A Survey on Behaviour Based Spam and Fake user Detection on Social Network

Chandan. M. B\textsuperscript{1}, Dr. Annapurna. V. K\textsuperscript{2}
M.Tech Student (Information Technology)\textsuperscript{1}, Professor\textsuperscript{2}
Department of CS&E
The National Institute of Engineering, Mysuru, Karnataka, India

Abstract:
Social networking sites such as Twitter and Facebook attract millions of users across the world and their interaction with social networking has affected their life. Dependence on platforms for seeking opinions, news, updates, etc. is increasing. While it is true that Online Social networks (OSNs) have become a new medium for disseminate information, at the same time, they are also fast becoming a playground for the spread of misinformation, propaganda, fake news, rumors, unsolicited messages, etc. In this approach, first represent each user of the social network with a unique ID (UID) that track’s the user behavior and interactions with other participants. Next, based on the estimated user’s behavior, we propose a statistical framework that uses the Naïve Bayes algorithm in order to identify spammers. The proposed approach is able to automatically discriminate between spammers and legitimate users, while existing unsupervised approaches require human intervention in order to set informal threshold parameters to detect spammers.

Keywords: online social Network (OSN), Application Programming Interface (API).

I. INTRODUCTION

Social networking sites now-a-days have become the most popular medium for people to spend their time and make more friends. Online Social networks (OSNs), such as Twitter, Facebook, and many enterprise social networks, have become very much popular in the last few years. People are spending a huge amount of time on OSNs making friends, with people whom they know or with people who are interested in. Twitter, which came into existence in 2006, became one of the most popular site. Now-a-days, over 300 million Twitter users create around 600 million new micro blogs which are called as “tweets” per day. Twitter is an Online Social Network (OSN) where users can share everything and anything, 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. 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 tricked by the fraudsters. Hence, there is a need to detect spam’s, which contain URLs and are redirecting people to different locations. The work in this paper deals with detecting such spam tweets and many new features based on language models that help to improve spam detection.

II. RELATED WORK

Tingmin et al\cite{1}. Provide a survey on 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 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 and despite all the existing works there is still a gap in the existing literature. Therefore, to bridge the gap, they 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. Concone et al\cite{2}, presented a methodology that provides malignant alerting by using a specified set of tweets in real-time conquered through the Twitter API. Afterwards the batch of tweets considering the same topic is sum up to generate an alert. The proposed architecture is used to evaluate Twitter posting, recognizing the advancement of admissible event, and reporting of that event. The proposed approach utilizes the information contained in the tweets when a spam or malware is recognized by the users or the report of security has been released by the certified authorities. Eshraqi et al\cite{3}, determined different features to detect the spam and then with the help of a den stream based clustering algorithm, recognize the spam tweets. Some user accounts were selected from various datasets and afterwards random tweets were selected from these accounts. The tweets are subsequently categorized as spam and non-spam. The authors claimed that the algorithm can divide the data into spam and non-spam with high accuracy and regret tweets maybe recognized with high accuracy and precision. Buntain et al\cite{4}, introduced a method for detecting fake news on Twitter automatically by predicting accurate assessment in two credibility-focused datasets. The method was applied on the Twitter fake news dataset and the model was trained against a crowd sourced worker based on the assessment of journalists. The two Twitter datasets were used to study the integrity in OSNs. Chen et al \cite{5}, performed an evaluation of machine learning algorithms to detect spam tweets. The authors analyzed the impact of various features on the performance of spam detection factor. Stafford et al \cite{6}, examined the degree to which the trending affairs in Twitter are exploited by spammers. Although numerous methods to detect...
the spam have been proposed, the research on determining the effects of spam on Twitter trending topics has attained only limited attention of the researchers. The authors presented a technique to cooperate with Twitter public Application programming interface (API). Mateen et al. [7]. Proposed a hybrid technique that utilizes user-based, content-based, and graph-based characteristics for spammer profiles detection. A model is proposed to differentiate between the non-spam and spam profiles using three characteristics. Gupta et al. [8]. Present a policy for the detection of spammers in Twitter and use the popular techniques, i.e., Naïve Bayes clustering and decision trees. The algorithm Classify an account as spam or non-spam. The dataset comprises 1064 Twitter users that contain 62 features, which are either user-specific or tweet-specific information. The spammer account contains almost 36% of the used dataset. As the behavior of spammers is different from non-spammers, some attributes or features are recognized in which both categories are different from one another. Feature identification is based on the number of features at user and tweet level such as followers or following, spam keywords, replies, hash tags, and URLs. Chauhan et al. [9]. Proposed a methodology for the detection of anomalous tweets. The type of abnormality that is distributed on Twitter is the type of URL anomaly. Anomalous users use various URL links for creating spams. The proposed methodology, which is used to identify various anomalous activities from social networking sites.

III. OUTCOME

1. Possible Solution

The prominent social networking sites have turned into a target platform for the spammers to disperse a huge amount of irrelevant and deleterious information. In this paper the techniques used for detecting spammers on Twitter are proposed. In addition, taxonomy for spam detection is categorized as fake content detection, URL based spam detection, spam detection in trending topics, and fake user detection techniques. The presented review will help researchers and the information on state-of-the-art spam detection techniques in a consolidated form.

a) The first major categorization in the taxonomy using the techniques such as regression prediction model, malware alerting system, and Lfun scheme approach proposed for detecting spam, which is injected in the Twitter platform.

b) The second categorization considers the technique through different machine learning algorithms for spam detection based on URLs.

c) The third category is identified through Naïve Bayes classifier and language model divergence in the proposed taxonomy contains approaches meant for spam identification from trending topics on Twitter. Various features for identifying spammers in trending topics have been classified with a variety of attributes.

d) The fourth category in the taxonomy is using based on detecting fake users through hybrid techniques for the identification of fake users to detect spams on Twitter.

In addition to the review techniques, the study also provides the comparison of miscellaneous spam detection features. These features are extracted from user accounts and that can help to identify spams. These features are categorized in to classes, namely user, content, graph, structure, and time.

- The user-based features incorporate the number of following and followers, account, age, reputation, and number of tweets.
- The content-based features contain number of retweets, number of URLs, number of replies and number of characters and digits, and spam words.
- The graph-based features include in/out degree and between centrality.
- Structure-based features include average tweet length, thread life time (number of times between first and last tweets), and tweet frequency.
- Time-based features include idle time in days and tweet sent in specifically time interval.

Therefore, the survey is assembled by the classes that are categorized according to different features that are used for analyzing and detecting spams in various groups.

IV. CONCLUSION

The fake user and spam on twitter is compared and analyzed by different classification algorithm, the paper presented a new possible methodology for detecting spam tweets from twitter which was different from the previous works which only focused on detecting spam accounts is introduced. Even after applying different technique to analyzing a fake user and spam. There is a scope for further work in this area.

V. REFERENCES


