IJCT

INTERNATIONAL JOURNAL OF COMPUTER TECHNIQUES

TWITTER SPAM DETECTION BASED ON SPOT ALGORITHM USING COUNT THRESHOLD AND PERCENTAGE THRESHOLD

R. SUGANYA[1], Dr.K. MUMTAZ[2]

[1] Research Scholar, Dept of Computer Science, ARIGNAR ANNA GOVT. ARTS COLLEGE FOR WOMEN, Walajapet

[2] Assistant Professor, Dept of Computer Science, ARIGNAR ANNA GOVT. ARTS COLLEGE FOR WOMEN, Walajapet

--***--

Abstract - Social media networks such as Twitter, Facebook, Instagram, LinkedIn allow users to freely distribute and share information to friends. People like to be in touch with the online platforms all the time. Twitter founded in 2006 has been one of the most popular microblogging service sites. The crucial problem in twitter is spam and spammers. The maximum limit placed on the number of characters in a tweet is 280. The proposed system uses SPOT to detect the spam messages. SPOT is used to monitor the spammers activities by tracking the IP address of the sender machine. The messages send by the sender are classified as either spam or non-spam by using GABOR filtering. In the content-based filtering the GABOR filtering is used to compare the tweet with the dataset. The dataset contains the spam user address or link. GABOR filtering filters the related terms since they are considered spam messages. The system status is classified as Compromised and Non-Compromised. Compromised represents the tweets which includes spam messages. The tweets without spam messages are called as Non-Compromised. Moreover, Count Threshold (CT) is used to count the total number of spam messages and Percentage Threshold (PT) is used to compute the overall percentage of the Spam messages. The CT calculates the spam messages which are send by the sender machine. Taking the threshold value to be 3, CT computes the spam messages and SPOT identifies the spammer and blocks it within the time duration. The system is implemented using MATLAB and NetBeans.

Key Words: Data Mining, Social Media, Twitter Spam Detection, SPOT Detection System, Count-threshold, Percentage-threshold.

I. INTRODUCTION

Data mining is the process of analysing a large set of data to find new, exact and hidden information. It is the process of discovering large volume of interesting data sets that involving machine learning statistics and database system etc. Data Mining is an ambiguous term that has been used to refer to the process of finding interesting information in large repositories of data. More precisely, the term refers to the application of special algorithm in a process built upon sound principles from numerous disciplines including statistics, artificial intelligence, machine learning, database. (Han and Kamber, 2011). Mining includes text mining, web mining, spatial mining, multimedia mining, biological mining. Most of the people used text messages for communication purpose in social networking sites. Social media is the biggest platform for gathering data. Data mining concept is analysing and extracting data from the data source. Twitter is one of the most famous communication platforms in worldwide. All the user's messages are in public. Twitter data is a rich source for text mining because all the messages are in text format. Gathering the data can be done by specific keywords, sentiment measuring, feedback, review about the products etc.

Spam is Unsolicited Commercial Email. Without getting permission from the user, unauthorized persons can send the spam messages to the user account. Spam is anywhere or everywhere in internet-based communication. Spammers mostly use the Internet-based communication to send the advertisement without user knowledge. Spam are found in instant messaging, text messages in mobile, group discussion and all the text communication medium. In Social networking, spams play a major role in twitter messages. In twitter, messages are called as tweets. Spammers use twitter account for sending advertisement, harmful links, unwanted posts, virus link, creating duplicate accounts, posting repeatedly trendy topics, unrelated tweets, tweet jacking etc., In Real time communication SPIM is a one type of spam. SPIM stands for Spam Over Instant Messaging. Spammers use the public profile to send or receive the unsolicited messages. Instant messaging spam are called as Spimmers. The main purpose of SPIM message is sending the advertisement, hyperlink etc.,

The main objective of the research is to eliminate the spam messages in a tweet message using SPOT algorithm. SPOT algorithm is used to improve the traditional Naïve Bayesian classifier to classify the tweets. The main focus of this research is to identify the spam messages, using the CT and PT values and filter them using GABOR filtering.

II. RELATED WORK

Opinion reviews are important source in online product. Not even virus harm full links, fake reviews are also called spam message. Because it may affect the production or productivity and original customer feedback about the product. Supervised classification method evaluates and detect the spammers automatically. Loopy Belief Propagation algorithm and kernel density estimation technique are used to identify the fake review on online shopping (Arjun Mukherjee et. al, 2011). The author describe by what method can all the more likely email worms be made? By what means would spyware be able to assert more casualties? By what method can all the more likely spam be sent? Incidentally, these three inquiries are altogether related. The author takes a gander at the future risk presented by email worms, spyware, and spammers that utilization zombie machines in another way: modern information mining of their casualties' spared email. The end result skirts numerous barriers, and could trap even experienced clients. More successful spam can be sent by utilizing malware on zombie machines to mine information from email corpora. This enables spam to be sent that naturally emulates honest to goodness email sent by the genuine proprietors of the zombie machines, and confirmation of-idea execution exhibits that the outcome can be persuading even to prepared clients. While this more viable spam has not, as far as anyone is concerned, been found in the wild, there are protective advances that can be taken currently to confine its effect when this spam makes its presentation (J. Aycock and N. Friess, 2006).

Twitter has become a target platform on which spammers spread large amounts of harmful information. These malicious spamming activities have seriously threatened normal users' personal privacy and information security. An effective method for detecting spammers is to learn a classifier based on user features and social network information. However, social spammers often change their spamming strategies for evading the detection system. To tackle this challenge, latent user features factorized by text matrix are adopted to capture the consistency of users' behaviour. Also, a new social regularization based on users' interaction is introduced to distinguish different types of users. Finally, Supervised Spammer Detection method with Social Interaction is proposed, which jointly learn a classifier by using combine text content, social network information and labelled data. Experimental results on a real-world Twitter dataset confirm the effectiveness of the proposed method (F. Benevenuto et. al, 2010). The most unsolved problem in twitter is spam detection. Machine techniques are used in this paper to detect the spam in twitter. Machine Learning (ML) based detection involved in multiple steps. Tweets are split into spam and non-spam in twitter. Using the labelled samples, trained machine learning classifier can detect the spam. Supervised classification algorithm train it with 10k spam and 10k non-spam tweets per day. “Spam Drift” problem is solved by Lfun approach. Lfun scheme, added “changed spam” to reduce the “spam drift”. It improves accuracy in real world but it is time consuming and very expensive (Chao Chen et. al, 2017).

Twitter is using Google Safe Browsing to detect and block the spam links. The twitter has changed the way of communication in the daily life of the people. Despite that blacklists can block malicious URLs embedded in tweets, their lagging time hinders the ability to protect users in real-time. The author describes in the paper the different machine learning algorithm used to detect the twitter spam. The performance of the real-time twitter spam detection is the lack of messages. The large dataset of over 600 million public tweets are taken. The labelled tweets around 6.5 million spam tweets and extracted 12 light weight features used for online detection. (C. Chen et. al, 2015).

III. PREVIOUS IMPLEMENTATIONS

Network of spam zombies are diagnosed as one of the most serious security issues. Many types of algorithms used for spam detection in twitter. The need of this research is to present a new SPOT algorithm for detecting online spams in twitter. The study focuses on discovery of tweet spam using the spam detection technique. The tweets are classified as either spam or non-spam by GABOR based spam filter. SPOT detection algorithm takes the sender machine IP address and sends only the non-spam messages to the receiver. Characters are filtered by the keyword searching with the twitter character restriction count. Two different algorithms in detecting spam zombies are CT and PT. CT (Count Threshold) is based on the number of spam messages and PT (Percentage Threshold) is another algorithm for percentage of spam messages sent from sender side machine. The algorithm identifies which message is Spam and SPOT works to identify the IP address and block the spammers and eliminate the spam. The spam rates are mentioned in the statistical tool review graph model. SPOT is an effective and efficient system in automatically detecting spammers in a network.

In social networking, online spam detection is the most difficult problem because day by day more people are using the social media and spammers are also increased. In twitter, daily huge number of tweets is shared. In twitter approximately 25,000 spam tweets occur per day. A social tweet message is seen by all the followers of the users. The traditional character restriction for twitter is 140 characters per message. The aim of this system, is to use SPOT algorithm to remove the spam content from the messages and filtering the spam and non-spam messages.

Previous research work has been done in both machine learning and non-machine learning approaches for detection of spam. In the machine learning approach unified model filtering was used to classify received tweets as spam or non-spam. Some of the machine learning techniques have used were Artificial Neural Network (ANN), K-nearest, Naive Bayesian (NB) and Support Vector Machine (SVM). Content-based spam filtering technique was used to filter the spam messages. The main focus is detecting the spammer. A major security challenge on the internet is the large number of spammers developed in social communication. Security attacks include spamming, identity theft etc. Many existing systems are less effective and easily hacked by spammers.

The proposed system aims to delete tweet with attached viruses. It detects and blocks the spammers by using a SPOT detection algorithm. The SPOT algorithm detects the spam zombies and also detects the spam messages. The account reactivation test is provided by the system. The system receives a tweet message. The filter checks for virus in the attachment. The SPOT algorithm deletes tweets having spam in an attachment. If the spam is not found, then the tweet is a secure tweet. The research applies SPOT detection algorithm to detect each and every tweet for spam and uses filter to classify the spam. Count Threshold (CT) is used to count the total number spam tweets. Percentage Threshold (PT) is used to identify the overall percentage of the spam messages in the sender machine. Spam is Unsolicited Commercial Email. Without getting permission from the user unauthorized persons can send the spam messages to the user account. Spam is anywhere or everywhere in internet-based communication. Spammers mostly use the Internet-based communication to send the advertisement without user knowledge. Spam are found in instant messaging, text messages in mobile, group discussion and all the text communication medium. In Social networking, spams play a major role in twitter messages. In twitter, messages are called as tweets. Spammers use twitter account for sending advertisement, harmful links, unwanted posts, virus link, creating duplicate accounts, posting repeatedly trendy topics, unrelated tweets, tweet jacking etc.,

In Real time communication SPIM is a one type of spam. SPIM stands for Spam Over Instant Messaging. Spammers use the public profile to send or receive the unsolicited messages. Instant messaging spam are called as Spammers. The main purpose of SPIM message is sending the advertisement, hyperlink etc.,

IV. SYSTEM IMPLEMENTATION

A network analysis-based system is used for spam filter in Twitter. By analysing the network and relations between senders and receivers, this spam filter does not require large data collection up 280 Characters, in real-time spam detection, there is some comparability between the spams on Twitter and other traditional spams on email, such as sending advertisement, tracking user account information and spreading spam messages. Twitter have its own characteristics: The limitation of each tweet length is 280 characters, which is usually not long enough for spammers to put in the desired spam information. Using the character restriction spammers usually add a short length spam messages attached with tweets. Due to the lack of incomplete information or even the spurious information in a spam tweet's messages, most of the users are harmed. In previous study describes 45% of the users in social medium ready to click the links sent by friends or in the friend list, even though the user knows the person or they don't know in the real life

DATA SET COLLECTION

To properly evaluate the spam filter and spam detection in twitter, here utilized an existing Twitter network dataset is used. This offline dataset contains 81,306 users and 1,768,149 follow relationships formatted as a directed Email. However, this dataset is unlabelled and was collected from over 2 years ago. Therefore, the collected information contains the Twitter users using Twitter API.

DATA LABELLING

In order to correctly label nodes in the dataset, follower-friend ratio is used to rank all nodes in the network. Starting from the lowest-ranked node, it manually selected 100 spammer nodes by checking out their Twitter profile and tweet history. The 3,589 deleted / suspended accounts cannot be labelled as spammers because they were not able to be verified as true spammers. Out of the set of all users followed by these 100 spammers, it randomly selected 100 nodes who are considered as legitimate users being spammed by the spammers. Lastly, out of the set of all users following these 100 normal users, randomly another 100 nodes are selected as control set for comparison purpose. It manually verified that these randomly selected 200 legitimate users are indeed not spammers. The SPOT algorithm is used to separate the messages as spam or not by identifying the sender messages.

MESSAGE HANDLING

In most current twitter tweets implementations, the user has almost no control over which tweet messages enter and exit their twitter account. A limited amount of control is sometimes provided through the use of filters or rapidly clicking on the delete key. Many available filters are either hard coded rules, or simple pattern matchers, directing messages to specific folder destinations. This approach as a simple text classification problem do not consider the overall picture of the user's usage of the message system. In addition, some of the filters operate on a rule-based system and need to be frequently updated with new rules to remain effective.

The following features can all be used alone or in combination to constrain the data view.

- Date - All messages between a set of dates.
- User, Direction - We can choose a specific user to view all their email, and also define which direction (inbound, outbound, or both) we would like to view.
- Label - We can view specific emails, such as spam or virus.
- MYSQL - An MYSQL statement can be defined to specifically choose a subset of the data. This allows users to extend the schema and use those extensions within the SPAM framework. Views exist to allow the system to scale to arbitrarily large amounts of messages without taking over all of the system resources. The message window also allows old views to be viewed by using a back and forth button near the top of the GUI.

Machine learning approaches have succeeded in text categorization problems, these techniques have been adopted in spam filtering systems. GABOR filters seem well message in document text and character analysis. But to succeed this

Spam message, it is necessary to make a good parameterization of these filters. The filtering work is derived and it extracts the GABOR features, it is important to calculate them from each GABOR response matrix. So, each binary value in the input tweet will be characterized by 280 values.

Fig 1 shows the steps carried out in filtering. Text filtering is always based on content-based filtering method. Machine learning is classified as supervised machine learning and unsupervised machine learning. Supervised learning is done in the context of classification. Unsupervised learning is clustering, representation learning and density estimation. GABOR filtering is under the supervised machine learning.

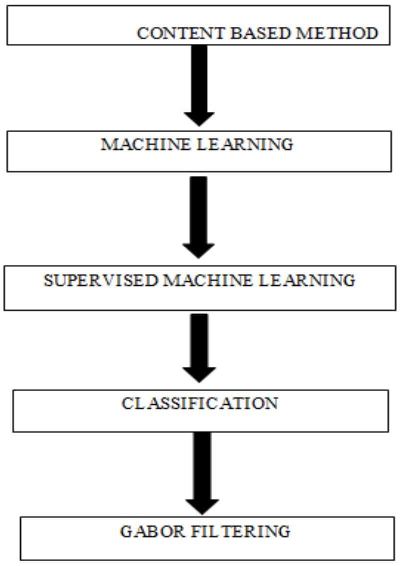


Fig. 1. Filter

GABOR Filtering

Message queries are represented as vectors.

$$d_j = (w_{1j}, w_{2j}, ...w_{tj})$$

$$q = (w_{1q}, w_{2q}..... wnq)$$

Each twitter can be corresponding to a separate term. If a term occurs in the tweet, its value of the vector is non-zero.

There are many ways of calculating these values, also known as (term) weights, have been developed. For a given Message want to find similar twitter message using cosine as in vector space model

$$d1{\cdot}d2/(\|d1\|\|d2\|)$$

tf have been normalized using augmented frequency, to prevent a bias towards longer Message as in this tf-idf:

Have pre-calculated all $\|d\|$ Have the values for the denominator pre-calculated So for a given $d1$ need to score over 1 million $d2$ Have a threshold of 0.6 cosine for similarity.

I can observe that for a given $\|d1\|$ there is a fairly narrow range of $\|d2\|$ for cosine ≥ 0.6

For example in one search for similar for a cosine of ≥ 0.6 and a $\|d1\|$ of 7.7631 then $\|d2\|$ range from 7.0867 to 8.8339

Where outside the threshold of cosine 0.6 $\|d2\|$ range from to 0.7223 to 89.3395 This was with standard tf Message normalization It is looking at a LOT of $\|d2\|$ that don't have a chance of being a cosine 0.6 match For the purpose assume the Message is normalized on raw tf Sorry but I am just not good with mark-up is used to make the equations So in my notation

||d1|| = sqrt(sum(w1 x w1))

d1 dot d2 = sum(w1 X w2)

Assume d1 is the shorter Message The very best d1 dot d2 that can be achieved is d1 dot d1If d1 is marry 100 paul 20 And d2 is marry 100 paul 20 peter 1 Normalized d1 is marry 1 paul 1/5 d2 is marry 1 paul 1/5 peter 1/100 Clearly marry and paul have the same idf in both Twitter Message

The best possible d1 dot d2 is d1 dot d1

The maximum possible match to d1 is d1

cos = d1 dot d1 / ||d1|| ||d2||

square both sides

cos X cos = (d1 dot d1) X (d1 dot d1) / ((d1 dot d1) X (d2 dot d2)) cos X cos = (d1 dot d1) / (d2 dot d2)

take the square root of both side

cos = ||d1|| / ||d2||

is ||d2|| not bounded by the cos?

If I just use ||d2|| >= cos ||d1|| and ||d2|| <= ||d1|| / cos I get the computational speed I need

- Simple model based on linear algebra

- Term weights not binary
- Allows computing a continuous degree of similarity between queries and message
- Allows twitter message according to their possible relevance
- Allows partial matching

In the proposed system, GABOR filters have five different values for Spatial frequency (f = 0.0625, 0.125, 0.25, 0.5, 1.0) and seven different values for orientation θ = (0, 30, 60, 90,120, 150, 180) are chosen, the total combinations are giving in the GABOR filtering is 35. The output of each GABOR filter, Spam and Real Absolute part are computed and then each part of mean and standard deviation is calculated, which assist as GABOR features. Thus, for each character to get a feature vector of dimensionality 280.

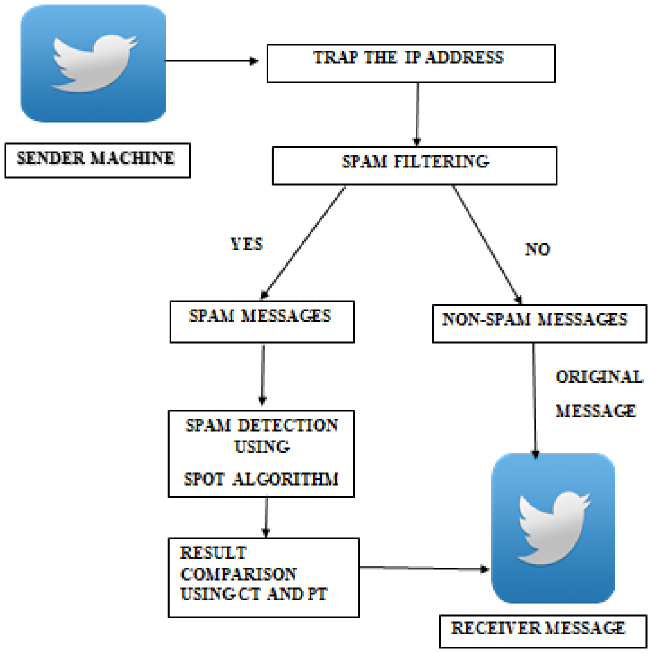


Fig. 2. System Architecture

The architecture fig 2 describes the work follow of the spam detection based on twitter. In step 1 SPOT algorithm accept the sender messages and tracking the IP address. The step 2 classify the spam and non-spam messages using spam filtering. In step 3, spam messages in the sender IP address are blocked using the SPOT detection algorithm and the spam messages are deleted**.**

SPOT Algorithm

```
Step 1. An outgoing message arrives at SPOT
Step 2: Get IP address of sending machine m
// all following parameters are specific to machine m
Step 3: Let n be the message index
Step 4: Let Xn = 1 if message is spam, otherwise Xn = 0 otherwise
Step 5: if (Xn = = 1) then
//spam, 3
Step 6: Λn += ln (θ1/ θ0)
Step 7: else
// non-spam
Step 8: Λn += ln ((1- θ1) / (1- θ0))
Step 9: end if
Step 10: if (Λn>= B) then
Step 11: Machine m is compromised. Test terminates form.
Step 12: else if (Λn< = A) then
Step 13: Machine m is normal. Test is reset for m
Step 14: Λn = 0
Step 15: Test continues with new observations values
Step 16: else
Step 17: Test continues with an additional observation
Step 18: end if
```

SPOT is intended supported the statistical tool SPRT. Eliminating the content affected by spam tweets in SPOT, H1 is a detection parameter and H0 is an original parameter. If H1 is true it is compromised and H0 is true it is not compromised. Xi=1 i^{th} tweet attached with spam messages and Xi=0 is non-spam messages. Λn for every IP address of the sender machine Parameters are passed by the user namely, α, β, 1, and 0. α are the desired false positive rate and β are desired false negative rates. The tweet attached with spam message when H1 is true denotes 1, the tweet without the spam H0 is true denotes 0. The value of the parameter α=0.01, β=0.01, θ1=0.9, and θ0=0.2To assign the value to θ1 and θ0 for false positive rate of the spam filter. The high detection rate and low false positive rate in spam filter the values of 1 and 0 assigned closed to the true probability.

The sender machine IP address is noted and also the messages are assessed as either spam or non-spam by GABOR filtering. For every noted IP address, SPOT maintains the logarithm value corresponding probability ration supported the relation between Λn, A and B, the rules determine the corresponding machine is compromised, traditional, or can’t be reached. First step of the algorithm, it checks the outgoing tweet messages from the sender message. The outgoing messages reaches the SPOT detection system, it tracks the sender IP address. Second step of the algorithm is classified the spam and non-spam by the spam filter technique. The logarithm value of the probability ratio is n, Xn=1 means it consider as spam messages and Xn=0 means it is considered as non-spam messages. The relation between the n, A and B algorithm show the machine is compromised, normal or the decision cannot be reached.

The smaller value of 1 and 0 it required a larger number of SPOT observe the reach detection. The value 1 indicates the compromise machine and 0 indicate the normal machine. The positive and negative rate of the spam filter together to detect the spam tweets.

Machine is identified by SPOT, the records of the machine in SPOT are reset so that a new monitoring phase starts for the machine. It will define the overall SPOT operations

$$Pr\,(X_i=1/\,H_0) = 1\text{-}Pr\,(_{Xi}=\,0/\,H_0) = H_0$$

$$Pr\,(X_i=1/\,H_1) = 1\text{-}Pr\,(X_i=\,0/\,H_1) = H_1$$

For any positive integer n=1, 2…

$$x(n) = \sum_{n=1} \left(\frac{Pr(x1, x2 \ldots \ldots . . \frac{xn}{H0})}{Pr(\, x1, x2 \ldots \ldots . \frac{xn}{H1})} \right)$$

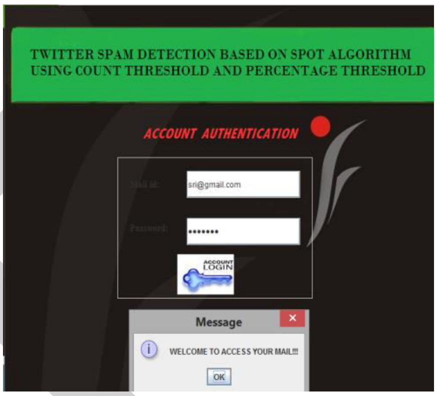
Within these two lines all the SPOT algorithm work can be done. H1 denotes the compromised machine and H0 denotes the normal machine. ith message indicate the network spam, Xi=1 represent the tweet with spam messages and Xi=0 is message not attach with any spam. Using the GOBAR filtering is classified either spam or non-spam messages. Λn denotes the IP address of each sender machine.

V. COMPUTATIONAL RESULT

MATLAB is one of the high-level techniques for computing language, developing algorithm, data visualization, data analysis and numerical calculation. Compared to the classical programming languages MATLAB is a powerful visualization tool in support of matrices and matrix computation. MATLAB have the excellent graphic visualization capacity. In MATLAB have some different parameters like a string, a graph, a string etc.,

The standard datatype in MATLAB is matrix, all the MATLAB data are considered as matrices format by some order. MATLAB main support the GUI (Graphical User Interface), Image processing, visualization, algorithm is developed by graphical tool. Using custom plots, the visualizing data are built.

Authentication is one of the process of cybersecurity. Authentication is the process of verifying someone entry in the device or user account. In all communication medium, first security is stop and not allowed the unauthorized person to access the other user profile. The system verifies the user name and password for security issue. The first step in network communication is checking the valid user. Fig 3 shows the account authentication images, which it has two fields for validation and verification, namely mail id and password. The user account is activated only when twitter user mail id and password are validated. It shows the message box with the content access denied and stop the person to enter into the system.


TWITTER SPAM DETECTION BASED ON SPOT ALGORITHM
USING COUNT THRESHOLD AND PERCENTAGE THRESHOLD

ACCOUNT AUTHENTICATION

Mail id: sri@gmail.com

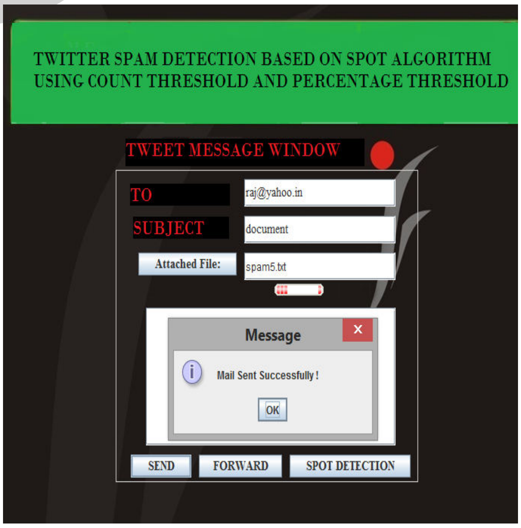
Password: •••••••

Message

WELCOME TO ACCESS YOUR MAIL!!!

Fig. 3. Account Authentication

Fig 4 is the twitter message window which shows the tweet messages send by the sender to single or multiple receiver. This window shows the to address, which must be in valid form. The subject describing the tweet attachment content. The Attached file used to attach the tweet messages in the link. Twitter having the character restriction limitation for each tweet messages as 280 characters. Use three options buttons they are send, forward and SPOT detection. 11111111111111111Forward options are used for sending same attachment to multiple receivers. SPOT detection option is used for detecting the spam.


TWITTER SPAM DETECTION BASED ON SPOT ALGORITHM
USING COUNT THRESHOLD AND PERCENTAGE THRESHOLD

TWEET MESSAGE WINDOW

TO raj@yahoo.in

SUBJECT document

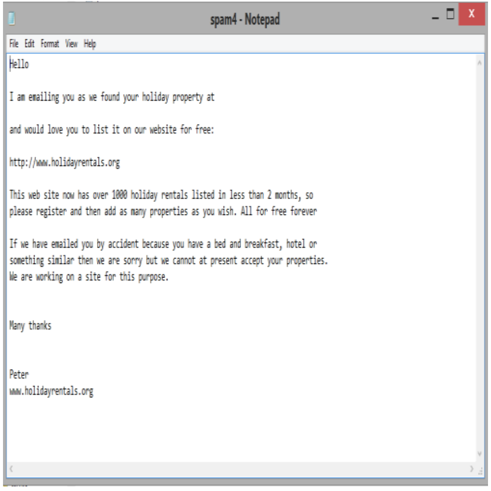
Attached File: spam5.txt

Message

Mail Sent Successfully !

Fig. 4. Tweet Message Window

Open the one set of datasets, the tweets are like in the fig 5. Within this tweet, the sender original tweets and additional spam tweets are also occurred. The length of the messages is restricted. If the message length exceeds 280 characters, the entire message is considered as spam.



Hello

I am emailing you as we found your holiday property at

and would love you to list it on our website for free:

http://www.holidayrentals.org

This web site now has over 1000 holiday rentals listed in less than 2 months, so
please register and then add as many properties as you wish. All for free forever

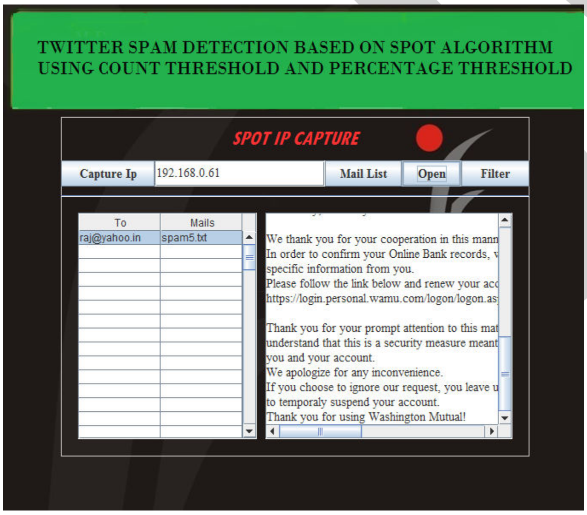
If we have emailed you by accident because you have a bed and breakfast, hotel or
something similar then we are sorry but we cannot at present accept your properties.
We are working on a site for this purpose.

Many thanks

Peter
www.holidayrentals.org

Fig. 5. Tweet Content

Compare to the all algorithm SPOT is the best algorithm for detecting spam on twitter. SPOT main purpose is to identify the sender IP address. One’s the message received by the spam it checks the IP address and record it. After the classification of the spam and non-spam messages, the SPOT algorithm verifies with the spam messages, and where the spam message comes from which IP address. Identify the spammer IP address and block the spammer. It shows the sender capture IP. The mail list lists out the number of messages send to the various receivers along with their mail id. Open option is used to see the content of the message


TWITTER SPAM DETECTION BASED ON SPOT ALGORITHM
USING COUNT THRESHOLD AND PERCENTAGE THRESHOLD
SPOT IP CAPTURE
Capture Ip
Mail List
Open
Filter

192.168.0.61

To	Mails
raj@yahoo.in	spam5.txt

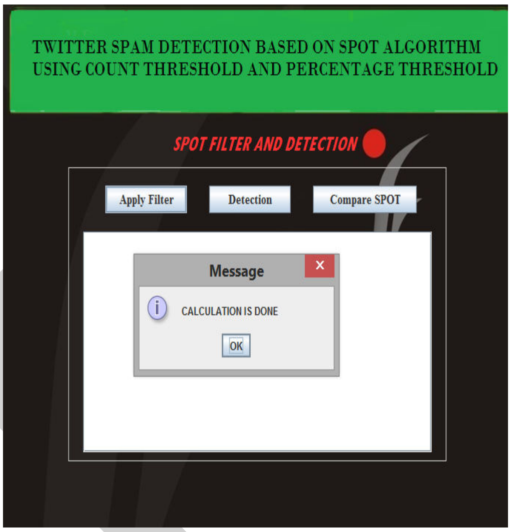
We thank you for your cooperation in this mann
In order to confirm your Online Bank records, v
specific information from you.
Please follow the link below and renew your acc
https://login.personal.wamu.com/logon/logon.as

Thank you for your prompt attention to this mat
understand that this is a security measure meant
you and your account.
We apologize for any inconvenience.
If you choose to ignore our request, you leave u
to temporaly suspend your account.
Thank you for using Washington Mutual!

Fig. 6. SPOT Detection

SPOT algorithm classifies the tweet into spam tweet and non-spam tweet. The tweet is classified by using context-based filtering is called GOBAR filtering. Also, checks the character length, the tweet character is more than the limitation then it is considering as spam messages. The default spams links or messages are stored in the database.

The sender tweet was receiver by the SPOT algorithm it may checks the GOBAR filtering conditions the value is one (1) or less than three (3) means the message is attached with spam, whether the value is zero (0) means it is non-spam messages.


TWITTER SPAM DETECTION BASED ON SPOT ALGORITHM
USING COUNT THRESHOLD AND PERCENTAGE THRESHOLD
SPOT FILTER AND DETECTION
Apply Filter
Detection
Compare SPOT
Message
OK

CALCULATION IS DONE

Fig. 7. Spam Filter

CT denotes the Count Threshold. In the time window T is the threshold it specifies the time, the time set to 1 hour. The threshold value is represented by c and it denotes the fixed length of the spam tweet. It shows the threshold value is three, monitor is used to analyse the spam messages send with the hour how many times.

Percentage Threshold (PT), the minimum and maximum value of threshold is mentioned. PT is used to count the spam messages in the time and show the value in the percentage. Within the hour the spam tweet are more than 30 times, it considers the spam zombies are more than 50%.

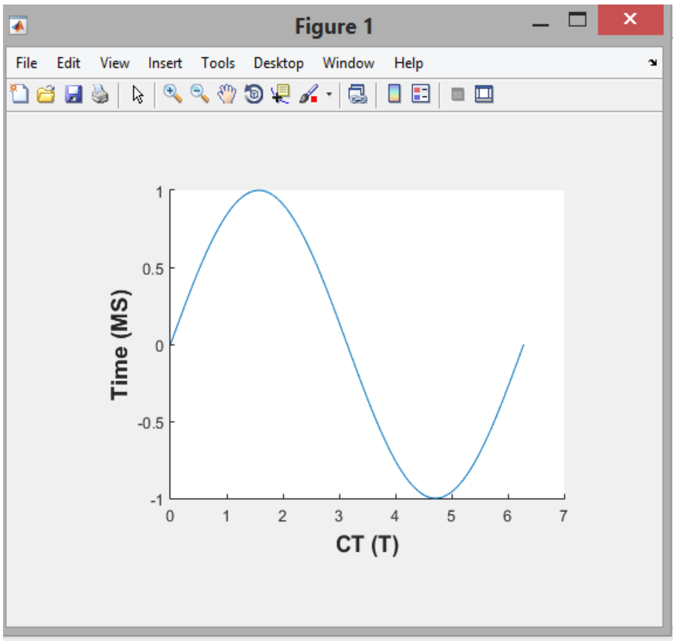


Fig. 8. CT and Time

Fig 8 shows the comparison of spam messages count CT with respect to time. The time taken for the proposed system is one hour. The curve reaches the peak point the sender machine send large volume of spam tweets.

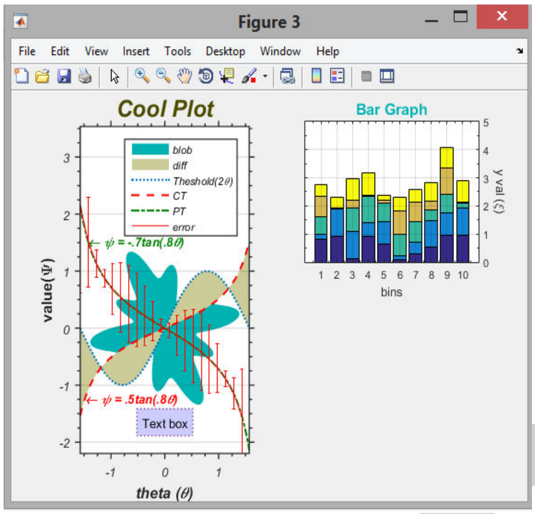


Fig. 9. SPOT Analysis

CONCLUSION

This paper analysed the spam detection in social communication especially in twitter. The system utilized GOBAR filtering (content filtering), SPOT algorithm CT and PT algorithm for the detection of spam. The GOBAR filtering technique classify the spam and non-spam tweets. SPOT algorithm is based on the Sequential Probability Ratio Test (SPRT) which keeps track of the sender messages. SPOT records the IP address of the sender machine, identify the spam messages send from the spammer and blocked it. Finding the CT and PT value depending upon the threshold limitation. The percentage of spam messages are computed by PT. The experimental result proved that the proposed method is efficient compared to the traditional spam detection. SPOT algorithm detects the spam messages in the network through online with exact accuracy and speed. The main focus of the work is on the detection of spam zombies and identifying the number of spam tweet counts and percentage of the spam tweets.

In online social medium, twitter is a very efficient medium of communication. The famous persons share their thought in the twitter page. Spammers activities are increased in twitter. The goodwill of famous persons is collapsed by the spammers. The proposed system is implemented using the SPOT detection algorithm and GOBAR filtering, performance and testing used the twitter dataset. In future, different algorithms can be used to classify the tweet spam messages and blocking the online spammers.

REFERENCES

[1] Ann Nosseir, Khaled Nagati and Islam Taj-Eddin (2013), “Intelligent Word-Based Spam Filter Detection Using Multi-Neural Networks”, IJCSI International Journal of Computer Science Vol. 10.

[2] Arjun Mukherjee1 Bing Liu Meichun Hsu2 Malu Castellanos2 Riddhiman Ghosh2 (2011), “Exploiting Burstiness in Reviews for Review Spammer Detection”, Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media.

[3] Asmeeta Mali (2013), “Spam Detection Using Baysian with Pattren Discovery”, International Journal of Recent Technology and Engineering (IJRTE), Vol 2.

[4] Aycock. J and Friess. N (2006), “Spam zombies from outer space”, Department of Computer Science, University of Calgary.

[5] Basavaraju. M (2010), “A Novel Method of Spam Mail Detection using Text Based Clustering Approach”, International Journal of Computer Application, vol 5.

[6] Benevenuto. F, Magno. G, T. Rodrigues, V. Almeida (2010), “Detecting spammer on twitter”, Seventh Annual Collaboration Electronic Messaging Anti-Abuse and Spam Conference.

[7] Chao Chen, Yu Wang, Jun Zhang, Yang Xiang (2017), “Statistical features-based real-time detection of drifted twitter spam”, IEEE Trans. On Information Forensics and Security, vol. 12.

[8] C. Chen, J. Zhang, X. Chen, Y. Xiang, W. Zhou (2015), “A large ground truth for timely twitter spam detection”, IEEE International Conference on Communications”.

[9] C. Chen, J. Zhang, X. Chen, Y. Xiang, W. Zhou (2015), “6 million spam tweets: A large ground truth for timely twitter spam detection”, IEEE International Conference on Communications.

[10] C. Chen, J. Zhang, Y. Xiang, W. Zhou (2015), “Asymmetric self-learning for tackling twitter spam drift”, IEEE conference on Computer Communications Workshops.

[11] Duan. K. Gopalan and X. Yuan (2007), “Behavioural Characteristics of Spammers and their Network Reachability Properties”, IEEE International Conference on Communications.

[12] Duan. Z and X. Yuan (2007), “Behavioural characteristics of spammers and their network reachability properties”, IEEE International conference on communication.

[13] Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu and Hady W. Lauw (2011), “Detecting Product Review Spammers using Ratib300-nbng Behaviours”, Human Language Technologies, Vol.1.

[14] Geerthik. S and Anish. T.P (2013), “Filtering Spam: Current Trends and Techniques”, International Journal of Mechatronics, Electrical and Computer Technology Vol.3.

[15] Guanjun Lin, Nan Sun1, Surya Nepal (2017), “Statistical Twitter Spam Detection Demystified: Performance, Stability and Scalability”, IEEE.

[16] Hao xue, Fengjun li, hyunjin seo and rosesnmpluretti (2015), “Trust-Aware Review Spam Detection”, IEEE Trustcom/bigdatase/ispa, Vol. 3.

[17] M. A. Hema (2014), “Effective Discovery of E-mail Spam Using SPOT Detection System”, Internal journal of Computer Science and Mobile Computing, Vol.3.

[18] Hua Shen, Xinyue Liu (2016), “Detecting Spammers on Twitter Based on Content and Social Interaction, International conference on Network and Information System for Computers IEEE.