Advanced Characteristic Analysis of Real Time Junk Occurrences in Twitter

Ancy S
Assistant Professor
Jeppiaar Institute of Technology
Chennai, India

Aruna Jasmine.J
Assistant Professor
Jeppiaar Institute of Technology
Chennai, India

Abstract: Spam on twitter is a major threat in recent days. To overcome these problems we take many steps to work on this. This work uses twitter as the input data source to address the problem of real-time. As twitter data contains a lot of spam, we built a dictionary of words to remove spam from the tweet social media. In order to solve these problem, we firstly carry out a deep analysis on the statistical features of taking training sets of data to differentiate spam tweet and non-spam tweet. Then we propose a approach called “NLTK(Natural Language Tool Kit). The proposed approach can discover “changed” spam posts from unlabeled posts and incorporate them into classifier’s training process. To evaluate the proposed scheme many experiments were carried out. The results show that our proposed NLTK can remarkably improve the spam detection accuracy in real-world scenario

Keywords: cloud computing; Twitter ; Natural Language Tool Kit; Spam; statistical features;

1. INTRODUCTION

In the recent years, the micro blogging social networking service twitter has become a very popular tool for broadcasting news, expressing opinions and communicating with friends. People can publish short text-messages (of 140 characters) which can be viewed by their followers. The usability and ease of using this NLTK(Natural Language Tool Kit) tool contributed to its wide growth. This leads to both efficiency and high throughput rate deployments that uses the computing capacity provided by the network of nodes. Twitter spam, referred to as unsolicited tweets containing malicious link directs victims to external sites containing malware downloads, phishing, drug sales, or scams, etc. As a result of that, security companies, as well as Twitter itself, are combating spammers to make Twitter as a spam-free platform. For instance, Trend Micro uses a blacklisting service called Web Reputation Technology system to filter spam URLs for users who have its products installed. Twitter also implements blacklist filtering as a component in their detection system called BotMaker . However, blacklist fails to save victims from new spam due to its time lag.Before the sites are blocked by the black list ,researches proved that 90% of the users tend to visit the malicious links. In order to give a solution to the drawbacks of blacklists, researchers have come up with some machine learning based schemes which can make use of statistical features from sspammers’ or spam tweets’ to detect spam without checking the URLs. However, the observation made in our collected data set shows that the characteristics of spam tweets are varying over time. We name this problem as “Twitter Spam Drift”. As earlier Machine Learning dependent classifiers are not updated with the “changed” spam tweets, the performance of such classifiers are dramatically influenced by “Spam Drift” when detecting new coming spam tweets. Why do spam tweets drift over time? It is because that the spammers are struggling with security companies and researchers. The researchers are working to detect spam and on the other side spammers are trying not to be detected. This leads spammers to evade current detection features through posting more tweets or creating spam with the similar semantic meaning but using different text such social media is a good resource to obtain informal names of places such as acronyms, abbreviations, or nicknames. Additionally, words which indicate specific locations other than place names such as names of local foods or products and regional dialects and the words which indicate specific locations only temporarily such as the names of events can also be obtained. All of these local words, which indicate specific locations at some point, would enrich the geographical dictionary; however, most of the existing methods extract the local words from the posts accumulated for a long period of time in a batch process, which makes it impossible to handle the temporal changes of the local words or of the locations indicated by them.In this work, we firstly illustrate the “Twitter spam drift” problem through analyzing the statistical properties of Twitter spam in our collected dataset and then its impact on detection performance of several classifiers.

2. RELATED WORKS

Many research works have been carried out for solving this spam problem. Some of the most recent works by Abdullah Talha Kabakus,Resul Kara in the year: 2017 . Twitter is one of the most popular social media platforms that has 313 million monthly active users which post500 million tweets per day. This popularity attracts the attention of spammers who use Twitter for their malicious aims such as phishing legitimate users or spreading malicious software and advertises through URLs shared within tweets, aggressively follow/unfollow legitimate users and hijack trending topics to attract their attention, propagating pornography. In August of2014, Twitter revealed that 8.5% of its monthly active users which equals approximately 23 million users have automatically contacted their servers for regular updates. Thus, detecting and filtering spammers from legitimate users are mandatory in order to provide a spam-free environment in Twitter. In this paper, features of Twitter spam detection presented with discussing their effectiveness. Also, Twitter spam detection methods are categorized and discussed with their pros and cons. The outdated features of Twitter which are commonly used by Twitter spam detection approaches are highlighted. Some new features of Twitter which, to the best of our knowledge, have not been mentioned by any other works are also presented.

Chao Yang, Robert Harkreader, Jialong Zhang worked in Analyzing Spammer’s Social Networks for Fun and Profit in the year: 2012. In this paper, we perform an empirical analysis of the cyber criminal ecosystem on Twitter. Essentially, through analyzing inner social relationships in the criminal account community, we find that criminal accounts tend to be socially connected, forming a small-world network. We also find that criminal hubs, sitting in the center of the social graph, are more inclined to follow criminal accounts. Through analyzing outer social relationships between criminal accounts and their social friends outside the criminal account community, we reveal three categories of accounts that have close friendships with criminal accounts. Through these analyses, we provide a novel and effective criminal account inference algorithm by exploiting criminal accounts' social relationships and semantic co ordinations.

Hongyu Gao, Yan Chen, Kathy Lee,Diana Palsetia,Alok Choudhary works in Towards Online Spam Filtering in Social Networks in the year :2008. Online social networks (OSNs) are extremely popular among Internet users. Unfortunately, in the wrong hands, they are also effective tools for executing spam campaigns. In this paper, we present an online spam filtering system that can be deployed as a component of the OSN platform to inspect messages generated by users in real-time. We propose to reconstruct spam messages into campaigns for classification rather than examine them individually. Although campaign identification has been used for offline spam analysis, we apply this technique to aid the online spam detection problem with sufficiently low overhead. Accordingly, our system adopts a set of novel features that effectively dis-tinguish spam campaigns. It drops messages classified as“ spam” before they reach the intended recipients, thus protecting them from various kinds of fraud. We evaluate the system using 187 million wall posts collected from Face- book and 17 million tweets collected from Twitter. In different parameter settings, the true positive rate reaches 80.9%while the false positive rate reaches 0.19% in the best case. In addition, it stays accurate for more than 9 months after the initial training phase. Once deployed, it can constantly secure the OSNs without the need for frequent re-training. Finally, tested on a server machine with eight cores (XeonE5520 2.2Ghz) and 16GB memory, the system achieves an average throughput of 1580 messages/sec and an average processing latency of 21.5ms on the Facebook dataset.

Then the Detecting and Characterizing Social Spam Campaigns by Hongyu Gao, Jun Hu, Christo Wilson in the year 2009. Said that Online social networks (OSNs) are popular collaboration and communication tools for millions of users and their friends. Unfortunately, in the wrong hands, they are also effective tools for executing spam campaigns and spreading malware. Intuitively, a user is more likely to respond to a message from a Facebook friend than from a stranger, thus making social spam a more effective distribution mechanism than traditional email. In fact, existing evidence shows malicious entities are already attempting to compromise OSN account credentials to support these “high-return” spam campaigns. In this paper, we present an initial study to quantify and characterize spam campaigns launched using accounts on online social networks. We study a large anonymized dataset of asynchronous “wall” messages between facebook users. We analyze all wall messages received by roughly 3.5 million facebook users (more than 187 million messages in all), and use a set of automated techniques to detect and characterize coordinated spam campaigns. Our system detected roughly 200,000 malicious wall posts with embedded URLs, originating from more than 57,000 user accounts. We find that more than 70% of all malicious wall posts advertise phishing sites. We also study the characteristics of malicious accounts, and see that more than 97% are compromised accounts, rather than “fake” accounts created solely for the purpose of spamming. Finally, we observe that, when adjusted to the local time of the sender, spamming dominates actual wall post activity in the early morning hours, when normal users are asleep.

Also in Detecting Spammers on Twitter by Fabŕıcio Benevenuto, Gabriel Magno, Tiago Rodrigues ,Virǵılio Almeida. In 2010.With millions of users tweeting around the world, real time search systems and different types of mining tools are merging to allow people tracking the repercussion of event sand news on Twitter. However, although appealing as mechanisms to ease the spread of news and allow users to discuss events and post their status, these services open opportunities for new forms of spam. Trending topics, the most talked about items on Twitter at a given point in time, have been seen as an opportunity to generate traffic and revenue. Spammers post tweets containing typical words of a trending topic and URLs, usually obfuscated by URL shorteners , that lead users to completely unrelated websites. This kind of spam can contribute to de-value real time search services unless mechanisms to fight and stop spammers can be found. In this paper we consider the problem of detecting spammers on Twitter. We first collected a large dataset of Twitter that includes more than 54 million users, 1.9 billion links, and almost 1.8 billion tweets. Using tweets related to three famous trending topics from 2009, we construct a large labelled collection of users, manually classified into spammer sand non-spammers. We then identify a number of characteristics related to tweet content and user social behaviour, which could potentially be used to detect spammers. We used these characteristics as attributes of machine learning process for classifying users as either spammers or non spammers. Our strategy succeeds at detecting much of the spammers while only a small percentage of non-spammer are misclassified. Approximately 70% of spammers and 96%of non-spammers were correctly classified. Our results also highlight the most important attributes for spam detection on Twitter.

3. PROPOSED WORK

Better result is obtained through the proposed methodology,Real-world dataset is collected and labelled, which contains 10 consecutive days' tweets with 100k spam tweets and 100k non-spam tweets in each day (2 million tweets in total). This dataset is available for researchers to study Twitter spam. we analyze the “Twitter Spam Drift” problem from both data analysis and experimental evaluation aspects. We are the first to study this problem in Twitter spam detection to the best of our knowledge,. We propose a NLTK approach which learns from unlabelled tweets to deal with “Twitter Spam Drift”. Through our evaluations, we show that this proposed NLTK can effectively detect Twitter spam by cutting down the influence of “Spam Drift” issue.

NLTK will aid you with everything from splitting sentences from paragraphs, splitting up words, recognizing the part of speech of those words, highlighting the main subjects, and then even with helping your machine to understand what the text is all about. In this series, we're going to tackle the field of opinion mining, or sentiment analysis.

Advantages: 1.NLTK can effectively detect twitter spam by reducing the impact of “Spam Drift” issue. 2. If user post unwanted tweets more times it will be remove the followers

3.1 Architecture

The system architecture is given in figure 1. The user initially registers the login page. Once the registration process is over the registration details are stored in the database which is maintained by admin. The next time when the user logs in the username and the mail id is validated and if correct the user will be able to log in to the twitter account. As soon as the user logs in he or she can view the various tweets updated. All the tweets are classified using the RF algorithms. The result of the classification says whether it is a spam mail or not spam. If it is found to be spam tweet then NLTK is applied and the particular source is blocked or un followed.

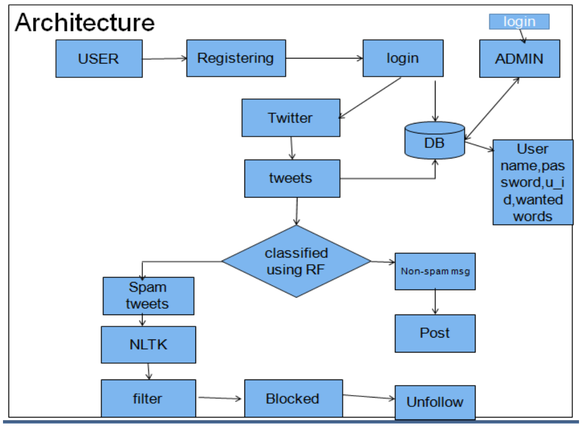

Figure. 1 Syste, Architecture

3.2 Authentication

In the module it we authenticate a valid user to enter into a twitter web page. Users are usually provided with a user ID, and authentication is accomplished when the user provides a valid credential, for example a password, that matches with that user ID. Most users are most familiar with using a password, which, as a piece of information that should be known only to the user, is called a knowledge factor. During authentication, credentials provided by the user are compared to those on file in a database of authorized users' information either on the local operating system or through an authentication server. If the credentials matches, and the authenticated user is authorized to use the resource, the process is completed and the user is granted access. The permissions and folders returned define both the environment the user sees and the way user can interact with it, including hours of access and other rights such as the amount of resource storage space. The username and the password is validated with data in the DB .

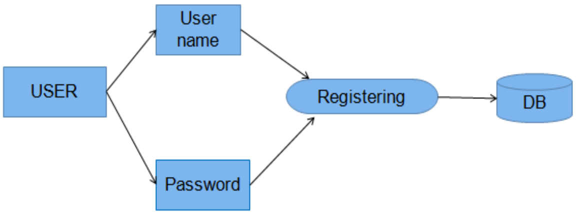

Figure. 2 Authentication process

3.3 Discovering

One of the efficient ways to end up in a subscriber's spam folder or junk folder is to load up your own email with words that have been identified as common words in spam mails by most of the email service providers. Spam words and phrases are that which can set off email service provider spam filters.Have in mind that it doesn't mandatorily mean that you can't use these words in variation. However, too many of them or too much repetition of one of them can land in you the spam In order to validate the performance of twitter spam, we replicated a spammer's behavior by building a spam campaign generator that mimics a commercially available spamming tool.or junk folder.This has a source of good and bad words and based on which the tweet is classified as spam tweet or non spam tweet

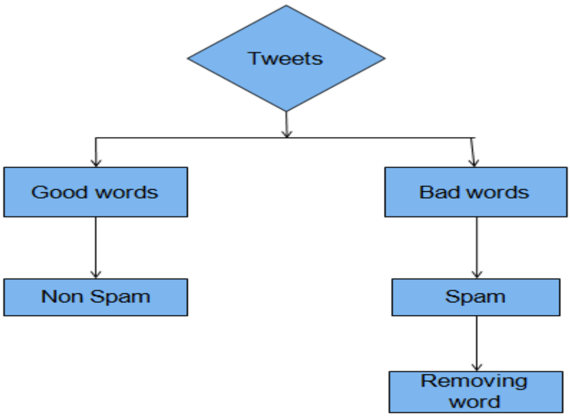

Figure. 3 Discovering process

3.4 Filtering

In this module we are separating the spam and all other mails. So that it will be easy for us to send it to trash. A spam filter looks for certain criteria on which it bases its judgments. For example, the simplest and earliest versions (such as the one available with Microsoft's Hotmail) can be set to watch in the subject line of messages for particular words and to exclude these from the user's inbox.

The method that employed here is not much effective, very often leaving out perfectly legitimate messages which are called false positives and letting actual spam through. Many well established programs, like Bayesian filters or other heuristic filters, attempt to identify spam through doubtful word patterns or word frequency. The several different types of spam filters which are already existing are as follows.

1. Content filters – the message content is checked to find out if there is spam or not.
2. Header filters – the email header is verified to find the presence of fraudulent content.
3. General blacklist filters – matches the sender address with the blacklisted spammer mail ids and intimate.
4. Rules-based filters – use user-defined criteria – such as notified senders or user defined specific wording in the subject line or body – to block spam
5. Permission filters – require anyone sending a message to be pre-approved by the recipient
6. Challenge-response filters – require anyone sending a message to enter a code in order to gain permission to send email

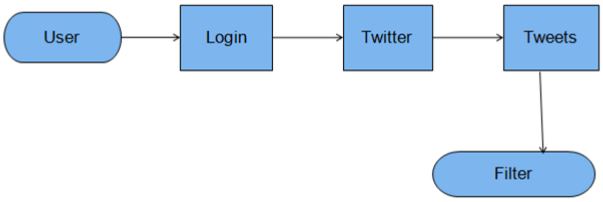

Figure.4 Filtering

3.5 Removing the spam

We believe these results to show clearly that Big data spam detection technique are ripe for in-production deployment. The spam detection mechanism currently uses the email body only.

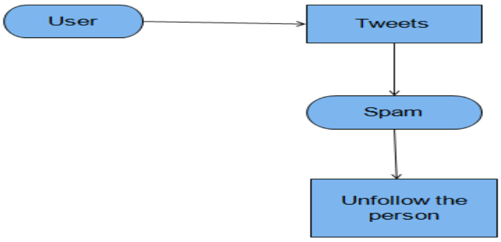

Figure. 5 Removing the spam

4. CONCLUSION

In this paper, we firstly sympathize the “Spam Drift” problem in statistical features based Twitter spam detection. In order to solve this problem, we propose a NLTK approach. In our NLTK scheme, classifiers will be trained again by the added “changed spam” tweets which are learnt from unlabelled samples, thus it can reduce the impact of “Spam Drift” to a great extent. We evaluate the performance of NLTK approach in terms of Detection Rate and F-measure. Experimental results show that both detection rate and F-measure are improved a lot when applying with our NLTK approach We also compare NLTK to four traditional machine learning algorithms, and find that our NLTK outperforms all four algorithms in terms of overall accuracy, F-measure and Detection Rate.

5.FUTURE SCOPE

The future work could be the revolutionizing trend of average value of each feature for two classes in 10 days. In routine, the variation of average value of feature from spam tweets is greater than that of non-spam tweets.

6.REFERENCES

[1] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, “Detecting spammer on twitter,” in Proc. 7th Annu. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf., Jul. 2010, p. 12.

[2] L. Breiman, “Random forests,” Mach. Learn., vol. 45, no. 1, pp. 532, 2001.

[3] C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on twitter,” in Proc. 20th Int. Conf. World Wide Web, 2011, pp. 675–684.

[4] C. Chen, J. Zhang, X. Chen, Y. Xiang, and W. Zhou, “6 million spam tweets: A large ground truth for timely twitter spam detection,” in Proc. IEEE Commun. Inf. Syst. Security Symp. (ICCCISS), Jun. 2015, pp. 8689–8694.

[5] C. Chen, J. Zhang, Y. Xiang, and W. Zhou, “Asymmetric self-learning for tackling twitter spam drift,” in Proc. 3rd Int. Workshop Security. Privacy Big Data (BigSecurity), Apr. 2015, pp. 237–242.

[6] C. Chen, J. Zhang, Y. Xiang, W. Zhou, and J. Oliver, “Spammers are becoming ‘smarter’ on twitter,” IT Prof., vol. 18, no. 2, pp. 14–18, Apr. 2016.

[7] E. M. Clark, J. R. Williams, C. A. Jones, R. A. Galbraith, C. M. Danforth, and P. S. Dodds, “Sifting robotic from organic text: A natural language approach for detecting automation on twitter,” J. Comput. Sci., vol. 16, p. 1–7, Sep. 2016.

[8] (2016). Whole Product Dynamic Real-World Protection Test, Av Comparatives, accessed on Aug. 1, 2015. [Online]. Available: http://www.avcomparatives. org/wp-content/uploads/2016/07/avc_prot_2016a_en.pdf

[9] I. Csiszar and J. Körner, Information Theory: Coding Theorems For Discrete Memoryless Systems. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[10] T. Dasu, S. Krishnan, S. Venkatasubramanian, and K. Yi, “An Information-theoretic approach to detecting changes in multidimensional data streams,” in Proc. Symp. Interface Statist., Comput. Sci., Appl., 2006, pp. 1–24.