Spam Detection in Twitter Stream

[1]Akanksha. S.Nagdeve, [2]Prof. M.M Ambekar

[1]M-Tech (CSE) Student, [2]Associate Professor
[1,2]Department of Computer Science & Engineering,
P.E.S. College of Engineering Aurangabad, Maharashtra, India

***Abstract*: There are many most popular social media sites in today's world one of them is Twitter. These sites may contain many of the spam tweets on it as it important to detect spam tweets or save us from the spammer ones from their tweets as this means to be identifying the spam fake users who post the spam tweets on the twitter. As there are several methods used to detect spam ones we are using a machine learning method in our work to detect the spam tweets. This method consists of spam detectors module, detectors to detect the blacklist domain in the form of URL's called as blacklist URL's or any other spam tweets which may contain some spammy words by those we can identify the spam tweets this can also help us to detect the trusted user or the spammer user one. As it is difficult to detect the spammer one user as there are many more accounts and there data available on the twitter. As for this we purpose a machine learning method to detect the spam tweets on the tweeter by their tweets using Semi- Supervised and Supervised method in the Machine learning**

***Keywords*: Machine learning method, Semi-supervised, Supervised, Twitter, Detectors.**

INTRODUCTION

When we talk about the online interaction with the world or any new updates that what has been going on in today's world we first take a look on social media. As one of the most popular social media site is Twitter it consists an account which is a user profile on Twitter. An social media site twitter provides most talked topics list which we can called them as Trending topics for the users #(Hash tag) is the term which is used in a particular topic or that is to mention a tending topics. It immediately reflects the events going on in real time. The Users can share their thoughts, photos, news, information, or ideas on it as well as many of the companies also uses Twitter to measure the customer satisfaction for their products as this type of popularity also attracts the spammers .As due to the effective message propagation makes the twitter and most attractive platform. These profiles are usually public - meaning that everyone has access to what users have tweeted or Re-tweeted where the user can share in order to spread advertise to generate sales, propagate pornography, share malicious links in the form of tweets which direct users to malicious software, hijack trending topics for their purposes, abuses reply or mention functions to post unsolicited messages to legitimate users to attract their attention, and phish legitimate users. Twitter users who access through mobile devices should more care about spam than the Twitter users who access through the web browsers since it may consists of excessive amount of personal information such as, call history, user location, bank account details, SMS, calendar events, data located in the device's memory or SD card, premium send -rate SMS messages, capture key-strokes by key logging, and detect user's location via Internet or GPS and share If the twitter is mostly using social media in today's world so we can say it may contain large number of tweets in it as well as in that it may also have many of the spam tweets .A part from the user there are also some users which we can said them as a block users as we can block the users because of the such spam tweets but as spammers can also change their tweet content and strategies to make their tweets and activities like a legitimate. Although for identifying and blocking spammer accounts remain a crucial and challenging task, tweet level spam is essential to fight against the spam tweets as to detection the block spammers who's tweets are fake or when can say that they post a spam tweets which may be in the form of URL's and also it may contain some spammy words but we can identify by working on spam tweets that what spam tweets contains. As identifying the spam users is too crucial and challenging task so first of all more important is to work on spam tweets means to be to detect the spam tweets that how can we found a tweets. We have to deal with all dataset of the tweeter to understand the form of the tweets of the user as there is large amount of dataset present so they are in many of the different forms. By using the term Machine learning we can distinguish between the spam tweets and un-spammed tweets and solve all of the above mention problems related to the social media site twitter by filtering the tweets we can Identify the entire well know Spam tweets.

Figure 1 Twitter views

RELATED WORK

There is a serious problem of Spam on almost all online sites, for this spam detection studies had been done for many years. As different techniques are been used by spammers on different platform so spam detection technique for one platform cannot be used or applicable for other platform. The spam may be in any form just like in the form of URL or text comment in twitter now we have to work for both of the format to detect the spam tweets in the twitter and due to the spam activities makes user trust decrease in the messages distribution and increasing computation overhead. On Twitter many of the people follow spam accounts as they found to be that the news and the comments are quiet interesting and real one. Thomas et al reported that spam targeting twitter is significantly different form the spam targeting email. Twitter consists of different types of spamming activities like link farming, phishing, spamming trending topics, and aggressive posting. Lee et al analyzed tweets content that what exactly the tweets contain and the user behavior that about what the comment has been post to identify content polluters .Hu et al uses social graph tweets of the user to detect the spam tweets as by the graph we can verify the accuracy of the tweets. They show the spam and non-spam pattern of the user tweets based on social graph to detect the pattern of spam tweets. To identify the spam tweets by clustering based on the similarities text and URL we can detect the difference between the spam and No-spam ones. Clustering based method is been used for the detection of spam tweets in twitter based on the large cluster used for the spam tweets. As when we remove spam users from the tweeter or when we block the spammers so it is not important that the spam tweets has also been removed the spam users can also create another twitter account and post spam tweets so it is more important to focus on the spam tweets not on spam users. Castillo et al focus on the trending topics tweets, content behavior and the various links present in the comments so to understand the difference between the non-spam and spam comments. The online learning of many different concept attract spammers towards it as there are many of the users who makes use of online learning concept and this makes spammer to post there fake advertise and spam comments for the user. As we are doing a different work now focus on tweet-level detection of spam comments. For tracking the changing spam activities the semi-supervised is the best machine learning method to be used for the spam detection in the twitter stream.

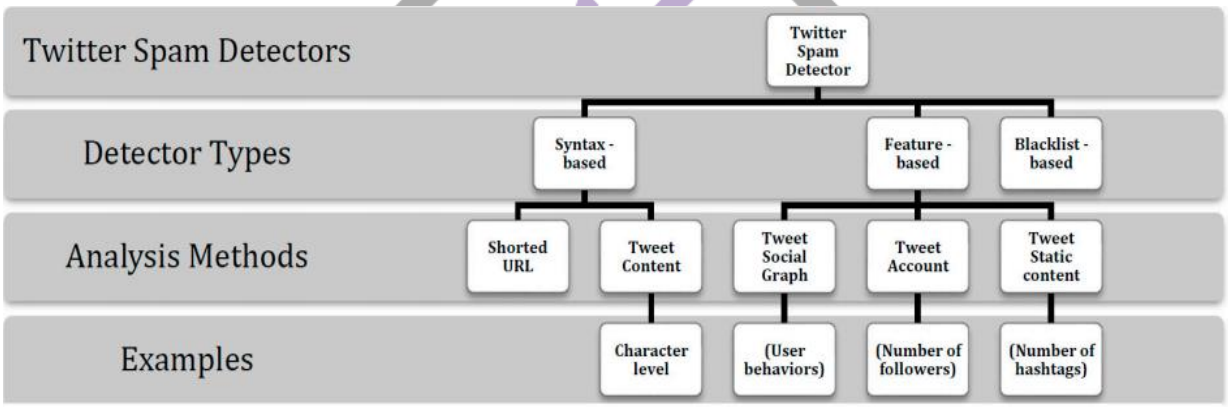

Figure 2: Spam detection in Twitter stream

SPAM DETECTION IN TWITTER STREAM

In this paper we used a machine learing frame work to detect the spam tweets on social networking site Twitter. In an machine learning concept we used Semi-supervised and Supervised machine learning method for spam detection.

This method consists of identifying the blacklist domains in that by using the classification model by this the tweets is lable as spam and non-spam tweets we can also called the non-spam tweet as Ham tweet. Here they uses four types of detctors in it.1) Blacklist Domain Dector - As we know that the fake users promate their products/services by posting a URL links in their tweet as the effective way for the spam detection is to detect the blacklist domains for this fake links posted by the spammers. We utilize the information of domain persent in our dataset to identify whether the tweets are from blacklist domain or not.As they may in the form of URL or text in the given tweets of the spammers we can take a look form the dataset to identify the truth or for the spam detection.

2) Near- Duplicate Detector: There are many more tweets present on the twitter and they may be near-duplicates which are assigned as the same lables of spam and ham tweet the duplicate tweets can be detected using the Minhash algorithn in it by concatenating the three minimum hash values which are computed form the tweet's unigram,bigram and trigram representations . If there are two or more than two tweets are similar than they should be considered as a duplicate tweets and if the near-duplicates are formed in cluster having the same signature label as spam or ham than the new tweet of same signature also receives the same label.

3) Reliable Ham Tweet Detector: On the twitter there are many of the tweets present which can be also considered as a ham tweets however for the spammers after gaining the response or we can say acceptance form the other users they post more spam tweets. We can consider a tweet is not spam or ham tweets by statisfying these two conditions a) The tweets on the twitter must not contain any of the spammy words in it which can indicate that that the tweet is spam. b) And the second one the posted tweet must be posted by the trusted user.Spammy words are the words which is used in the spam tweets and which indicates that the tweet is spam but the word which is used in the spam tweet can also be available in ham tweets. So for this we can find it by the total number of spam tweets containing the spammy words and the number of ham tweets conatining the spammy word we can find the probalibility of the trusted user who can never post the spam tweet which the user post at least 5 to 6 confident tweets.Classification can be used to

classify the spammy words in the given tweets for spam and ham tweets the spam tweet which contain spammy words in it consider as a spam tweets on twitter for the clusters we can also predict them in the form of clusters of spam and ham cluster by using the multiple classifier.
4) Multiclassifier-Based Detector: Tweets that are not been labeled in previous steps are labeled and processed in this step. In this detector we develop a spam detector by using efficient namely random forest (RF). The classifiers is based on decision tree-based classification models. As below there is a table which consists of some different types of tweet which are present on the twitter it includes the spammy words in the different form of sentence in it there may a type of word called as categorical words used in the tweets which are categories in the top- level tweets.

Tweets which are labeled by the first three detectors (i.e., blacklisted domain, nearduplicate, and reliable ham tweet) they are considered as confidently labeled tweets. For the classifier based , that we use classifiers based on a different classification technique compare to other. Tweets are labeled as spam by above classifiers are considered as confidently labeled spam tweets. Similarly, tweets that do not contain any spammy words are labeled as ham by the above classifiers are confidently labeled ham tweets. An confidently labeled spam tweets are considered which are utilized for the given blacklist domains for spam dection , and confidently labeled ham tweets are utilized to identify trusted users. For the duplicate tweets that do not match prelabeled clusters but having the same signature are grouped into a new cluster, i.e., nearduplicate tweets is the collection of each cluster.Next, if there are 10 tweets groups we label them as a clusters of high confidence, then the signatures of these newly labeled confident clusters will be used by the near-duplicate detector in the next time window. We are using the Porter Stemmer Algorithm in this paper which plays an important role in the machine learning method of spam detection An Porter Stemmer works as a mining of attitudes, views, opinions from the text and tweets with the help of the NLP (Natural Language processing) as it classifies the text in the form of 'positive' or 'negative' term. This has been used in the term of spam detection in twitter stream by analyzing or we can say by mining the tweets on the twitter it classifies in the form of 'True' or 'False' for the given spam or ham tweets on the twitter by using this algorithm we can find the spammy words being used in the text or URL tweets in the twitter and we can also calculate the accuracy of the given spam tweets that about how much it is been said to as spam tweet during the time of the result of spam and ham tweets on the twitter. As there are many researchers to make research on the spam problems in today's world as this is the one of the method being used to solve the related spam problem for the one of the most popular site that is twitter.

Some types of spam comments on Twitter.

Your free ringtone is waiting to be collected. You can text password \XYZ\" on 123 to verify. Get Usher and Britney. FML
Free Msg Hello there it's been 4 week's now! I'd like to be have fun still? Tb ok! XYZ had charges you to send, å£1.90 to rcv
WINNER!! As a valued network customer you have been selected to receivea å£900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
Had your mobile 11 months or more? U R had a offer to Update to the latest mobiles of different colors with camera for Free! Call for the FREE mobile update on 08455667733
SIX chances to win CASH! From 500 to 30,000 pounds txt> CSH123 and send to 83535. Per day cost 200 rs, 6days, 16+ TsandCs apply Reply HL 4 info
SMS. ac Sptv: The New Jersey Devils and the Detroit blue Wings play Ice Cricket. Correct or Incorrect? End? Reply END SPTV
Please make a call to our customer service representative on 08006007000 at time between 09am-10 pm as you have WIN guaranteed å£5000 cash or å£10,000 prize!
PRIVACY FOR YOUR ACCOUNT ! 2007 Account Statement for 07443355600 shows 788 Bonus Points. To claim you can call on 08100200555 Identifier Code: 20044 Expires
Are you unique enough? Find out from 30th August. www.areyouunique.co.uk
Will u meet ur dream partner soon? Is ur career off 2 a flyng start? You can find it out free, txt STAR followed by ur star sign, e. g. STAR ARIES

CONCLUSION

In this paper, we propose a semi-supervised spam detection framework, which utilizes four lightweight detectors for the spam dection in twitter stream. The experiment demonstrate the effectiveness of the given result approach of semi-supervised in our spam detection framework. By using the PorterStemmer algorithm we found that by mining on data makes a easy concept for the detection of spam tweets in twitter thus it also consists of confidently labeled clusters and tweets make the system effective in capturing new spamming patterns can also be represented by the social graph for spam detection in twitter.Tweet-level spam detection approach is a very fine grained on the real time system ones However,only limited information can be obtained for the given tweets We believe that tweet-level spam detection complements user-level spam detection. Due to the limited user information in our data set, we have used the simple technique to deal with user-level spam detection.

REFERENCES

1) I. Santos, I. Miñambres-Marcos, C. Laorden, P. Galán-García, A. Santamaría-Ibirika, and P. G. Bringas, "Twitter content-based spam filtering," in Proc. Joint Conf. (SOCO-CISIS-ICEUTE), 2013,
2) I. Santos, J. Nieves, and P. G. Bringas, Int. Symp. Distrib. Comput. Artif. Intell., 2011.
3) F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on Twitter," in Proc. CEAS, 2010, p. 12.
4) E. Tan, L. Guo, S. Chen, X. Zhang, and Y. Zhao, "Unik: Machine learning method unsupervised spam detction," in Proc. CIKM, 2013.
5) . Thomas, C. Grier, J. Ma, V. Paxson,Spam filtering service on real time," in Proc. IEEE Symp. Secur. Privacy, May 2011, pp. 447–462.