Real-time tweet Spam Filtering

Md Abdullah Al Mamun
Qatar University
Doha, Qatar
mm1620594@qu.edu.qa

ABSTRACT

An upsurge in spam volume and sophistication is the result of the quick advancement of Twitter. Spam tweet is an important and difficult problem for a long time. The misuse of undeniable Twitter services such as shortened URLs, shortened URLs, "hashtags", and "mentions", permits spammers to activate proficiently. Similar activities, however, may be a fundamental element in finding new spam accounts as demonstrated by earlier researchers. To defend Twitter users from spamming behaviors, many IT specialist, and researchers have proposed many revealing and filtering mechanism so far. Therefore, we are motivated to propose a new model for filtering out the non-relevant/span tweets that have a certain hashtag in real-time using sparkStreaming.

KEYWORDS

Twitter, Tweets, Hashtag, Spam, SprkStreaming

ACM Reference Format:
Md Abdullah Al Mamun. 1997. Real-time tweet Spam Filtering. In *Proceedings of ACM Woodstock conference (WOODSTOCK'97),* Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Online social media became more and more accepted cooperation and interaction tools for Billions of Internet users due to itâĂŹs easy to use and full-time availability. One of the famous social media platforms, Twitter draws attention by delivering free microblogging services for clients to announce or broadcast short messages in 140 characters through dissimilar devices, for instance, mobile phones, tab or desktops, follow other users and so on, [3]. Many tweeter users post their messages or share their suitable moments, for example, breaking news or events to the fans and followers [4]. Conversely, the easiness and accessibility of this platform also attract hackers and criminals to make money dishonestly or damage by keeping busy the twitter server. These attacks are listed as spam, scam, and phishing [1], [14]. Due to having the character limits on tweets, the attackers frequently post the fake URLs that redirect users to an external malicious webpage which can be harmful to the browser or even computer [9]. Twitter junk is further unsafe and high-level in tempting its users to get cheated compared to the conventional spam that spread through emails [12]. A contemporary study shows that the rate of click on spam through Twitter achieves 0.13%, while it only ranges 0.0003% 0.0006% in usual email spam [6]. In order to address the problem of Twitter spam, we propose a real-time spam filtering method using sparkStreaming. More specifically, given a set of related hashtags, filter out non-relevant/spam tweets that have that hashtag in real-time is our main objective.

WOODSTOCK'97, July 1997, El Paso, Texas USA

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

2 RELATED WORKS

In current years, there are many spam filtering mechanisms has been proposed. Specifically, in last two years, many innovative ideas have been advanced that significantly enhanced the filtering accuracy and efficiency compared to those which were proposed earlier. Current twitter spam filtering approaches can be divided into three categories such as feature analysis, syntax analysis, and blacklisting techniques. Since, text the only format of tweets, many scientists rely on semantically detection technique [3, 7, 8] while some other works were proposed detection algorithm using feature extraction from both account and messages and applied them on a statistical analysis [2, 7, 10, 11, 13]. Others use blacklisting technique as a third party service to block malicious and anonymous posting [5, 13].

3 METHOD

In this project, the real-time tweet stream will be used as a dataset. Hadoop and SparkStreaming tools will be used for experimentation.

4 RESEARCH PLAN

The time line of this project is shown in the Table 1. The tentative date of the project completion is Jun 11.

Table 1: Project Time-line

Sl no.	Deliverable Item	Date
1	Literature Review	May 10
2	Experimental Work	May 22
3	Result and Discussion	May 28
4	Introduction and Conclusion	Jun 3
5	Demo	Jun 6
6	Report Submission	Jun 11

ACKNOWLEDGMENTS

The authors would like to thank Dr. Tamer Elsayed for providing the resourceful lectures and practical training on Bigdata Analytics.

REFERENCES

[1] Kayode Sakariyah Adewole, Nor Badrul Anuar, Amirrudin Kamsin, Kasturi Dewi Varathan, and Syed Abdul Razak. 2017. Malicious accounts: dark of the social networks. *Journal of Network and Computer Applications* 79 (2017), 41–67.

[2] Chao Chen, Jun Zhang, Yang Xiang, and Wanlei Zhou. 2015. Asymmetric self-learning for tackling twitter spam drift. In *Computer Communications Workshops (INFOCOM WKSHPS), 2015 IEEE Conference on*. IEEE, 208–213.

[3] Zi Chu, Indra Widjaja, and Haining Wang. 2012. Detecting social spam campaigns on twitter. In *International Conference on Applied Cryptography and Network Security*. Springer, 455–472.

[4] Saptarshi Ghosh, Bimal Viswanath, Farshad Kooti, Naveen Kumar Sharma, Gautam Korlam, Fabricio Benevenuto, Niloy Ganguly, and Krishna Phani Gummadi. 2012. Understanding and combating link farming in the twitter social network. In *Proceedings of the 21st international conference on World Wide Web*. ACM, 61–70.

[5] Zafar Gilani, Reza Farahbakhsh, and Jon Crowcroft. 2017. Do Bots impact Twitter activity?. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 781–782.

[6] Chris Grier, Kurt Thomas, Vern Paxson, and Michael Zhang. 2010. @ spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 27–37.

[7] Xia Hu, Jiliang Tang, and Huan Liu. 2014. Online Social Spammer Detection.. In *AAAI*, Vol. 14. 59–65.

[8] Xia Hu, Jiliang Tang, Yanchao Zhang, and Huan Liu. 2013. Social Spammer Detection in Microblogging.. In *IJCAI*, Vol. 13. 2633–2639.

[9] Sangho Lee and Jong Kim. 2013. Warningbird: A near real-time detection system for suspicious urls in twitter stream. *IEEE transactions on dependable and secure computing* 10, 3 (2013), 183–195.

[10] Shigang Liu, Yu Wang, Jun Zhang, Chao Chen, and Yang Xiang. 2017. Addressing the class imbalance problem in twitter spam detection using ensemble learning. *Computers & Security* 69 (2017), 35–49.

[11] Carl Sabottke, Octavian Suciu, and Tudor Dumitras. 2015. Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits.. In *USENIX Security Symposium*. 1041–1056.

[12] Kurt Thomas, Chris Grier, Justin Ma, Vern Paxson, and Dawn Song. 2011. Design and evaluation of a real-time url spam filtering service. In *Security and Privacy (SP), 2011 IEEE Symposium on*. IEEE, 447–462.

[13] Xianchao Zhang, Zhaoxing Li, Shaoping Zhu, and Wenxin Liang. 2016. Detecting spam and promoting campaigns in Twitter. *ACM Transactions on the Web (TWEB)* 10, 1 (2016), 4.

[14] Yin Zhu, Xiao Wang, Erheng Zhong, Nathan Nan Liu, He Li, and Qiang Yang. 2012. Discovering Spammers in Social Networks.. In *AAAI*.