Machine Learning Approach For Spammer Detection on Social Networks

[1] G.Srinivasarao, Assistant Professor, Department of CSE, Chalapathi Institute of Technology, Guntur.
[2] Maraka S V R Siva Swamy, B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.
[3] Maddu Teja Sri Sandhya, B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.
[4] Marriboina Aravind Pavan Kalyan, B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.
[5] Akhil Nandam, B.Tech, Department of CSE, Chalapathi Institute of Technology, Guntur.

Abstract: Social networking sites engage millions of users around the world. The users' interactions with these social sites, such as Twitter have a tremendous impact and occasionally undesirable repercussions for daily life. The prominent social networking sites have turned into a target platform for the spammers to disperse a huge amount of irrelevant and deleterious information. Twitter, for example, has become one of the most extravagantly used platforms of all times and therefore allows an unreasonable amount of spam. Fake users send undesired tweets to users to promote services or websites that not only affect legitimate users but also disrupt resource consumption. Moreover, the possibility of expanding invalid information to users through fake identities has increased that results in the unrolling of harmful content. Recently, the detection of spammers and identification of fake users on Twitter has become a common area of research in contemporary online social Networks (OSNs). In this paper, we perform a review of techniques used for detecting spammers on Twitter. Moreover, a taxonomy of the Twitter spam detection approaches is presented that classifies the techniques based on their ability to detect: (i) fake content, (ii) spam based on URL, (iii) spam in trending topics. The presented techniques are also compared based on various features, such as user features, content features, graph features, structure features, and time features. We are hopeful that the presented study will be a useful resource for researchers to find the highlights of recent developments in Twitter spam detection on a single platform.

1. INTRODUCTION

It has become quite unpretentious to obtain any kind of information from any source across the world by using the Internet. The increased demand of social sites permits users to collect abundant amount of information and data about users. Huge volumes of data available on these sites also draw the attention of fake users [1]. Twitter has rapidly become an online source for acquiring real-time information about users. Twitter is an Online Social Network (OSN) where users can share anything and everything, such as news, opinions, and even their moods. Several arguments can be held over different topics, such as politics, current affairs, and important events. When a user tweets something, it is instantly conveyed to his/her followers, allowing them to outspread the received information at a much broader level [2]. With the evolution of OSNs, the need to study and analyze users' behaviors in online social platforms has intensified. Many people who do not have much information regarding the OSNs can easily be tricked by the fraudsters. There is also a demand to combat and place a control on the people who use OSNs only for advertisements and thus spam other people's accounts. Recently, the detection of spam in social networking sites attracted the attention of researchers. Spam detection is a difficult task in maintaining the security of social networks. It is essential to recognize spams in the OSN sites to save users from various kinds of malicious attacks and to preserve their security and privacy. These hazardous maneuvers adopted by spammers cause massive destruction of the community in the real world. Twitter spammers have various objectives, such as spreading invalid information, fake news, rumors, and spontaneous messages. Spammers achieve their malicious objectives through advertisements and several other means where they support different mailing lists and subsequently dispatch spam messages randomly to broadcast their interests. These activities cause disturbance to the original users who are known as non-spammers. In addition, it also decreases the repute of the OSN platforms. Therefore, it is essential to design a scheme to spot spammers so that corrective efforts can be taken to counter their malicious activities [3]. Several research works have been carried out in the domain of Twitter spam detection. To encompass the existing state-of-the-art, a few surveys have also been carried out on fake user identification from Twitter. Tingmin *et al.* [4] provide a survey of new methods and techniques to identify Twitter spam detection. The above survey presents a comparative study of the current approaches. On the other hand, the authors in [5] conducted a survey on different behaviors exhibited by spammers on Twitter social network. The study also provides a literature review that recognizes the existence of spammers on Twitter social network. Despite all the existing studies, there is still a gap in the existing literature. Therefore, to bridge the gap, we review state-of-the-art in the spammer detection and fake user identification on Twitter. Moreover, this survey presents a taxonomy of the Twitter spam detection approaches and attempts to offer a detailed description of recent developments in the domain. The aim of this paper is to identify different approaches of spam detection on Twitter and to present a taxonomy by classifying these approaches into several categories. For classification, we have identified four means of reporting spammers that can be helpful in identifying fake identities of users. Spammers can be identified based on: (i) fake content, (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. Table 1 provides a comparison of existing techniques and helps users to recognize the significance and effectiveness of the proposed methodologies in addition to providing a comparison of their goals and results. Table 2 compares different features that are used for identifying spam on Twitter. We anticipate that this survey will help readers find diverse information on spammer detection techniques at a single point.

2. LITERATURE SURVEY

1) Detection of online review spam: a literature review

Abstract: Online reviews have become an important resource for customers. It has become a habit for customers to first read a review before deciding to make a purchase. But it can be used by fraudsters to make review spam. This activity can result in the wrong customer purchase decision. Automatic opinion mining methods can also provide inaccurate conclusions due to this activity. This paper aims to provide a literature review on the online review spam detection topic. We identify papers relevant to related topics since 2015, understanding each paper to extract findings, similarities, and research gaps. We find that studies on this topic can be categorized into three focus groups. Focus on review spam detection methods, studies on individuals who write review spam, and studies that examine the spammer groups. Each focus of research has its strengths and weaknesses method which provide benefits in the field of review spam detection.

2) A review on social spam detection: Challenges, open issues, and future directions

Abstract: Online Social Networks are perpetually evolving and used in plenteous applications such as content sharing, chatting, making friends/followers, customer engagements, commercials, product reviews/promotions, online games, and news, etc. The substantial issues related to the colossal flood of social spam in social media are polarizing sentiments, impacting users' online interaction time, degrading available information quality, network bandwidth, computing power, and speed. Simultaneously, groups of coordinated automated accounts/bots often use social networking sites to spread spam, rumors, bogus reviews, and fake news for targeted users or mass communication. The latest developments in the form of artificial intelligence-enabled Deepfakes have exacerbated these issues at large. Consequently, it becomes extremely relevant to review recent work concerning social spam and spammer detection to counter this issue and its effect. This paper provides a brief introduction to social spam, the spamming process, and social spam taxonomy. The comprehensive review entails several dimensionality reduction techniques used for feature selection/extraction, features used, various machine learning and deep learning techniques used for social spam and spammer detection, and their merits and demerits. Artificial intelligence and deep learning empowered Deepfake (text, image, and video) spam, and their countermeasures are also explored. Furthermore, meticulous discussions, existing challenges, and emerging issues such as robustness of detection systems, scalability, real-time datasets, evade strategies used by spammers, coordinated inauthentic behavior, and adversarial attacks on machine learning-based spam detectors, etc., have been discussed with possible directions for future research.

3) An integrated approach for malicious tweets detection using NLP.

Abstract: Many previous works have focused on detection of malicious user accounts. Detecting spams or spammers on Twitter has become a recent area of research in social network. However, we present a method based on two new aspects: the identification of spamtweets without knowing previous background of the user; and the other based on analysis of language for detecting spam on twitter in such topics that are in trending at that time. Trending topics are the topics of discussion that are popular at that time. This growing micro blogging phenomenon therefore benefits spammers. Our work tries to detect spam tweets in based on language tools. We first collected the tweets related to many trending topics, labelling them on the basis of their content which is either malicious or safe. After a labelling process we extracted a many features based on the language models using language as a tool. We also evaluate the performance and classify tweets as spam or not spam. Thus our system can be applied for detecting spam on Twitter, focusing mainly on analysing of tweets instead of the user accounts

4) Twitter spam detection: Survey of new approaches and comparative study.

Abstract: Twitter spam has long been a critical but difficult problem to be addressed. So far, researchers have proposed many detection and defence methods in order to protect Twitter users from spamming activities. Particularly in the last three years, many innovative methods have been developed, which have greatly improved the detection accuracy and efficiency compared to those which were proposed three years ago. Therefore, we are motivated to work out a new survey about Twitter spam detection techniques. This survey includes three parts: 1) A literature review on the state-of-art: this part provides detailed analysis (e.g. taxonomies and biases on feature selection) and discussion (e.g. pros and cons on each typical method); 2) Comparative studies: we will compare the performance of various typical methods on a universal testbed (i.e. same datasets and ground truths) to provide a quantitative understanding of current methods; 3) Open issues: the final part is to summarise the unsolved challenges in current Twitter spam detection techniques. Solutions to these open issues are of great significance to both academia and industries. Readers of this survey may include those who do or do not have expertise in this area and those who are looking for deep understanding of this field in order to develop new methods.

3. EXISTING SYSTEM

In the field of Twitter spam detection, several studies have been conducted. A few polls on false user identification from Twitter were also conducted to cover the current state-of-the-art. Present a review of new methodologies and techniques for detecting Twitter spam. The survey above provides a comparative analysis of existing techniques. The authors of conducted a survey on the various behaviours displayed by spammers on the Twitter social network. The research also includes a literature analysis that acknowledges the existence of spammers on Twitter. Despite all of the studies that have been done, there is still a void in the literature. As a result, we examine the state-of-the-art in spammer detection and fake user identification on Twitter in order to close the gap. Furthermore, this study gives taxonomy of Twitter spam detection methods and strives to provide a comprehensive overview of current developments in the field.

4. PROPOSED SYSTEM

The goal of this work is to discover several ways to spam detection on Twitter and to offer a taxonomy that categorises these approaches into different groups. For the purposes of classification, we've identified four methods for reporting spammers that can assist in detecting user impersonation. Spammers can be detected using the following methods: I false content, (ii) URLbased spam detection, (iii) spam detection in popular subjects, and (iv) fake user identification. Table 1 compares existing procedures and aids users in recognising the significance and effectiveness of the proposed methodology, as well as comparing their goals and outcomes. Table 2 examines the many features used to identify spam on Twitter. We hope that by conducting this poll, readers will be able to find a wealth of information on spammer detection strategies in one place. The taxonomy for spammer detection approaches on Twitter is presented in Section II of this article. In Section III, we compare and contrast various strategies for detecting spammers on Twitter. Section IV contains an overview analysis and debate, while Section V brings the paper to a close and suggests some future research topics.

SYSTEM ARCHITECTURE

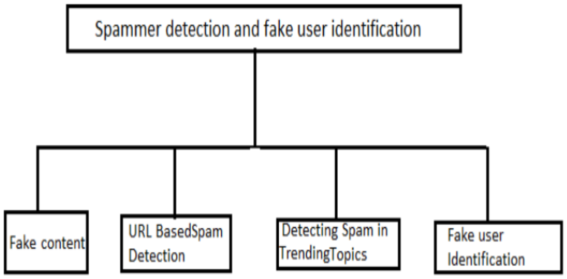

Spammer detection and fake user identification
Fake content
URL BasedSpam Detection
Detecting Spam in TrendingTopics
Fake user Identification

Fig 1: System Architecture

5. ALGORITHMS

5.1 DECISION TREE CLASSIFIERS

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2… Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

5.2 SVM

In classification tasks a discriminate machine learning technique aims at finding, based on an independent and identically distributed (iid) training dataset, a discriminate function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminate classification function takes a data point x and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminate approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space. SVM is a discriminate technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyper plane parameter—in contrast to genetic algorithms (GAs) or perceptions, both of which are widely used for classification in machine learning. For perceptions, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perception and GA classifier models are different each time training is initialized. The aim of GAs and perceptions is only to minimize error during training, which will translate into several hyper planes' meeting this requirement.

5.3 RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of over fitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance. The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg. An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance. Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

6. RESULTS

6.1 Output Screens

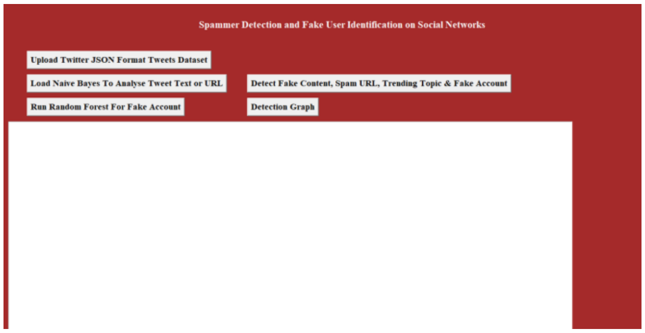
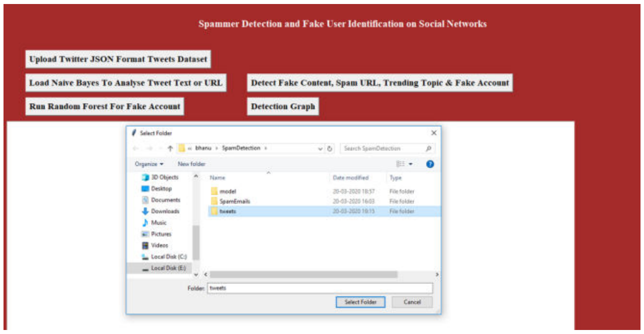

Spammer Detection and Fake User Identification on Social Networks
Upload Twitter JSON Format Tweets Dataset
Load Naive Bayes To Analyse Tweet Text or URL
Detect Fake Content, Spam URL, Trending Topic & Fake Account
Run Random Forest For Fake Account
Detection Graph

Fig 6.1 Upload dataset

In above screen click on ‘Upload Twitter JSON Format Tweets Dataset’ button and upload tweets folder

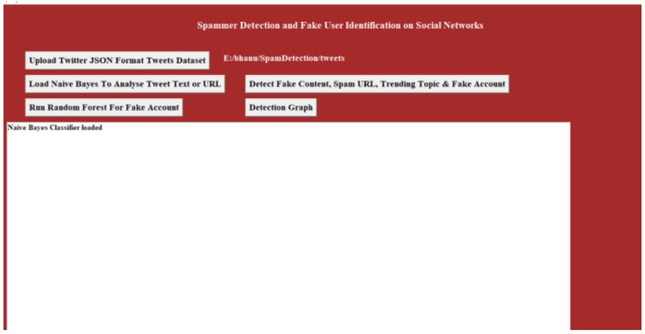

Spammer Detection and Fake User Identification on Social Networks
Upload Twitter JSON Format Tweets Dataset
Load Naive Bayes To Analyse Tweet Text or URL
Detect Fake Content, Spam URL, Trending Topic & Fake Account
Run Random Forest For Fake Account
Detection Graph
Select Folder
« bhanu › SpamDetection ›
Search SpamDetection
Organize ▾
New folder
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Local Disk (C:)
Local Disk (E:)
Name
Date modified
Type
Select Folder
Cancel

Name	Date modified	Type
model	20-03-2020 18:57	File folder
SpamEmails	20-03-2020 16:03	File folder
tweets	20-03-2020 18:15	File folder

Folder: tweets

Fig 6.2 Uploading tweets data

In above screen I am uploading ‘tweets’ folder which contains tweets from various users in JSON format. Now click open button to start reading tweets

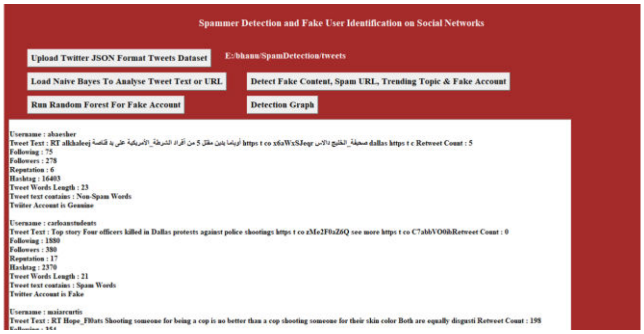

Spammer Detection and Fake User Identification on Social Networks
Upload Twitter JSON Format Tweets Dataset
Load Naive Bayes To Analyse Tweet Text or URL
Detect Fake Content, Spam URL, Trending Topic & Fake Account
Run Random Forest For Fake Account
Detection Graph

E:/bhanu/SpamDetection/tweets

```
Naive Bayes Classifier loaded
```

Fig 6.3 Run Naivy Bayes Algorithm

In above screen naïve bayes classifier loaded and now click on ‘Detect Fake Content, Spam URL, Trending Topic & Fake Account’ to analyse each tweet for fake content, spam URL and fake account using Naïve Bayes classifier and other above mention technique


Spammer Detection and Fake User Identification on Social Networks
Upload Twitter JSON Format Tweets Dataset
Load Naive Bayes To Analyse Tweet Text or URL
Detect Fake Content, Spam URL, Trending Topic & Fake Account
Run Random Forest For Fake Account
Detection Graph

E:/bhanu/SpamDetection/tweets

```
Username : abaesher
Tweet Text : RT alkhaleej أوباما يابين مقتل 5 من أفراد الشرطة_الأمريكية على يد قناصة https t co x6aWxSJeqr دالاس صحيفة_الخليج dallas https t c Retweet Count : 5
Following : 75
Followers : 278
Reputation : 6
Hashtag : 16403
Tweet Words Length : 23
Tweet text contains : Non-Spam Words
Twiiter Account is Genuine

Username : carloanstudents
Tweet Text : Top story Four officers killed in Dallas protests against police shootings https t co zMe2F0aZ6Q see more https t co C7abbVO0ibRetweet Count : 0
Following : 1880
Followers : 380
Reputation : 17
Hashtag : 2370
Tweet Words Length : 21
Tweet text contains : Spam Words
Twitter Account is Fake

Username : maiarcurtis
Tweet Text : RT Hope_F0ats Shooting someone for being a cop is no better than a cop shooting someone for their skin color Both are equally disgusti Retweet Count : 198
Following : 354
```

Fig 6.4 Spammer detection result

In above screen all features extracted from tweets dataset and then analyze those features to identify tweets is no spam or spam. In above text area each records values are separated with empty line and each tweet record display values as TWEET TEXT, FOLLOWERS, FOLLOWING etc with account is fake or genuine and tweet text contains spam or non-spam words. Now click on ‘Run Random Forest Prediction’ button to train random forest classifier with extracted tweets features and this random forest classifier model will be used to predict/detect fake or spam account for upcoming future tweets. Scroll down above text area to view details of each tweet

7. CONCLUSION

Here the paper is a implementation of analysis method utilized on behalf of distinguishing spammers on Twitter. We additionally exhibited taxonomy of Twitter spam identification method are considered as false contented recognition, URL built spam identification, spam location in inclining points, and phony client recognition strategies. We likewise analyzed the introduced strategies dependent on a few features, for example, client features, content features, chart features, structure features, and time features. Besides, the procedures were likewise looked at regarding their predefined objectives and datasets utilized. It is foreseen that the introduced audit will assist scientists with finding the data on best in class Twitter spam discovery procedures in a united structure. Notwithstanding the improvement of proficient and viable methodologies for the spam discovery and phony client distinguishing proof on Twitter, there are as yet certain open zones that need extensive consideration by the analysts.

8. REFERENCES

1. B. Erçahin, Ö. Aktaş, D. Kilinç, and C. Akyol, ‘‘Twitter fake account detection,’’ in Proc. Int. Conf. Comput. Sci. Eng.(UBMK), Oct. 2017, pp. 388–392.

2. F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, ‘‘Detecting spammersonTwitter,’’ in Proc. Collaboration,Electron.Messaging,AntiAbuse Spam Conf. (CEAS), vol. 6, Jul. 2010, p. 12.

3 Kiran Kumar Kommineni, Ratna Babu Pilli, K. Tejaswi, P. Venkata Siva, Attention-based Bayesian inferential imagery captioning maker, Materials Today: Proceedings, 2023, , ISSN 2214-7853, https:// doi.org/ 10.1016/ j.matpr. 2023.05.231.

4. K. K. . Kommineni and A. . Prasad, “A Review on Privacy and Security Improvement Mechanisms in MANETs”, Int J Intell Syst Appl Eng, vol. 12, no. 2, pp. 90–99, Dec. 2023.

5. S. J. Soman, ‘‘A survey on behaviors exhibited by spammers in popular social media networks,’’ in Proc. Int. Conf. Circuit, Power Comput. Technol. (ICCPCT), Mar. 2016, pp. 1–6.

6. A. Gupta, H. Lamba, and P. Kumaraguru, ‘‘1.00 per RT #BostonMarathon # prayforboston: Analyzing fake content on Twitter,’’ in Proc. eCrime Researchers Summit (eCRS), 2013, pp. 1–12.

7. F. Concone, A. De Paola, G. Lo Re, and M. Morana, ‘‘Twitter analysis for real-time malware discovery,’’ in Proc. AEIT Int. Annu. Conf., Sep. 2017, pp. 1–6.

8. N. Eshraqi, M. Jalali, and M. H. Moattar, ‘‘Detecting spam tweets in Twitter using a data stream clustering algorithm,’’ in Proc. Int. Congr. Technol., Commun. Knowl. (ICTCK), Nov. 2015, pp. 347–351.

9. C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, ‘‘Statistical features-based real-time detection of drifted Twitter spam,’’ IEEE Trans. Inf. Forensics Security, vol. 12, no. 4, pp. 914–925, Apr. 2017.

10.C.BuntainandJ.Golbeck,‘‘Automaticallyidentifyingfake newsinpopular Twitter threads,’’ in Proc. IEEE Int. Conf. Smart Cloud (SmartCloud), Nov. 2017, pp. 208–215.

11. K. K. Kumar, S. G. B. Kumar, S. G. R. Rao and S. S. J. Sydulu, "Safe and high secured ranked keyword searchover an outsourced cloud data," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 2017, pp. 20-25, doi: 10.1109/ICICI.2017.8365348.