and Twitter are considered to be the most popular ones. Twitter was founded in 2006 and it enables its users to send and receive posts or messages of up to 280 characters in the form of tweets. By using Twitter, people can express their views, follow their favorite personalities and share multimedia content. When a user follows an account on Twitter, they will be notified whenever the account publishes a status update. This open nature of the OSN is being exploited by spammers and cybercriminals who use bots to carry out disruptive attacks for personal gain.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
Anirudh Sankaranarayanan*, Pursuing,Bachelor's Degree, Sri Krishna College of Technology, Coimbatore, Tamilnadu, India.
Kanshiram U., Pursuing,Bachelor's Degree, Sri Krishna College of Technology, Coimbatore, Tamilnadu, India.
Gokuladharshan T.P., Pursuing,Bachelor's Degree, Sri Krishna College of Technology, Coimbatore, Tamilnadu, India.
Suganya T., Assistant Professor, Department of Computer Science, Engineering, Sri Krishna College of Technology, Coimbatore, TamilNadu

These cybercriminals are using OSNs as a platform to perform complex and advanced attacks such as stalking, cyberbullying, product promotion, phishing, and identity deception. Hijacking, Short URLs, Follower Fraud, and Tweet jacking are some of the methods used by them to execute those attacks. These attacks are evolving every day and are becoming harder to detect and solve. According to a study done by the University of Southern California, 9% to 15% of active Twitter accounts are spambots. These bots impersonate human behavior to gain the support of the victims and then use it to carry out the attacks. These studies indicate how OSNs are becoming a safe haven for these bots. Researchers at the OSNs are actively working to reduce the number of spammers and the spambots present in the system. Although many new approaches and schemes have been proposed to deal with this situation, the spammers are equally updating their way of approach to evade these detection mechanisms and are giving the engineers a hard time. Computer scientists are working toward abolishing cyber threats to make social networks a fun and beneficial activity for the user. As a result, many methods of detecting such threats have been formulated. Yet, spammers are utilizing even better methods to avoid capture. This results in a back and forth motion of progress on either side. On examining the techniques used by spammers in the past and comparing them to those in the present, we observe that the complexity of the technique has increased greatly over time. Earlier spam detection techniques use feature-based strategies that use certain attributes, such as the number of followers or number of tweets to differentiate between actual users and spammers. Since these features are obtained from observing regular user activity this can simply be emulated by the spammer and thus remain undetected. Normal users are usually part of many user ecosystems, thus encouraging frequent communication and participation. On the other hand, spammers only follow random accounts and don't have the same interactions that normal users have. Thus, to avoid capture spammers form groups among themselves at the cost of reducing their intended target quantity.

II. RELATED WORK

Various research papers have been published about spam detection on twitter. The rise in the number of spammers in the past couple of years is a contributing factor to the amount of research done in this topic.

Claudia Meda, et al. [8] presented the first novel approach of a framework that uses an non-uniform feature selection in a Machine Learning system that is used for the classification of users on twitter into spammers or otherwise.

Gerard Biau, et al. [3] provides an excellent refresher course of everything that is required to know about the random forest algorithm. The article also reviews the latest developments and places stress on the mathematical modelling behind the algorithm. It is intended to give amateurs an easy understanding of the main ideas involved.

Sahami et al. [7] gives us a view into the past of spam and how it was an issue even before the rise of conventional social media websites. Email spamming was much more of a threat in that time period, and this paper discusses using the learned Naive Bayes classifier to partition spam emails from real ones. This paper shows how spam has been an ever prevalent problem on the internet which is all the more reason to invest time into researching the issue and stopping spam for good.

Gao et al. analyzes spam attacks on Facebook using similarity graphs and checking the semantic closeness of posts and URLs leading to a common page. Upon further investigation it was discovered by them that the majority of spammers on Facebook were operating using hijacked accounts which were in turn used to hijack more accounts by manipulating the trust of users and goading them to visit more malicious web links.

III. IMPLEMENTATION OF THE PROPOSED CHATBOT

In our study, we propose using a feature specific version of the Random Forest algorithm, which is an ensemble machine learning method used to classify by creating multiple decision trees in the training period and presenting the output generated by the maximum number of trees. It brings together methods such as bagging along with random selection of features. Random forests may be classified as a type of bagging where each model corresponds to a tree. The random tree is created as per CART guidelines [1] but due to feature importance, the algorithm has been slightly configured to better suit the problem.

3.1. Dataset

The goal of these datasets is to test the algorithm against accounts who could be spammers or real users. There were two stages in the process of obtaining the other dataset. Initially, the traffic on twitter had to be procured using an API called Twitter Streaming, and several different user typologies were obtained based on features such as the number of tweets per day, URLs per day, the ratio between followers and user's friends (following). Twitter streamed dataset was also obtained from the WebArchive [13].

The dataset used for this experiment consists of ten primary attributes used for the detection of spammers. In order to prove the effectiveness of the algorithm, two different sets of data were used to test it. The main dataset [12] is commonly used in twitter spam experiments and consists of 1065 users along with 62 attributes. The secondary dataset is from kaggle and consists of 665 users. Three machine learning algorithms including Naive Bayes, Decision Tree and finally Random Forests are contrasted against each other in terms of performance. The datasets are uniform among these tests so as to show a comparative study of how each algorithm performs and display the drawbacks of the inferior algorithms such as Naive Bayes.

3.2. Implementation and Result

Exploratory Data Analysis was performed on the data set in order to better understand the relationships between features. Through various plots and graphs we obtained the relation between Friend vs. Follower count. The spearman coefficient as described in the image (Fig 2.) is obtained, and using this we are easily able to visualize the important features and their relations to classifying a spammer. This helps to improve the accuracy of the already existing algorithm. Online Social Networks are exploited by spammers for various purposes. However, because of the lack of identity, they couldn't gain the trust of normal users. Thus, spam accounts usually have no followers and their following-followers ratio difference would be huge. The spam detection systems could exploit this trait of spam accounts to improve their accuracy and efficiency.

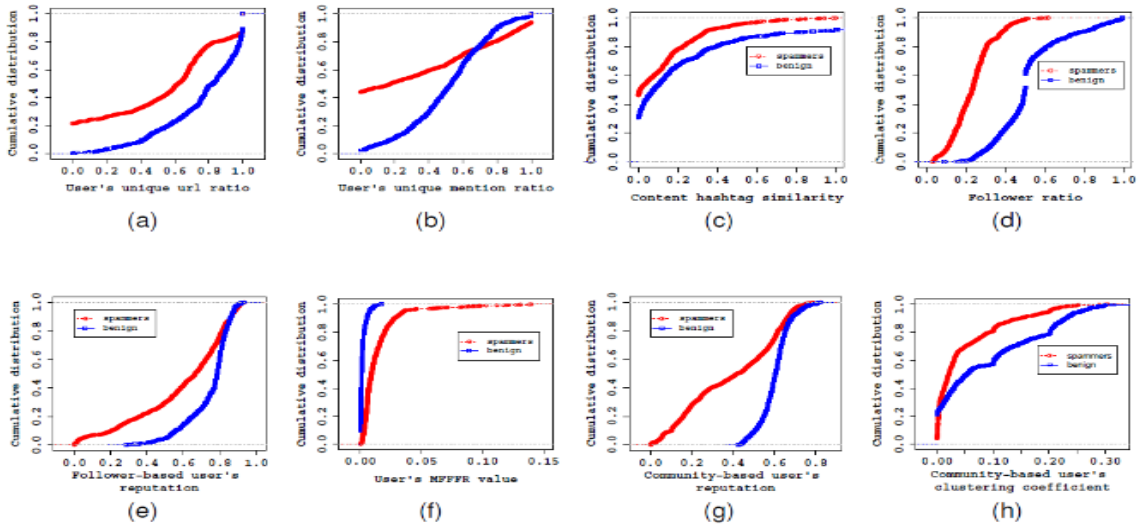

Fig.1 Data set graph

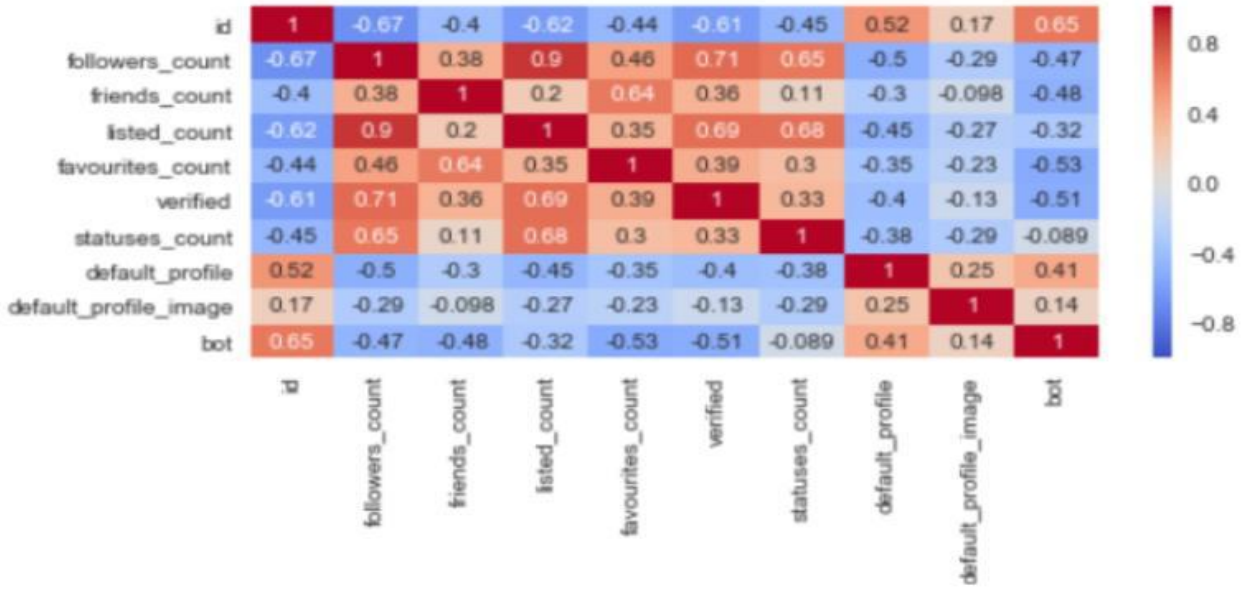

Fig.2 Spearman Correlation Heatmap for feature relations

Our approach is relatively new and it uses an account's interaction network to distinguish spammers, unlike other approaches that rely on profiles for identifying the spam accounts.

Spammers are using new and improved algorithms every day, with which they can easily evade metadata-based features that are related to their accounts. So the OSNs need to make use of interaction-based and community-based features that are harder to evade.

However, achieving absolute precision is tough as hackers frequently change their operating behavior. Hence, it is important to create a model with a complete log of spam accounts.

We obtain results that are conclusive with our research and allow us to achieve a higher accuracy than what is possible with the conventional Random Forest Algorithm.

In both trials the training and test sets do not have any tweet in common.

The results in the test set have been utilized for the measurement of general errors and have not been used for setting up the parameters for the experiment.

The results shown in Table- I consist of multiple runs with different couples of training and test data.

Table I : Performance Comparison of Classifiers

Accuracy	Training Set	Testing Set
Decision Tree	0.8824	0.8785
Naive Bayes	0.5421	0.5631
Random Forest	0.8252	0.7916
Modified Algorithm	0.9646	0.9385

3.3. Future Work

Future objectives of the research are to uncover organized spam attacks by analyzing the spammers' network. Furthermore, new patterns and models can be found by analyzing the ephemeral progression of spammers' followers, which can then be used in characterizing the spammers. All current systems only detect singular spammers among those present in the input.

The ability to uncover networks of spammers would be a much more fruitful technique to eradicate them.

This would also be useful in thwarting coordinated mass spam attacks that happen so often on OSNs.

IV CONCLUSION

The paper introduces a different approach in the Random Forests features sampling process, in a spammer detection context; the proposed framework assigns a different probability value to a subset of features selected by domain experts: different weights assigned to the features allow it to reach significant results during the classification process. Specifically, the approach lets the analyst use a certain sample of features with a low scope of performance reduction in terms of accuracy using probability distribution instead of a complete removal of non relative features. Experimental evidence supports the idea, displaying that an uneven feature selection method obtains a more effective predictor respect to conventional approaches. Finally, the e

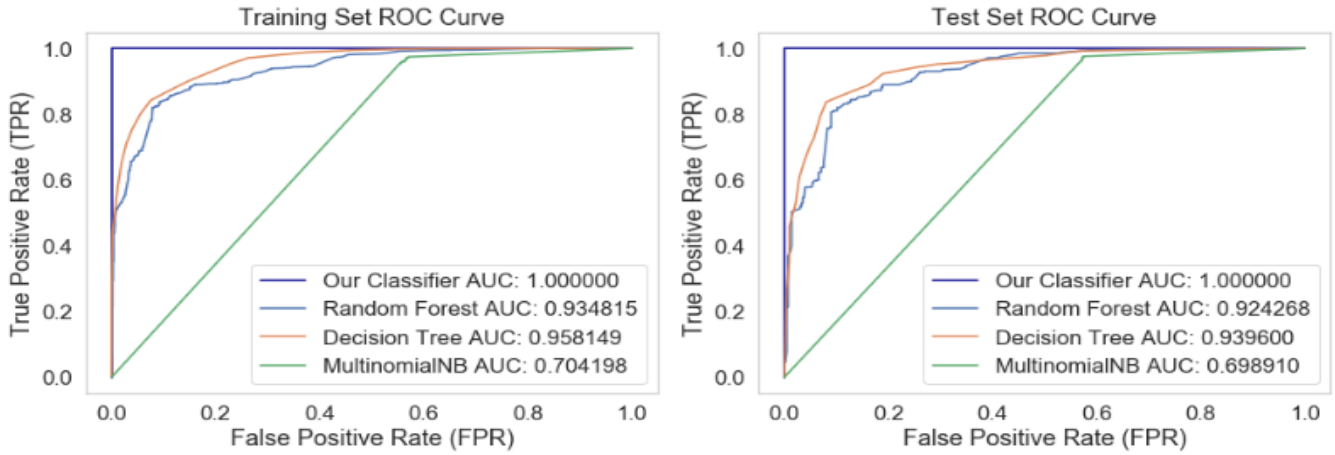

Fig.3 Receiver Operating Characteristic (ROC) and Area under Curve (AUC) Comparison

features is chosen, the performance compares to the usual Random Forest and does not degrade it in any way. To sum up, many classification algorithms such as Naive Bayes Classifier, Decision Tree Classifier and Random Forest Classifier have been tried and tested in this experiment and none of them seem to reach the accuracy level (as demonstrated in the Receiver Operator Characteristic Curve and the area under the curve in Figure 3.) of the proposed algorithm. The area under the curve of the proposed classifier is the highest and hence it gives us a better accuracy in classifying spam bots as compared to the other algorithms. In addition to this we can see the accuracy values for the testing set and training sets that have been tabulated in Table I. It is clear through these evidences that the proposed system is superior to the existing ones.

REFERENCES

1. Hayes T, Usami S, Jacobucci R, McArdle JJ. Using Classification and Regression Trees (CART) and random forests to analyze attrition: Results from two simulations. Psychol Aging. 2015;30(4):911–929. doi:10.1037/pag0000046
2. Chen, Chao, et al. "6 million spam tweets: A large ground truth for timely Twitter spam detection." 2015 IEEE international conference on communications (ICC). IEEE, 2015.
3. Biau, Gérard, and Erwan Scornet. "A Random Forest Guided Tour." arXiv preprint arXiv:1511.05741 (2015).
4. Breiman, Leo. "Random forests."Machine learning 45.1 (2001): 5-32.
5. Chen, C., Zhang, J., Chen, X., Xiang, Y., & Zhou, W. (2015, June). 6 million spam tweets: A large ground truth for timely Twitter spam detection. InCommunications (ICC), 2015 IEEE International Conference on (pp. 7065-7070). IEEE.
6. H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, “Detecting and characterizing social spam campaigns,” inProc. IMC, Melbourne,2001, pp. 35–47.
7. M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, “A Bayesian approach to filtering junk e-mail,” inProc. of Workshop on Learning For Text Categorization, Madison, Wisconsin, 1998, pp. 98–105
8. Meda, C., Bisio, F., Gastaldo, P., & Zunino, R. (2014, October). A machine learning approach for Twitter spammers detection. In Security Technology (ICCST), 2014 International Carnahan Conference on (pp. 1-6). IEEE.
9. Wang, B., Zubiaga, A., Liakata, M., & Procter, R. (2015). Making the most of tweet-inherent features for social spam detection on Twitter. arXiv preprint arXiv:1503.07405.
10. Zhang, X., Li, Z., Zhu, S., & Liang, W. (2016). Detecting Spam and Promoting Campaigns in Twitter. ACM Transactions on the Web (TWEB), 10(1),
11. The Twitter Rules, https://support.twitter.com/articles/18311
12. https://archive.org/details/archiveteam-twitter-stream-2019-05
13. https://www.kaggle.com/c/twitter-spam/data

AUTHORS PROFILE

Anirudh Sankaranarayanan is pursuing his Bachelor’s Degree in Engineering at Sri Krishna College of Technology, Coimbatore, Tamilnadu, India. His research interests include Data mining and Machine Learning. He has a year of experience working as a freelance software developer for online clients. His interest in machine learning was ignited at a young age by reading many articles and novels on the same. It is his goal to someday create a general intelligence neural network.

Gokuladharshan T.P is pursuing his Bachelor’s Degree in Engineering at Sri Krishna College of Technology, Coimbatore, Tamilnadu, India. His research interests include Machine Learning. He has two years of experience in editing and making short films and feature films. He is also a certified Search Engine Optimization Expert. It is his goal someday to direct a Hollywood film produced by Blumhouse Pictures or WB.

Kanshiram U is pursuing his Bachelor’s Degree in Engineering at Sri Krishna College of Technology, Coimbatore, Tamilnadu, India. His research interests include Machine Learning. He is also currently working in Wallins, an IT firm in Coimbatore as a digital media and marketing analyst. He one day hopes to complete his PhD in Computer Science and become a Professor at some well known institution.

Ms.T. Suganya., M.E. obtained her bachelor’s degree and Masters Degree in Computer Science and Engineering from Anna University. She has more than 9 years of teaching experience and currently, she is working as Assistant Professor in Department of Computer Science and Engineering, in Sri Krishna College of Technology, Coimbatore, TamilNadu. Her areas of interest include Pattern analysis and Recognition and Image Processing. She has published 10 papers in reputed international, national level conferences and International journals.