INTERNATIONAL JOURNAL FOR ENGINEERING RESEARCH
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 19th Dec 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=ISSUE-12

DOI: 10.48047/IJIEMR/V11/ISSUE 12/34

Title SPAMMER DETECTION AND FAKE USER IDENTIFICATION ON TWITTER USING MACHINE LEARNING ALGORITHMS

Volume 11, Issue 12, Pages: 253-259

Paper Authors

Eerla Vishwanath, Dr.G.Venkata Rami Reddy

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

SPAMMER DETECTION AND FAKE USER IDENTIFICATION ON TWITTER USING MACHINE LEARNING ALGORITHMS

1 **Eerla Vishwanath,** M.Tech in Data Sciences SIT JNTUH

2 **Dr.G.Venkata Rami Reddy**, Professor of It, M.Tech, Ph.D School of IT, Jntuh

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.

1. INTRODUCTION

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. 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).

Contacts
General Groups
Social Media
Alliances
Pervasive Online Social Networks
Virtual Community
Business Opportunities
Connections
Career
Internet
Relationships

Fig.1: Online social networks

2. LITERATURE REVIEW

Statistical features-based real-time detection of drifted Twitter spam

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.

Automatically identifying fake news in popular Twitter threads

Information quality in social 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 BuzzFeed'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.

A performance evaluation of machine learning-basedstreaming spam tweets detection

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 nonspam 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.

A model-based approach for identifying spammers in social networks

In this paper, we view the task of identifying spammers in social networks from a mixture modeling perspective, based on which we devise a principled unsupervised approach to detect spammers. In our approach, we first represent each user of the social network with a feature vector that reflects its behaviour and interactions with other participants. Next, based on the estimated users feature vectors, we propose a statistical framework that uses the Dirichlet distribution in order to identify spammers. The proposed approach is able to automatically discriminate between spammers and legitimate users, while existing unsupervised approaches require human intervention in order to set informal threshold parameters to detect spammers. Furthermore, our approach is general in the sense that it can be applied to different online social sites. To demonstrate the suitability of the proposed method, we conducted experiments on real data extracted from Instagram and Twitter.

Spam detection of Twitter traffic: A framework based on random forests and non-uniform feature sampling

Law Enforcement Agencies cover a crucial role in the analysis of open data and need effective techniques to filter troublesome information. In a real scenario, Law Enforcement Agencies analyze Social Networks, i.e. Twitter, monitoring events and profiling accounts. Unfortunately, between the huge amount of internet users, there are people that use microblogs for harassing other people or spreading malicious contents. Users' classification and spammers' identification is a useful technique for relieve Twitter traffic from uninformative content. This work proposes a framework that exploits a non-uniform feature sampling inside a gray box Machine Learning System, using a variant of the Random Forests Algorithm to identify spammers inside Twitter traffic. Experiments are made on a popular Twitter dataset and on a new dataset of Twitter users. The new provided Twitter dataset is made up of users labeled as spammers or legitimate users, described by 54 features. Experimental results demonstrate the effectiveness of enriched feature sampling method

3. IMPLEMENTATION

Tingmin*et al.* provide a survey of new methods and techniques to identify Twitter spam detection. The above survey presents a comparative study of the current approaches.

On the other hand, S. J. Somanet. al. conducted a survey on different behaviors exhibited by spammers on Twitter social network. The study also provides a literature review that recognizes the existence of spammers on Twitter social network.

- ❖ Despite all the existing studies, there is still a gap in the existing literature. Therefore, to bridge the gap, we review state-of-the-art in the spammer detection and fake user identification on Twitter

DRAWBACKS:

- ❖ No efficient methods used.
- ❖ No real time data used.
- ❖ More complex

In this paper author is describing concept to detect spam tweets and fake user account from online social network called twitter. To perform detection author is using twitter dataset and 4 different techniques called Fake Content, Spam URL Detection, Spam Trending Topic and Fake User Identification. Using above 4 techniques we can identify whether tweet is normal or spam and then using Random Forest data Mining algorithm we will train above dataset to classify number of spam and non-spam tweets or fake or non-fake accounts. For each technique author is using different data mining techniques to classify tweets as spam or non-spam but here we are using Random Forest classifier.

ADVANTAGES:

- ❖ This study includes the comparison of various previous methodologies proposed using different datasets and with different characteristics and accomplishments.
- ❖ Tested with real time data.

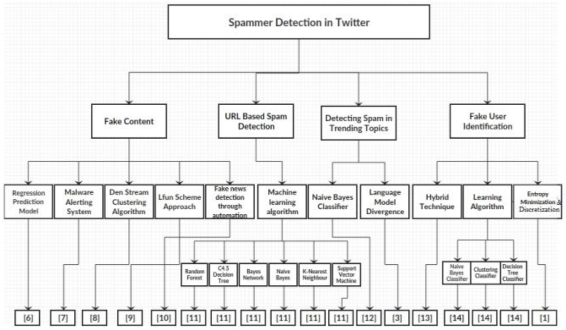


Fig.2: System architecture

Description of 4 techniques to detect tweet is spam or normal.

The presented techniques are also compared based onvarious features, such as user features (retweets, tweets, followers etc.), content features (tweet content messages).

1) Fake Content: If the number of followers is low in comparison with the number of followings, the credibility of an account is low and the possibility that the account is spam is relatively high. Likewise, feature based on content includes tweets reputation,

HTTP links, mentions and replies, and trending topics. For the time feature, if many tweets are sent by a user account in a certain time interval, then it is a spam account.

2) Spam URL Detection: The user-based features are identified through various objects such as account age and number of user favourites, lists, and tweets. The identified user-based features are parsed from the JSON structure. On the other hand, the tweet-based features include the number of (i) retweets, (ii) hashtags, (iii) user mentions, and (iv) URLs. Using machine learning algorithm called Naïve Bayes we will check whether tweets contains spam URL or not.
3) Detecting Spam in Trending Topic: In this technique tweets content will be classified using Naïve Bayes algorithm to check whether tweet contains spam or non-spam words. This algorithm will check for spam URL, adult content words and duplicate tweets. If Naïve Bayes detect tweet as SPAM then it will return 1 and if not detected any SPAM content then Naïve Bayes will return 0.
4) Fake User Identification: These attributes include the number of followers and following, account age etc. Alternatively, content features are linked to the tweets that are posted by users as spam bots that post a huge amount of duplicate contents as contrast to non-spammers who do not post duplicate tweets. In this technique features (following, followers, tweet contents to detect spam or non-spam content using Naïve Bayes Algorithm) will be extracted from tweets and then classify those features with Naïve Bayes Algorithm as spam or non-spam. Later this features will be train with random forest algorithm to determine account is fake or non-fake. All extracted features will be saved inside features.txt file. Naïve Bayes classifier saved inside 'model' folder.

4. ALGORITHMS

ALGORITHMS:

RANDOM FOREST:

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

NAÏVE BAYES:

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

5. EXPERIMENTAL RESULTS

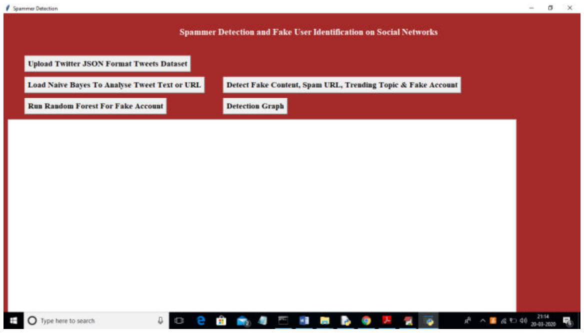

Spammer Detection
Spammer Detection and Fake User Identification on Social Networks
Upload Twitter JSON Format Tweets Dataset
Load Naive Bayes To Analyse Tweet Text or URL
Detect Fake Content, Spam URL, Trending Topic & Fake Account
Run Random Forest For Fake Account
Detection Graph
Type here to search
21:14
20-03-2020

Fig.3: Home screen

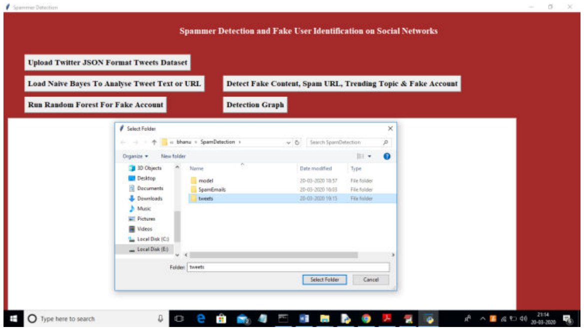
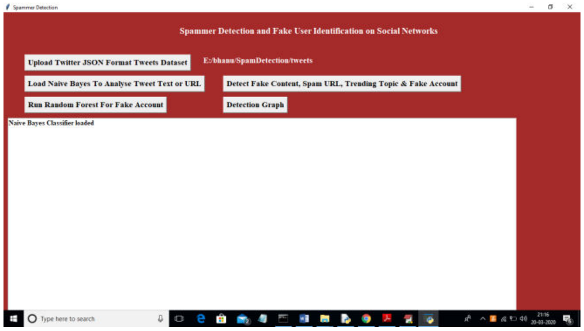


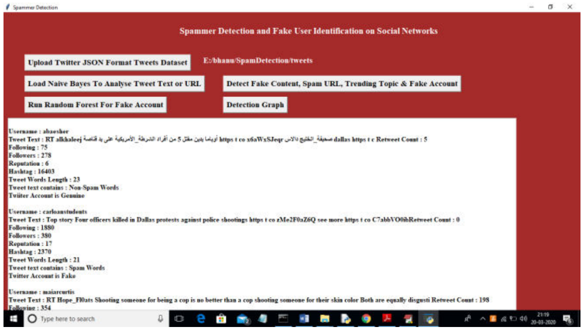
Fig.4: Dataset uploading


Spammer Detection
Spammer Detection and Fake User Identification on Social Networks
Upload Twitter JSON Format Tweets Dataset
Load Naive Bayes To Analyse Tweet Text or URL
Detect Fake Content, Spam URL, Trending Topic & Fake Account
Run Random Forest For Fake Account
Detection Graph
Type here to search

E:/bhanu/SpamDetection/tweets

Naive Bayes Classifier loaded

Fig.5: Load Naive Bayes To Analyse Tweet Text or URL


Spammer Detection
Spammer Detection and Fake User Identification on Social Networks
Upload Twitter JSON Format Tweets Dataset
Load Naive Bayes To Analyse Tweet Text or URL
Detect Fake Content, Spam URL, Trending Topic & Fake Account
Run Random Forest For Fake Account
Detection Graph
Type here to search

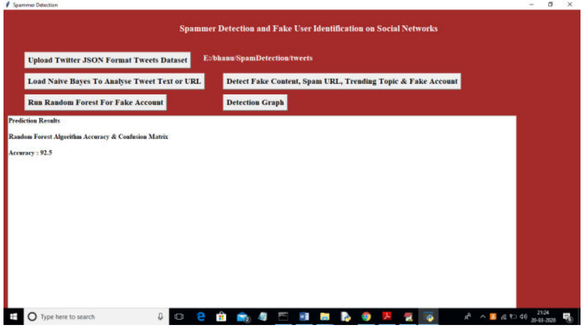
E:/bhanu/SpamDetection/tweets

```
Username : abaesher
Tweet Text : RT alkhaleej أوباما يدين مقتل 5 من أفراد الشرطة_الأمريكية على يد قناصة https t co x6aWxSJeqr صحيفة_الخليج دالاس dallas https t c Retweet Count : 5
Following : 75
Followers : 278
Reputation : 6
Hashtag : 16403
Tweet Words Length : 23
Tweet text contains : Non-Spam Words
Twiiter Account is Genuine

Username : carloanstudents
Tweet Text : Top story Four officers killed in Dallas protests against police shootings https t co zMe2F0aZ6Q see more https t co C7abbVO0ibRetweet Count : 0
Following : 1880
Followers : 380
Reputation : 17
Hashtag : 2370
Tweet Words Length : 21
Tweet text contains : Spam Words
Twitter Account is Fake

Username : maiarcurtis
Tweet Text : RT Hope_Fl0ats Shooting someone for being a cop is no better than a cop shooting someone for their skin color Both are equally disgusti Retweet Count : 198
Following : 354
```

Fig.6: Detect Fake Content, Spam URL, Trending Topic & Fake Account


Spammer Detection
Spammer Detection and Fake User Identification on Social Networks
Upload Twitter JSON Format Tweets Dataset
Load Naive Bayes To Analyse Tweet Text or URL
Detect Fake Content, Spam URL, Trending Topic & Fake Account
Run Random Forest For Fake Account
Detection Graph
Type here to search

E:/bhanu/SpamDetection/tweets

Prediction Results

Random Forest Algorithm Accuracy & Confusion Matrix

Accuracy : 92.5

Fig.7: Random forest prediction

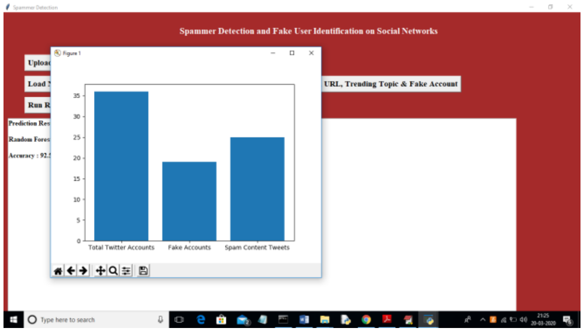


Fig.8: Detection graph

6. CONCLUSION

Using above techniques we can detect whether tweets contains normal message or spam message. By detecting and removing such spam messages help social networks in gaining good reputation in the market. If social networks did not remove spam messages then its popularity will be decreases. Now a days all users are heavily dependent on social networks to get current news and business and relatives information and thus protecting it from spammer help it to gain reputation.

7. FUTURE ENHANCEMENTS

Although a few studies based on statistical methods have already been conducted to detect the sources of rumors, more sophisticated approaches, e.g., social networkbased approaches, can be applied because of their proven effectiveness.

REFERENCES

[1] B. Erçahin, Ö. Aktaş, D. Kilinç, and C. Akyol, ''Twitter fake account detection,'' in Proc. Int. Conf. Comput. Sci. Eng. (UBMK), Oct. 2017, pp. 388–392.

[2] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, ''Detecting spammers on Twitter,'' in Proc. Collaboration, Electron. Messaging, AntiAbuse Spam Conf. (CEAS), vol. 6, Jul. 2010, p. 12.

[3] S. Gharge, and M. Chavan, ''An integrated approach for malicious tweets detection 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: Survey of 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 popular social media networks,'' in Proc. Int. Conf. Circuit, Power Comput. Technol. (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. eCrime Researchers Summit (eCRS), 2013, pp. 1–12.

[7] F. Concone, A. De Paola, G. Lo Re, and M. Morana, ''Twitter analysis for real-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 in Twitter 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, ''Statistical features-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 popular Twitter threads,'' in Proc. IEEE Int. Conf. Smart Cloud (SmartCloud), Nov. 2017, pp. 208–215.