International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Spammer Detection and Fake User

Mr. B. B. K. Prasad[1], MD. INTHIYAZ[2], K. SATYA SRI[3], B.S.S. N GAYATRI[4]
1 Associate Professor Dept. of IT, NRI Institute of Technology, A. P, India-521212
2,3,4 UG Scholar, Dept. of IT, NRI Institute of Technology, A.P, India - 521212

ABSTRACT
Social networking sites engage millions of users around the world. The users' interactions with these social sites, such as Twitter and Facebook have a tremendous impact and occasionally undesirable repercussions for daily life. The prominent social networking sites have turned into a target platform for the spammers to disperse a huge amount of irrelevant and deleterious information. Twitter, for example, has become one of the most extravagantly used platforms of all times and therefore allows an unreasonable amount of spam. Fake users send undesired tweets to users to promote services or websites that not only affect legitimate users but also disrupt resource consumption. Moreover, the possibility of expanding invalid information to users through fake identities has increased that results in the unrolling of harmful content. Recently, the detection of spammers and identification of fake users on Twitter has become a common area of research in contemporary online social Networks (OSNs). In this paper, we perform a review of techniques used for detecting spammers on Twitter. Moreover, a taxonomy of the Twitter spam detection approaches is presented that classifies the techniques based on their ability to detect: (i) fake content, (ii) spam based on URL, (iii) spam in trending topics, and (iv) fake users. The presented techniques are also compared based on various features, such as user features, content features, graph features, structure features, and time features. We are hopeful that the presented study will be a useful resource for researchers to find the highlights of recent developments in Twitter spam detection on a single platform.

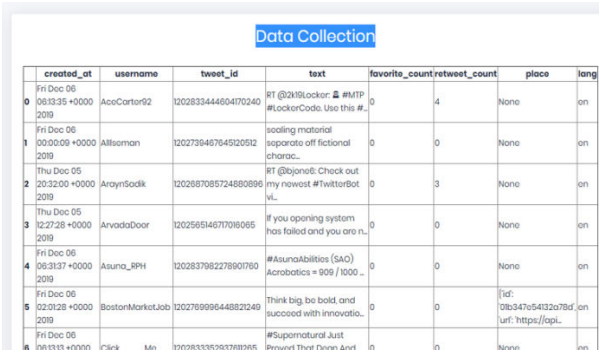
Introduction

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 labeled tweets data set, 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 first 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 unlabeled 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.Information quality in social

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

media is an increasingly important issue, but web-scale data hinders experts' ability to assess and correct much of the inaccurate content, or "fake news," present in these platforms. This paper develops a method for automating fake news detection on Twitter by learning to predict accuracy assessments in two credibility-focused Twitter datasets: CREDBANK, a crowdsourced dataset of accuracy assessments for events in Twitter, and PHEME, a dataset of potential rumors in Twitter and journalistic assessments of their accuracies. We apply this method to Twitter content sourced from Buzz Feed's fake news dataset and show models trained against crowdsourced workers outperform models based on journalists' assessment and models trained on a pooled dataset of both crowdsourced workers and journalists. All three datasets, aligned into a uniform format, are also publicly available. A feature analysis then identifies features that are most predictive for crowdsourced and journalistic accuracy assessments, results of which are consistent with prior work. We close with a discussion contrasting accuracy and credibility and why models of non-experts outperform models of journalists for fake news detection in Twitter.The popularity of Twitter attracts more and more spammers. Spammers send unwanted tweets to Twitter users to promote websites or services, which are harmful to normal users. In order to stop spammers, researchers have proposed a number of mechanisms. The focus of recent works is on the application of machine learning techniques into Twitter spam detection. However, tweets are retrieved in a streaming way, and Twitter provides the Streaming API for developers and researchers to access public tweets in real time. There lacks a performance evaluation of existing machine learning-based streaming spam detection methods. In this paper, we bridged the gap by carrying out a performance evaluation, which was from three different aspects of data, feature, and model. A big ground-truth of over 600 million public tweets was created by using a commercial URL-based security tool. For real-time spam detection, we further extracted 12 lightweight features for tweet representation. Spam detection was then transformed to a binary classification problem in the feature space and can be solved by conventional machine learning algorithms. We evaluated the impact of different factors to the spam detection performance, which included spam to non-spam ratio, feature discretization, training data size, data sampling, time-related data, and machine learning algorithms. The results show the streaming spam tweet detection is still a big challenge and a robust detection technique should take into account the three aspects of data, feature, and model.

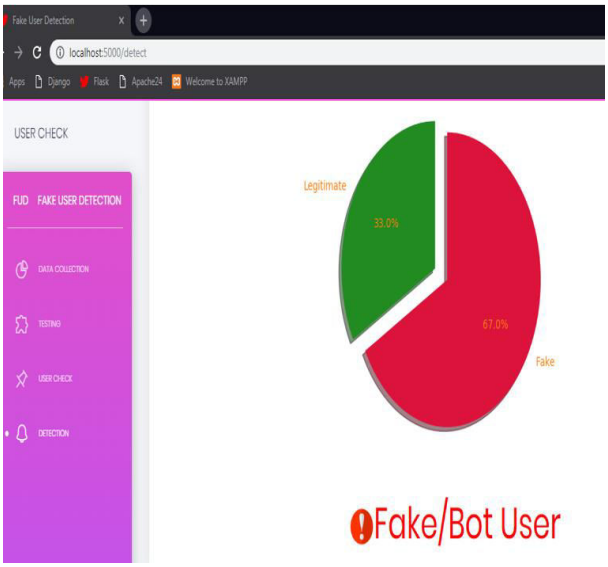
Results:


Data Collection

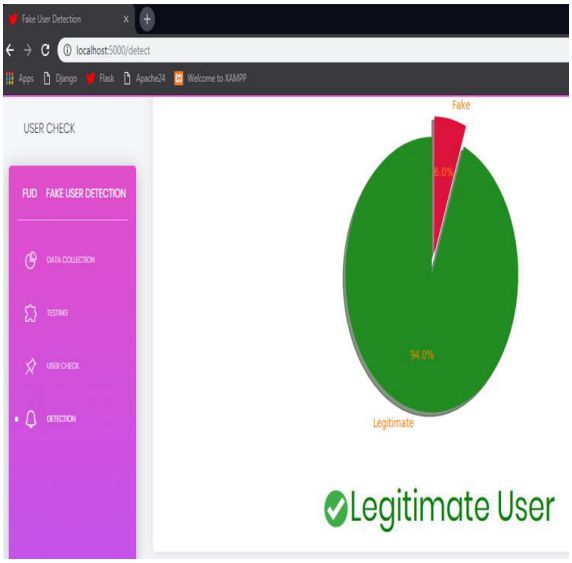
	created_at	username	tweet_id	text	favorite_count	retweet_count	place	lang
0	Fri Dec 06 06:13:35 +0000 2019	AceCarter92	1202833444604170240	RT @2k19Locker: 🚨 #MTP #LockerCode. Use this #..	0	4	None	en
1	Fri Dec 06 00:00:09 +0000 2019	Allsoman	1202739468764520512	scaling material separate off fictional charac..	0	0	None	en
2	Thu Dec 05 20:32:00 +0000 2019	AroynSadik	1202687085724880896	RT @bjone6: Check out my newest #TwitterBot vi..	0	3	None	en
3	Thu Dec 05 12:27:28 +0000 2019	ArvadaDoor	1202565146717016065	If you opening system has failed and you are n..	0	0	None	en
4	Fri Dec 06 06:31:37 +0000 2019	Asuna_RPH	1202837982278901760	#AsunaAbilities (SAO) Acrobatics = 909 / 1000 ..	0	0	None	en
5	Fri Dec 06 02:01:28 +0000 2019	BostonMarketJob	1202769996448821249	Think big, be bold, and succeed with innovatio..	0	0	{'id': '01b347e54132a78d', 'url': 'https://api..	en
6	Fri Dec 06	Click Me	1202833352937681265	#Supernatural Just Proved That Dean And	0	0	None	en

Datasets Trained

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



Fake Users in Twitter Graph



Legitimate User in Twitter Pie Graph

Conclusion

In this paper, we performed a review of techniques usedfor detecting spammers on Twitter. In addition, we also presenteda taxonomy of Twitter spam detection approaches and categorized them as fake content detection, URL based spam detection, spam detection in trending topics, and fake user detection techniques. We also compared the presentedtechniques based on several features, such as user features, content features, graph features, structure features, and time features. Moreover, the techniques were also compared in terms of their specified goals and datasets used. It is anticipated that the presented review will help researchers find the information on state-of-the-art Twitter spam detectiontechniques in a consolidated form. Despite the development of efficient and effectiveapproaches for the spam detection and fake user identification on Twitter [34], there are still certain open areas that require considerable attention by the researchers. The issues are briefly highlighted as under:False news identification on social media networks is an issue that needs to be explored because of the serious repercussions of such news at individual as well as collective level [25]. Another associated topic that is worth investigating is the identification of rumor sources on social media. Although a few studies based on statistical methods have already been conducted to detect the sources of rumors, more sophisticated approaches, e.g., social networkbasedapproaches, can be applied because of their proveneffectiveness.

References

[1] B. Erçahin, Ö. Akta³, D. Kilinç, and C. Akyol, ``Twitter fake accountdetection," in *Proc. Int. Conf. Comput. Sci. Eng. (UBMK)*, Oct. 2017,pp. 388_392.

[2] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, ``Detectingspammers on

Twitter," in *Proc. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf. (CEAS)*, vol. 6, Jul. 2010, p. 12.

[3] S. Gharge, and M. Chavan, ``An integrated approach for malicious tweetsdetection using NLP," in *Proc. Int. Conf. Inventive Commun. Comput.Technol. (ICICCT)*, Mar. 2017, pp. 435_438.

[4] T. Wu, S. Wen, Y. Xiang, and W. Zhou, ``Twitter spam detection: Surveyof new approaches and comparative study," *Comput. Secur.*, vol. 76,pp. 265_284, Jul. 2018.

[5] S. J. Soman, ``A survey on behaviors exhibited by spammers in popularsocial media networks," in *Proc. Int. Conf. Circuit, Power Comput. Tech-nol. (ICCPCT)*, Mar. 2016, pp. 1_6.

[6] A. Gupta, H. Lamba, and P. Kumaraguru, ``1.00 per RT #BostonMarathon# prayforboston: Analyzing fake content on Twitter," in *Proc. eCrimeResearchers Summit (eCRS)*, 2013, pp. 1_12.

[7] F. Concone, A. De Paola, G. Lo Re, and M. Morana, ``Twitter analysis forreal-time malware discovery," in *Proc. AEIT Int. Annu. Conf.*, Sep. 2017,pp. 1_6.

[8] N. Eshraqi, M. Jalali, and M. H. Moattar, ``Detecting spam tweets inTwitter using a data stream clustering algorithm," in *Proc. Int. Congr.Technol., Commun. Knowl. (ICTCK)*, Nov. 2015, pp. 347_351.

[9] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, ``Statisticalfeatures-based real-time detection of drifted Twitter spam," *IEEE Trans.Inf. Forensics Security*, vol. 12, no. 4, pp. 914_925, Apr. 2017.

[10] C. Buntain and J. Golbeck, ``Automatically identifying fake news in popularTwitter threads," in *Proc. IEEE Int. Conf. Smart Cloud (SmartCloud)*,Nov. 2017, pp. 208_215.

[11] C. Chen, J. Zhang, Y. Xie, Y. Xiang,W. Zhou, M. M. Hassan, A. AlElaiwi,and M. Alrubaian, ``A performance evaluation of machine learning-basedstreaming spam tweets detection," *IEEE Trans. Comput. Social Syst.*,vol. 2, no. 3, pp. 65_76, Sep. 2015.