

An Efficient Method for Spammer and Fake User detection on Social Networks

CHENNURI NAGENDRA SAI, Department of CSE, Siddhartha Institute of Technology & Sciences, Telangana

Dr. R. Dinesh Kumar, Professor, Department of CSE, Siddhartha Institute of Technology & Sciences, Telangana

Mallereddy Sowjanya Reddy, Assistant Professor, Dept of CSE, Siddhartha Institute of Technology and Sciences, Telangana

Abstract - Social networking sites engage millions of users around the world. The users' interactions with these social sites, such as Twitter and Facebook 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 those 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, and (iv) fake users. 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 intensity. 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 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 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

A Comprehensive Survey of Spam Profile Detection Methods in Online Social Networks

Social networks have grown into a popular way for internet surfers to interact with friends in addition to family members, reading news, and also discuss events. Users spend more time on popular social platforms (e.g., Facebook, Twitter, etc.) storing and sharing their personal information. This fact together with the prospect of communicating thousands of users fascinates the concentration of malicious users. They exploit the implicit trust interactions concerning users with the purpose of accomplishing their malicious objectives, for instance, create malicious links inside the posts/tweets, spread fake news, send out unsolicited messages to genuine users and so on. In this paper, we reviewed various existing techniques on spam profile detection in online social networks.

Fake Content and Fake User Identification on Social Networks

Millions of users are engaged with social networking sites around the world. Social sites like twitter, Facebook have a large impact on rare unwanted consequences caused in our regular life in user's interactions. In order to disperse a large amount of inappropriate and harmful data protruding social networking sites are made as a target platform for the spammers. Twitter is main example that has become one of the important platforms for unreasonable amount of spam in all the tomes for fake users to tweet and promote websites or services that crates a major effect for legitimate users and also it disturbs resource consumption. By resulting the opening for unusual and harmful information there is an increase of fake identities that expands invalid data. Research on current online social networks (OSN) is quit natural for identifying of spammers and also detection of fake users on twitter. This paper is a review paper that tells about detecting spammer techniques on twitter. Depending on the ability detection taxonomy of twitter spam identification methods are classified and presented as 1. Fake content 2.URL based on spam 3.trending topics in spam 4.fake users.

Identification of Spammers and Fake Users on Social Networks

A great many users are engaged with social networking sites the world over. Social sites like Twitter, Facebook have a large effect on rare unwanted consequences caused in our regular life in user's interactions. To disperse a large measure of inappropriate and destructive information distending social networking sites are made as a target stage for the spammers. Twitter is the fundamental example that has become one of the significant stages for an unreasonable measure of spam in all the tomes for fake users to tweet and promote websites or services that create a significant effect for legitimate users and furthermore it upsets resources utilization. By resulting in the opening for surprising and unsafe

data there is an increase of fake identities that expands invalid information. Research on current online social networks (OSN) is quite normal for identifying spammers and furthermore detection of fake users on twitter. Recently, the detection of spammers and the identification of fake users on Twitter has become a typical area of research in contemporary online social networks (OSNs). In this paper, we review the 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 capacity to detect: (I) fake content, (ii) spam based on URL, (iii) spam in trending points, and (iv) fake users. The presented techniques are likewise compared based on different features, for example, user features, content features, chart features, structure features, and time features. We are hopeful that the presented investigation will be a useful resource for researchers to and the features of recent developments in Twitter spam detection on a single stage.

3. SYSTEM ANALYSIS:

3.1 Existing System

- ❖ Shen investigated issues of detecting spammers on Twitter. The proposed method combines characteristics withdrawal from text content and information of social networks. The authors used matrix factorization to determine the underline feature matrix or the tweets and then came up with a social regularization with interaction coefficient to teach the factorization of the underline matrix. Subsequently, the authors combined knowledge with social regularization and factorization matrix processes, and performed experiments on the real-world Twitter dataset, i.e., UDI Twitter dataset.
- ❖ Washha described the Hidden Markov Model for filtering the spam related to recent time. The method supports the accessible and obtainable information in the tweet object to

recognize spam tweets and the tweets that are handled previously related to the same topic.

- ❖ Jeong analyzed the follow spam on Twitter as an alternative of dispersion of provoking public messages; spammers follow authorized users, and followed by authorized users. Categorization techniques were proposed that are used for the detection of follow spammers. The focus of the social relation is cascaded and formulated into two mechanism, i.e., social status filtering and trade significance
- ❖ Profile filtering, where each of which uses two-hop sub networks that are centered at each other. Assemble techniques and cascading filtering are also proposed for combining the properties of both trade significance profile and social status. To check whether a user is fake or not, a two-hop social network for each user is focused to gather social information from social networks.
- ❖ Meda presented a technique that utilizes a sampling of non-uniform features inside a machine learning system by the adaptation of random forest algorithm to recognize spammer insiders. The proposed framework focuses on the random forest and non-uniform feature sampling techniques. The random forest is a learning algorithm for the categorization and regression that works by assembling several decision trees at preparation time and selecting the one with the majority votes by individual trees. The scheme integrates bootstrap aggregating technique with the un-planned selection of features.

Disadvantages

There is no filtering system based on a preprocessing schedule and on Naïve Bayes algorithm to discard the tweets containing inaccurate information, Less security due No URL Based Spam Detection

3.2 Proposed System

- ❖ In the proposed system, the system elaborates a classification of spammer detection techniques. The system shows the proposed taxonomy for identification of spammers on Twitter. The proposed taxonomy is categorized into four main classes, namely, (i) fake content; (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. Each category of identification methods relies on a specific model, technique, and detection algorithm.
- ❖ The first category (fake content) includes various techniques, such as regression prediction model, malware alerting system, and Lfun scheme approach. In the second category (URL based spam detection), the spammer is identified in URL through different machine learning algorithms. The third category (spam in trending topics) is identified through Naïve Bayes classifier and language model divergence. The last category (fake user identification) is based on detecting fake users through hybrid techniques.

Advantages

- The average numbers of verified accounts that were either spam or non-spam and (ii) the number of followers of the user accounts.
- The fake content propagation was identified through the metrics that include: (i) social reputation, (ii) global engagement, (iii) topic engagement, (iv) likability, and (v) credibility. After that, the authors utilized regression prediction model to ensure the overall impact of people who spread the fake content at that time and also to predict the fake content growth in future.

MODULES:

Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as View and Authorize Users, Add and View Spam Filters, View All User Posted Tweets, View All User Tweets Based On URLs, View Friend Request and Response, View All Tweets with Re-Tweets, View All Tweets, Re-Tweets and Comments, View All Spammers Detection, View All Fake User Identification, View Fake User Identification Results, View Fake Tweet Identification Results

- **User**

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he has to wait for admin to authorize him and after admin authorized him. He can login by using authorized user name and password. Login successful he will do some operations like My Profile, Search Friends ,Create Tweets, View My Friends, View Friend Requests, Search Tweets and Comment ,View My Tweets and Comments, View Friend's Retweets and Give Comments.

4. OUTPUT RESULTS:



Fig 4.3: All Reviewed Tweets Page

5. CONCLUSION

In this paper, we performed a review of techniques used for detecting spammers on Twitter. In addition, we also presented taxonomy of Twitter spam detection approaches and categorized them as fake content detection, URL based spam detection, spam detection in trending topics, and fake user detection techniques. We also compared the presented techniques based on several features, such as user features, content features, graph features, structure features, and time features. Moreover, the techniques were also compared in terms of their specified goals and datasets used. It is anticipated that the presented review will help researchers find the information on state-of-the-art Twitter spam detection techniques in a consolidated form.

Despite the development of efficient and effective approaches for the spam detection and fake user identification on Twitter, there are still certain open areas that require considerable attention by the researchers. The issues are briery highlighted as under: False news identification on social media networks is an issue that needs to be explored because of the serious repercussions of such news at individual as well as collective level. Another associated topic that is worth investigating is the identification of rumor sources on social media. Although a few studies based on statistical methods have already been conducted to detect the sources of rumors, more sophisticated approaches, e.g., social network based approaches, can be applied because of their proven effectiveness.



Fig 4.1: Home Page



Fig 4.2: View All Tweets Page

REFERENCES

- [1] B. Erçahin, Ö. Akta³, D. Kiliñç, and C. Akyol, "Twitter fake account detection," in Proc. Int. Conf. Comput. Sci. Eng. (UBMK), Oct. 2017, pp. 388392.
- [2] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on Twitter," in Proc. Collaboration, Electron. Messaging, Anti- Abuse Spam Conf. (CEAS), vol. 6, Jul. 2010, p. 12.
- [3] S. Gharge, and M. Chavan, "An integrated approach for malicious tweets detection using NLP," in Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT), Mar. 2017, pp. 435438.
- [4] T. Wu, S. Wen, Y. Xiang, and W. Zhou, "Twitter spam detection: Survey of new approaches and comparative study," *Comput. Secur.*, vol. 76, pp. 265284, Jul. 2018.
- [5] S. J. Soman, "A survey on behaviors exhibited by spammers in popular social media networks," in Proc. Int. Conf. Circuit, Power Comput. Tech-nol. (ICCPCT), Mar. 2016, pp. 16.
- [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. 112.
- [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. 16.
- [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. 347351.
- [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. 914925, Apr. 2017.
- [10] C. Buntain and J. Golbeck, "Automatically identifying fake news in popular Twitter threads," in Proc. IEEE Int. Conf. Smart Cloud (SmartCloud), Nov. 2017, pp. 208215.
- [11] C. Chen, J. Zhang, Y. Xie, Y. Xiang, W. Zhou, M. M. Hassan, A. AlElaiwi, and M. Alrubaian, "A performance evaluation of machine learning-based streaming spam tweets detection," *IEEE Trans. Comput. Social Syst.*, vol. 2, no. 3, pp. 6576, Sep. 2015.
- [12] R. Dineshkumar, J. Suganthi, "Sanskrit character recognition system using neural network," *Indian journal of science and technology*, Vol.8 (1), 2015.

