IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 26-29

Statistical Features Based Real-time Detection of Drifted Twitter Spam

Mr .M. Madhan .M.E., Assistant Professor
(Department Of Computer Science Engineering, Mailam Engineering College,Mailam)

Abstract : *The Twitter Spam has become a critical problem nowadays. Recent works focus on applying machine learning techniques for Twitter spam detection, which make use of the statistical features of tweets. In our labelled tweets dataset, however, we observe that the statistical properties of spam tweets vary over time, and thus the performance of existing machine learning based classifiers decreases. This issue is referred to as "Twitter Spam Drift". In order to tackle this problem, we firstly carry out a deep analysis on the statistical features of one million spam tweets and one million non-spam tweets, and then propose a novel Lfun scheme. The proposed scheme can discover "changed" spam tweets from unlabelled tweets and incorporate them into classifier's training process. A number of experiments are performed to evaluate the proposed scheme. The results show that our proposed Lfun scheme can significantly improve the spam detection accuracy in real-world scenarios.*

I. Introduction

The TWITTER has become one of the most popular social networks in the last decade. It is rated as the most popular social network among teenagers according to a recent report [17]. However, the exponential growth of Twitter also contributes to the increase of spamming activities. Twitter spam, which is referred to as unsolicited tweets containing malicious link that directs victims to external sites containing malware downloads, phishing, drug sales, or scams, etc. [1], not only interferes user experiences, but also damages the whole Internet. In September 2014, the Internet of New Zealand was melt down due to the spread of malware downloading spam. This kind of spam lured users to click links which claimed to contain Hollywood star photos, but in fact directed users to download malware to perform DDoS attacks [29]. Consequently, security companies, as well as Twitter itself, are combating spammers to make Twitter as a spam-free platform. For example, Trend Micro uses a blacklisting service called Web Reputation Technology system to filter spam URLs for users who have its products installed [27]. Twitter also implements blacklist filtering as a component in their detection system called BotMaker [21]. However, blacklist fails to protect victims from new spam due to its time lag [18]. Research shows that, more than 90% victims may visit C. Chen is with University of Electronic Science and Technology of China and the School of Information Technology, Deakin University, Australia email: chao.chen@deakin.edu.au. Y. Wang, J. Zhang, Y. Xiang and W. Zhou is with the School of Information Technology, Deakin University, Australia e-mail: fy.wang, jun.zhang, yang.xiang, wanlei.zhoug@deakin.edu.au. G. Min is with University of Electronic Science and Technology of China and University of Exeter, UK e-mail: g.min@exeter.ac.uk. a new spam link before it is blocked by blacklists [37]. In order to address the limitation of blacklists, researchers have proposed some machine learning based schemes which can make use of spammers' or spam tweets' statistical features to detect spam without checking the URLs [14], [40]. Machine Learning (ML) based detection schemes involve several steps. First, statistical features, which can differentiate spam from non-spam, are extracted from tweets or Twitter users (such as account age, number of followers or friends and number of characters in a tweet). Then a small set of samples are labelled with class, i.e. spam or non-spam, as training data. After that, machine learning based classifiers are trained by the labelled samples, and finally the trained classifiers can be used to detect spam. A number of ML based detection schemes have been proposed by researchers [1], [35], [39], [43]. However, the observation in our collected data set shows that the characteristics of spam tweets are varying over time. We refer to this issue as "Twitter Spam Drift". As previous UML based 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 spammers are struggling with security companies and researchers. While researchers are working to detect spam, spammers are also trying to avoid being 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 [34], [39]. In this work, we firstly illustrate the "Twitter spam drift" problem through analysing the statistical properties of Twitter spam in our collected dataset and then its impact on detection performance of several classifiers. By observing that there are "changed" spam samples in the coming tweets, we propose a novel Lfun (Learning from unlabelled tweets) approach, which

updates classifiers with the spam samples from the unlabelled incoming tweets. In summary, our contributions are listed below: We collect and label a real-world dataset, 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 1.

We investigate the "Twitter Spam Drift" problem from both data analysis and experimental evaluation aspects. To the best of our knowledge, we are the first to study this problem in Twitter spam detection.

We propose a novel Lfun approach which learns from unlabelled tweets to deal with "Twitter Spam Drift". Through our evaluations, we show that our proposed Lfun can effectively detect Twitter spam by reducing the

impact of "Spam Drift" issue. The rest of this paper is organized as follows. Section II presents a review on machine learning based methods for Twitter

spam detection. In Section III, the collection and labeling of the data used in our work is introduced. Meanwhile, the "Spam Drift" problem is illustrated and justified. Then we introduce our Lfun approach in Section IV, and analyze the performance benefit of our approach. Section V evaluates our

Lfun approach and compares it with four traditional machine learning algorithms. Finally, Section VII concludes this work and introduces our future work.

II. Existing System

Although there are a few works, such as and which are suitable to detect streaming spam tweets, there lacks of a performance evaluation of existing machine learning-based streaming spam detection methods.

In this paper, we aim to bridge the gap by carrying out a performance evaluation, which was from three different aspects of data, feature, and model. Others apply existing blacklisting service, such as Google Safe Browsing to label spam tweets.

Nevertheless, these services' API limits make it impossible to label a large amount of tweets. However, Twitter has around 5% spam tweets of all existing tweets in the real world.

DISADVANTAGE OF EXISTING SYSTEM

it is very time- and resource-consuming.

it is also costly and sometimes the results are doubtful ,Others apply existing blacklisting services, such as Google, Safe Browsing and URIBL to label spam tweets.

III. Existing System

Although there are a few works, such as and which are suitable to

IV. Proposed & Modification System

The Consequently, the research community, as well as Twitter itself, has proposed some spam detection schemes to make Twitter as a spam-free platform. For instance, Twitter has applied some "Twitter rules" to suspend accounts if they behave abnormally.

Those accounts, which are frequently requesting to be friends with others, sending duplicate content, mentioning others users, or posting URL-only content, will be suspended by Twitter.

Twitter users can also report a spammer to the official @spam account. To automatically detect spam, machine learning algorithms have been applied by researchers to make spam detection as a classification problem.

Most of these works classify a user is spammer or not by relying on the features which need historical information of the user or the exiting social graph. For example, the feature, "the fraction of tweets of the user containing URL" used in must be retrieved from the users' tweets list; features such as, "average neighbors' tweets" in and "distance" in cannot be extracted without the built social graph.

However, Twitter data are in the form of stream, and tweets arrive at very high speed. Despite that these methods are effective in detecting Twitter spam, they are not applicable in detecting streaming spam tweets as each streaming tweet does not contain the historical information or social graph that are needed in detection.

Advantage Of Proposed System

Advantage of online learning and incremental learning, i.e., it can be deployed without much training data at the beginning.

This competent is automatically updated with detected spam tweets with no human effort.

V. Machine Learning Algorithms

Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions expressed as outputs. Machine learning is closely related to and often overlaps with computational statistics; a discipline which also focuses in prediction-making through the use of computers.

Process of ML-Based Twitter Spam Detection This section describes the process of Twitter spam detection by using machine learning algorithms. illustrates the steps involved in building a supervised classifier and detecting Twitter spam. Before classification, a classifier that contains the knowledge structure should be trained with the pre labeled tweets. After the classification model gains the knowledge structure of the training data, it can be used to predict a new incoming tweet. The whole process consists of two steps: 1) learning and 2) classifying.

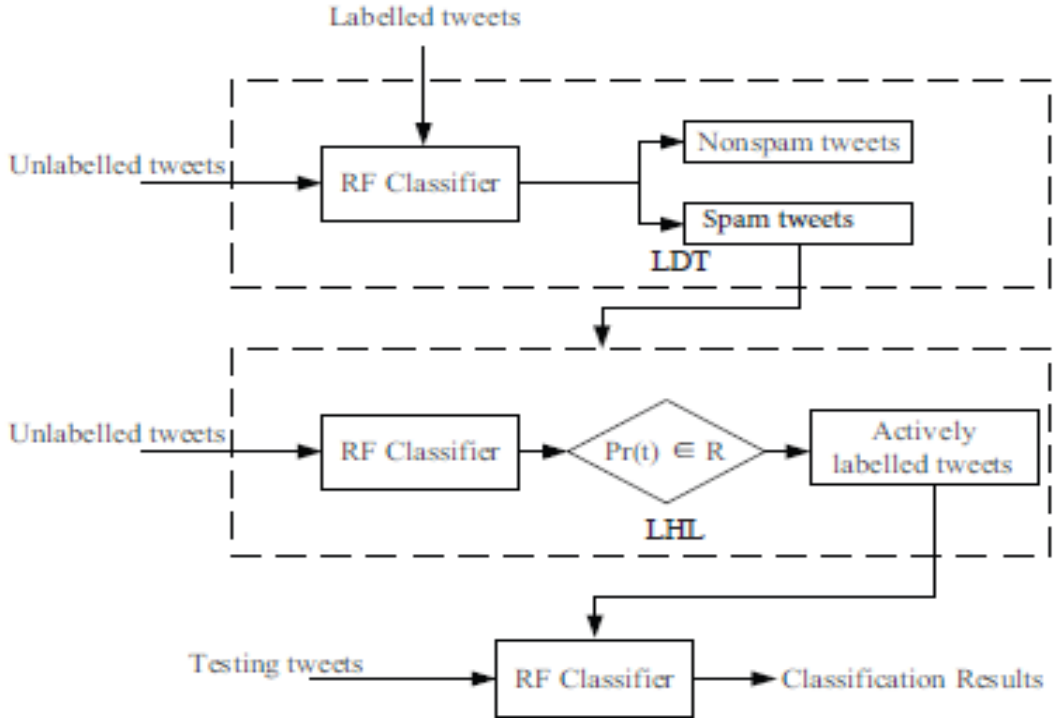
Lfun Algorithm

Require: labelled training set f 1; :::; Ng,
unlabelled tweets Tunlabelled,
a binary classification algorithm _,
Ensure: manually labelled selected tweets Tm
1: Tlabelled
SN
i=1 i
// Use _ to create a classifier Cls from Tlabelled:
2: Cls _ : Tlabelled
// Tunlabelled is classified as Tspam and Tnon□spam:
3: Tspam + Tnon□spam Tunlabelled
// Merge spam tweets Tspam classified by Cls into
Tlabelled:
4: Tex Tlabelled + Tspam
// use Tex to re-train the classifier Cls :
5: Cls _ : Tex
// determine the incoming tweet's suitability for

selection:

6: U ;
7: for i = 1 to k do
8: if Ui meet the selection criteria S then
9: U (U [Ui)
10: end if
11: end for
// manually labelling each ui in U
12: Tm ;
13: for i = 1 to k do
14: manually label each ui
15: Tm (Tm [ui)
16: end for

VI. ARCHITECTURE DIAGRAM



VII. Conclusion

Machine In this paper, we provide a fundamental evaluation of ML algorithms on the detection of streaming spam tweets. In our evaluation, we found that classifiers' ability to detect Twitter spam reduced when in a near real-world scenario since the imbalanced data brings bias. We also identified that Feature discretization was an important preprocesses to ML-based spam detection. Second, increasing training data only cannot bring more benefits to detect Twitter spam after a certain number of training samples. In this paper, we firstly identify the "Spam Drift" problem in statistical features based Twitter spam detection. In order to solve this problem, we propose a Lfun approach. In our Lfun scheme, classifiers will be re-trained by the added "changed spam" tweets which are learnt from unlabelled samples, thus it can reduce the impact of "Spam Drift" significantly. There is also a limitation in our Lfun scheme. The benefit of "old" labelled spam is to eliminate the impact of "spam drift" to classify more accurate spam tweets in future days. The effectiveness of "old" spam has been proved by our experiments during a short period. However, the effectiveness will decrease as the correlation of "very old" spam becomes less with the new spam in the long term run. In the future, we will incorporate incremental adjustment to adjust the training data, such as dropping the "too old" samples after a certain time. It can not only eliminate unuseful information in the training data but also make it faster to train the model as the number of training samples decrease.

References

[1] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna. Compa: Detecting compromised accounts on social networks. In Annual Network and Distributed System Security Symposium, 2013.

[2] J. a. Gama, I. ˇ Zliobait˙e, A. Bifet, M. Pechenizkiy, and A. Bouchachia. A survey on concept drift adaptation. ACM Comput. Surv., 46(4):44:1–44:37, Mar. 2014.

[3] H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary. Towards online spam filtering in social networks. In NDSS, 2012.

[4] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao. Detecting and characterizing social spam campaigns. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, IMC '10, pages 35–47, New York, NY, USA, 2010. ACM.

[5] H. Gao, Y. Yang, K. Bu, Y. Chen, D. Downey, K. Lee, and A. Choudhary. Spam ain't as diverse as it seems: Throttling osn spam with templates underneath. In Proceedings of the 30th Annual Computer Security Applications Conference, ACSAC '14, pages 76–85, New York, NY, USA, 2014. ACM.

[6] A. Greig. Twitter overtakes facebook as the most popular social network for teens, according to study. DailyMail, October 2013.

[7] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: the underground on 140 characters or less. In Proceedings of the 17th ACM conference on Computer and communications security, CCS '10, pages 27–37, New York, NY, USA, 2010. ACM.

[8] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier. TweetCred: Real-Time Credibility Assessment of Content on Twitter, pages 228–243. Springer International Publishing, Cham, 2014.

[9] K. Huang, Z. Xu, I. King, M. Lyu, and C. Campbell. Supervised selftaught learning: Actively transferring knowledge from unlabeled data. In Neural Networks, 2009. IJCNN 2009. International Joint Conference on, pages 1272–1277, June 2009.