A Survey on Features and Techniques of Blog Spammer Identification

Rupali Dohare[1], Dr. Pratima Gautam[2]
[1]Research Scholar, Rabindranath Tagore University, Bhopal (M.P.) India.
[2]Dean of CS/IT, Rabindranath Tagore University, Bhopal (M.P.) India.

ABSTRACT
Blogging becomes a popular way for a Web user to publish information on the Web. Bloggers write blog posts, share their likes and dislikes, voice their opinions. Activities happened in Blogosphere affect the external world. This attract many promoters hire some bloggers who post to increase heights of those brands or products. So spamming is a major problem in internet-based things as well as in social media. Different techniques have been proposed for spam filtering have been exposed across various platforms with varies degree of measures. This survey focused on some of the present strategies used for filtering social spam. Starting with different types of social spam, the paper has discussed about recent developments in the field of elimination of social spam. This paper gives a concise study of methods proposed by different researchers. Here various features of spammer profile identification were also done with a comprehensive and comparative understanding of existing literature.

Index Terms— Content filtering, Fake Profile, Online Social Networks, Spam Detection.

I INTRODUCTION

During the recent few years social media has evolved many folds and has become much more interactive and integral part of our lives. The interaction channels in the social media have changed from traditional media like newspapers and television to mobile phones, social media websites, micro blogging sites etc. It has changed the way people communicate with each other on the personal as well as on public from as described in [1]. There are varieties of social media sites that offer diverse functionality, some are for common people like Facebook, which started as an experimental social network in the Harvard University by some students, while others like LinkedIn is a network formed by professionals from every field. Many sites are exclusively for sharing videos and pictures media like YouTube, Instagram, Flickr etc. while others focused on blogs where people from varied domains express and share their views. There are even social tagging and news sites like Reddit, Delicious etc. which allow the user to rank the websites on the basis of quality of content and usefulness of the sites. Most recent trend of micro – blogging let people update the real – time status of their daily routine or happenings via app like Twitter which has more than 200 million users exchanging more than 400 million tweets per day [2] where the length of tweets is limited to 140 characters.

According to Teen, Social Media and Technology Overview 2015 [3], ―More than 24% of the teen are constantly online and 71% of them use more than one social networking site‖. This ease of sending and receiving data over Internet has resulted in some notorious people sending unwanted messages to large number of recipients over the network trying to take advantage by getting access to their privacy. Initial spread of spams started with email spam. According to M3AAWG report, the abusive email content amounts to 87.1% - 90.2% of the total email content during 2012 – 2014 [4] which has increased the financial burden by increasing the storage requirement and technological requirement for spam detection. Slowly spams started spreading in every digital media like from mobile network through mobile phone, social networking sites, blogs, review sites etc.

The rest of this paper is organized as follows: in the second section, the requirement of blog spamming was discussed. Third section list various techniques adopt by spammer to promote their target website, brand, product, etc. While fourth section provide related work of the current approaches applied by different researchers to identify blog spammers. Research problem is pointed out, and then the proposed problem is formalized in detail. The conclusion of the whole paper is made in the last section.

II REQUIREMENT OF BLOG SPAMMING

Due to machine-generated nature and its focus on search engines manipulation, spam shows abnormal properties such as high level of duplicate content and links; rapid changes of content; and the language models built for spam pages deviate significantly from the models built for the normal Web.

(a) Spam pages deviate from power law distributions based on numerous web graph statistics such as Page Rank or number of in-links.
(b) Spammers mostly target popular queries and queries with high advertising value.
(c) Spammers build their link farms with the aim to boost ranking as high as possible, and therefore link farms have specific topologies that can be theoretically analyzed on optimality.
(d) According to experiments, the principle of approximate isolation of good pages takes place: good pages mostly link to good pages, while bad pages link either to good pages or a few selected spam target pages. It has also been observed that connected pages have some level of semantic similarity – topical locality of the Web, and therefore label smoothing using the Web graph is a useful strategy.

(e) Numerous algorithms use the idea of trust and distrust propagation using various similarity measures, propagation strategies and seed selection heuristics.
(f) Due to abundance of "neponistic" links, that negatively affect the performance of a link mining algorithm, there is a popular idea of links removal and down weighting. Moreover, the major support is caused by the k-hop neighborhood and hence it makes sense to analyze local sub graphs rather than the entire Web graph.
(g) Because one spammer can have a lot of pages under one website and use them all to boost ranking of some target pages, it makes sense to analyze host graph or even perform clustering and consider clusters as a logical unit of link support.
(h) In addition to traditional page content and links, there are a lot of other sources of information such as user behaviour or HTTP requests. We hope that more will be developed in the near future. Clever feature engineering is especially important for web spam detection.
(i) Despite the fact that new and sophisticated features can boost the state-of-the-art further, proper selection and training of a machine learning models is also of high importance.

III TECHNIQUES OF SPAMMING

Term spamming techniques can be grouped based on the text field in which the spamming occurs [5]. Therefore, we distinguish:

(a) Body spam In this case, the spam terms are included in the document body. This spamming technique is among the simplest and most popular ones, and it is almost as old as search engines themselves [6].
(b) Title spam Today's search engines usually give a higher weight to terms that appear in the title of a document. Hence, it makes sense to include the spam terms in the document title [7].
(c) Meta tag spam The HTML meta tags that appear in the document header have always been the target of spamming. Because of the heavy spamming, search engines currently give low priority to these tags, or even ignore them completely. Here is a simple example of a spammed keywords meta tag:
(d) Anchor text spam Just as with the document title, search engines assign higher weight to anchor text terms, as they are supposed to offer a summary of the pointed document. Therefore, spam terms are sometimes included in the anchor text of the HTML hyperlinks to a page. Please note that this spamming technique is different from the previous ones, in the sense that the spam terms are added not to a target page itself, but the other pages that point to the target. As anchor text gets indexed for both pages, spamming it has impact on the ranking of both the source and target pages [7].
(e) URL spam Some search engines also break down the URL of a page into a set of terms that are used to determine the relevance of the page [8]. To exploit this, spammers sometimes create long URLs that include sequences of spam terms. For instance, one could encounter spam URLs.

Some spammers even go to the extent of setting up a DNS server that resolves any host name within a domain. Often, spamming techniques are combined. For instance, anchor text and URL spam is often encountered together with link spam. Another way of grouping term spamming techniques is based on the type of terms that are added to the text fields. Correspondingly:

(i) Repetition of one or a few specific terms. This way, spammers achieve an increased relevance for a document with respect to a small number of query terms.
(ii) Dumping of a large number of unrelated terms, often even entire dictionaries. This way, spammers make a certain page relevant to many different queries. Dumping is effective against queries that include relatively rare, obscure terms: for such queries, it is probable that only a couple of pages are relevant, so even a spam page with a low relevance/importance would appear among the top results.
(iii) Weaving of spam terms into copied contents. Sometimes spammers duplicate text corpora (e.g., news articles) available on the Web and insert spam terms into them at random positions. This technique is effective if the topic of the original real text was so rare that only a small number of relevant pages exist. Weaving is also used for dilution, i.e., to conceal some repeated spam terms.

IV RELATED WORK

Muhammad U. S. Khan et. al. [8] proposes a framework that separates the spammers and unsolicited bloggers from the genuine experts of a specific domain. The proposed approach employs modified Hyperlink Induced Topic Search (HITS) to separate the unsolicited bloggers from the experts on Twitter on the basis of tweets. The approach considers domain specific keywords in the tweets and several tweet characteristics to identify the unsolicited bloggers.

Y. Chen et. al. [9] utilize graph-based detection due to less security guarantee in feature-based detection. Assuming that fake profiles can establish limited number of intruded (attack) edges, the sub graph formulated by the set of all real accounts is sparsely connected to false account, that is, the cut over intruded edges is sparse. This method makes prediction and find out such sparse cut with formal guarantees. For example, Tuenti deploy SybilRank to rank accounts according to their perceived likelihood of being false, based on structural properties of its social graph and based on their formulation.

H. Gao et. al. [10] utlize graph-based detection provides comfortable security guarantees, real-world social graphs do not conform to the main assumption on which it depends. In particular, various surveys conform that intruders can interstices OSNs on a large scale by deceiving users into befriending their fake profile.

Taghi Javdani et. al. [11]apply the hybrid graph analysis method and behavior analysis, is to increase the diagnostic accuracy and detection rate with the help of appropriate classification algorithms and the most effective features. So, two scenarios were used to achieve higher accuracy level and lower false positive. The first scenario was based on using the entire data to build and evaluate the model. The results showed that despite the high precision of this approach, due to the high levels of false positive, this approach is not appropriate. In the second scenario, the ratio of the normal users to spammers was considered equal to 2 to 1 which led to satisfactory results. After reviewing the confusion matrix and false positives in different algorithms, the Logistic algorithm was chosen as an appropriate algorithm which meets the objective of this study.

Hailu Xu et al.[12]. Studied a methodology to detect spam across online social networks. This methodology focuses on combining spam in one soial network to another social network. They had used 1937 spam tweets and 10942 ham tweets and 1338 spam posts and 9285 ham posts. In TSD, out of 1937 spam tweets, 75.6% spam tweets contained in URL links, 24.4% spam tweets contained in words. From 10942 ham tweets, 62.9% tweets are in URL links and words, remaining 37.1% consist of only words.For the spam posts of FSD, 32.8% spam posts consists of URL links and words, 67.2% of spam posts consist of words. For ham posts 95.1% consist of URL links and 4.9% only consist of words. They had used top 20 word features from Twitter spam data and Face book spam data. They had split the TSD and FSD into training and test data sets .The training and test data sets of TSD, FSD are used to train and test various classifiers like Random forest, logistic, random tree, Bayes Net, Naïve bayes.

M. Okazaki et al. [13].presented an initial study to quantify and characterize spam campaigns launched using accounts on Face book. They studied a large anonym zed dataset of 187 million asynchronous wall messages between Face book users, and used a set of automated techniques to detect and characterize coordinated spam campaigns. Authors detected roughly 200,000 malicious wall posts with embedded URLs, originating from more than 57,000 user accounts.

Fire et al. **[14]**.developed the Social Privacy Safeguard (SPS) software, which is a set of applications for Face book that aim to improve user account privacy policies. The application examines a user's friends list in order to determine accounts that have a risk to the user's privacy. Such accounts could then be protected by users from accessing their profile information. Using these set of data from the SPS developed over Face book, the authors could test several machine learning classifiers to detect fake profiles, some algorithms are been used: Naïve Bayes, Rotation Forest and Random Forest are been used for fake profile detection.

Pern Hui et al. [15] Third-party applications capture the attractiveness of web and platforms providing mobile application. Many of these platforms accept a decentralized control strategy, relying on explicit user consent for yielding permissions that the apps demand. Users have to rely principally on community ratings as the signals to classify the potentially unsafe and inappropriate apps even though community ratings classically reflect opinions regarding supposed functionality or performance rather than concerning risks. To study the advantages of user-consent permission systems through a large data collection of Face book apps, Chrome extensions and Android apps. The study confirms that the current forms of community ratings used in app markets today are not reliable for indicating privacy risks an app creates. It is found with some evidences, indicating attempts to mislead or entice users for granting permissions: free applications and applications with mature content request; "look alike" applications which have similar names as that of popular applications also request more permissions than is typical. Authors find that across all three platforms popular applications request more permissions than average.

J. Kim et al. [16] Twitter can suffer from malicious tweets containing suspicious URLs for spam, phishing, and malware distribution. Attackers have limited resources and thus have to reuse them; a portion of their redirect chains will be shared. We focus on these shared resources to detect suspicious URLs. We have collected a large number of tweets from the Twitter public timeline and trained a statistical classifier with features derived from correlated URLs and tweet context information. Our classifier has high accuracy and low false- positive and false negative rates.

Malik Mateen et al.[17] studied an approach for spam detection in Twitter network. To detect spam in Twitter dataset used different kind of features like user based features, content based features and graph based features. User based features are based on users relationships and properties of user accounts. The spammers have to reach large number of profiles to spread misinformation. Different user account related features are Number of followers, Number of following, age of account, FF ratio and reputation. Content based features are related to tweets posted by user. Different features are total number of tweets, hash tag ratio, URL's ratio, mentions ratio, tweet frequency and spam words. Graph based features are used to identify spammer behaviour. Different features are in/out degree and between's. In the proposed methodology used Twitter dataset consist of 10,256 users and 467480 tweets. To develop a spam detection model used J48, decorate and Naive ayes classifiers. These three classifiers are individually trained on various dataset features and classify the dataset as spam or ham dataset. Out of these three classifiers J48 classifier highest accuracy to classify the data as spam or non spam.

Content based features are best suitable for classifying the dataset. To classify the dataset with highest accuracy combine the content, user based and graph based features. The combined feature set is given as input to the three classifiers. But decorate and J48 classifiers have given highest accuracy up to 97.6%.

Fire et al. [18] developed the Social Privacy Safeguard (SPS) software, which is a set of applications for Facebook that aim to improve user account privacy policies. The application examines a user's friends list in order to determine accounts that have a risk to the user's privacy. Such accounts could then be protected by users from accessing their profile information. Using these set of data from the SPS developed over Facebook, the authors could test several machine learning classifiers to detect fake profiles, some algorithms are been used:Naïve Bayes, Rotation Forest and Random Forest are been used for fake profile detection.

V CONCLUSION

With the rapid growth of social networks, people tend to misuse them for unethical and illegal conducts, fraud and phishing. Creation of a fake profile becomes such adversary effect which is difficult to identify without appropriate research. So this paper have summarize current solutions that have been practically developed and theorized to solve this issue of spam detection issue and spam identification of fake profiles. Here it was obtained that spammers develop high social networking sites than create fake profile on that and start there blogging for target product. It was obtained that most of work use clustering techniques for segregating spammer from real users by reading their behavior on sites. In future it is desired to develop the highly accurate algorithm which not only detects the spam but spammer profile as well.

REFERENCES

[1] Van Dijck, José. The culture of connectivity: A critical history of social media. Oxford University Press, 2013.

[2] Boyd, D., & Ellison, N. (2008). Social network sites: Definition, history, and scholarship. Journal of Computer Mediated Communication, 13(1), 210—23.

[3] Lenhart, Amanda. "Teens, social media & technology overview 2015." Pew Research Center 9 (2015).

[4] Gauri Jain* , 2Manisha, 3Basant Agarwal. "An Overview of RNN and CNN Techniques for Spam Detection in Social Media". Volume 6, Issue 10, October 2016 ISSN: 2277 128X.

[5] I. Drost and T. Sche↵er. Thwarting the nigritude ultramarine: Learning to identify link spam. In Proceedig of the 16th European Conference on Machine Learning, ECML'05, 2005.

[6] C. D. Manning, P. Raghavan, and H. Schtze. Introduction to Information Retrieval. Cambridge University Press, New York, NY, 2008.

[7] O. A. Mcbryan. GENVL and WWWW: Tools for taming the web. In Proceedings of the First World Wide Web Conference, WWW'94, Geneva, Switzerland, May 1994.

[8] Muhammad U. S. Khan, Mazhar Ali, Assad Abbas,Samee U. Khan, *and* Albert Y. Zomaya. "Segregating Spammers and Unsolicited Bloggers from Genuine Experts on Twitter". IEEE Computer Society, 2017.

[9] 28. H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. N. Choudhary. Towards Online Spam Filtering in Social Networks. In NDSS, 2012.

[10] 29. H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao. Detecting and Characterizing Social Spam Campaigns. In Internet Measurement Conference, pp 35–47. ACM, 2010.

[11] Mona Najafi Sarpiri, Taghi Javdani Gandomani, Mahsa Teymourzadeh, Akram Motamedi. "A Hybrid Method for Spammer Detection in Social Networks by Analyzing Graph and User Behavior". Journal of computers, Volume 13, Number 7, July 2018.

[12] Hailu Xu,Weiqing sun,Ahmad javaid: Efficient spam detection across online social networks,IEEE-2015.

[13] T. Sakaki, M. Okazaki, and Y. Matsuo: "Realtime event detection by social sensors", In Proceedings of the 19th international conference on World wide web ACM, 2010.

[14] M. Fire, D. Kagan, A. Elyashar, Y. Elovici, Friend or foe? fake profile identification in online social networks, Social Network Analysis and Mining 4 (1) 1–23, 2014.

[15] Chia, Pern Hui, Yusuke Yamamoto, and N. Asokan. "Is this app safe? a large scale study on application permissions and risk signals." Proceedings of the 21st international conference on World Wide Web. ACM, 2012.

[16] S. Lee and J. Kim, WarningBird: Detecting suspicious URLs in Twitter stream,‖ in Proc. NDSS, 2012.