TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 4467-4469

Spam Tweet Detection in Twitter

Suma Chandu[1], Tharun Kumar[2], Meghanath[3], Geetha B[4]

[4]Professor, [1,2,3,4]School of Computing and Information Technology,
REVA University, Bengaluru, India
[1]chandhu.chinna799@gmail.com, [2]gtharun012@gmail, [3]meghanadhreddy498@gmail.com,
[4]geetha.b@reva.edu.in

Article Info
Volume 83
Page Number: 4467-4469
Publication Issue:
May - June 2020

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *12 May 2020*

Abstract

As social networking sites are more popular the spammers uses these sites to keep spam tweets. Twitter is one of the platform for spammers, now the twitter bot is detecting some users and blocking them however it is difficult to block all the users .Nowadays so many users are using twitter, some of the users are spreading fake messages to the users. This type of messages will be seen mostly at the time of elections and any other occasions. The fake messages includes text like your account ********2314 is credited with 10000.00 and some other messages like you won a car, initially you have to deposit 20000 to avail this offer.
To overcome this we came with a approach for blocking the spam messages without reaching the user. This can be done by using the Machine Learning Algorithms. In this we are classified the machine learning algorithms and their performance is shown in a graph.

Keywords: *Twitter bot, tweet detection, machine learning algorithms*

1. Introduction

In this experiment we are identifying the fake messages and blocking them without reaching the user. For this we divided the message into three types spam, not spam and neutral messages. If the text in the message Is spam it will block the message without reaching the users. If text is neutral the message will deliver to the user but it will send a instruction to the developer to verify whether it is spam or not. Several machine learning algorithms such as SVM, NB, MLP, RF, KNN, DECISION TRESS are used to verify the messages as spam or not. The message will be divided into-1,0,1 based on its spam nature. It will show how the data is taken and how the data is preprocessing and the working models applied on the training set of examples. The spammers are targeting the twitter because the facebook and Instagram are blocking the third parties and because of this they are targeting in twitter.

To overcame this we came with this model which will help to block all the spam messages and helps the user to overcome with seeing the spam messages and it will block the messages without reaching the user and the spammer will be blocked from the particular mail ID using by user.

2. Related Work

We describe the related work in three areas: spam detection in long text data (e-mail, web, reviews), detection of spammers in Twitter, and spam detection at tweet level in Twitter. Drucker et al. studied the use of support vector machines (SVMs) to classify the e-mail as spam or non-spam. Androutsopoulos et al. showed that a Naive Bayesian classifier can be used to filter spam e-mails.

3. Literature Survey

Google safe browser and twitter bot are not identifying and blocking the spam messages because there are 321 millon twitter users are there and nearly 50 million tweets are made by the users so it is difficult to the google safe browser and twitter bot for identifying and blocking the spam messages.

So we are coming with a idea of machine learning algorithms which will detect the spam tweets and blocks them at user level the main aim is to classify whether a message is spam or not, previously some of them are designed them but they are having an efficiency of 89 % now we are implementing the algorithms and identifying the spam messages and also blocking them without reaching the user now we raised the efficiency upto 98%.

Generally spam messages can be seen at the time of elections and some of the users will use to pass some fake information like your account is credited with 10000 this type of messages are considered as spam messages.

We are also using NLTK which is a language processing unit to process the message and it will divide the paragraph into word and it will also read the total length of the word and space between the words and also some of the special symbols all will be read by using NLTK and the data is read by using the algorithms and it can be telling the message is spam or not by using NLTK.

4. Objectives

However the detection of word being positive or negative and analyse personality is a significant challenge. Classification can be done using various methods. To develop a system that detects sentence is spam or not spam based on the occurrence of the words.

Specific Objectives

There are some methods in this process. They are

1. Data Collection
2. Data Pre-Processing
3. Modelling and Training
4. Testing
5. Validation

1. Data Collection:-

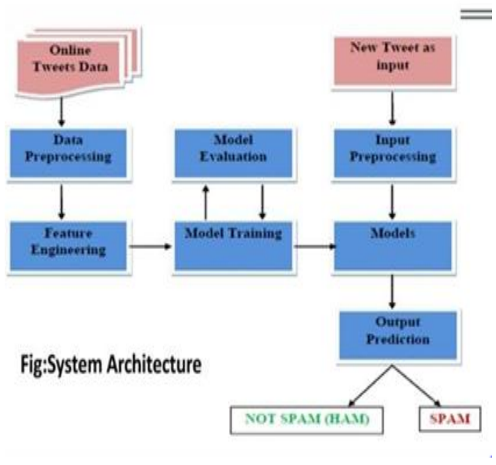

Figure 1: System Architecture

This is the first step in spam detection, in this step a large data set is collected which is in raw format. The data will be collected in XSJM format, it will use beautiful string to convert the message into reverse format.

2. Data Pre-Processing:-

Data pre-processing is a process of converting raw data into useful information. Because the unformated data is not useful for analysis. Generally a sentiment is used to divide the text into several parts. VADAR sentiment is used in this process. It will divide whether a message is positive or negative(-1,0,1).

3. Modelling and Testing

Several machine learning algorithms are used to train the program such as SVM, NB, MLP, BRF, KNN, DECISION TRESS. Each of the algorithms will be trained by giving the examples, and the algorithms will learn by the training. The training will be on the following terms

1. It will read the words and the gaps between the words.
2. It will also find out the length, capital letters, symbols and has tags of a tweet.
3. It will check the location of the user.
4. It will check the timeline and status of the user.
5. It will check the reputation count of the user, it calculated as:

Reputation count= #followers/#followers + #friends

4. Testing:-

This is the main phase in the project it will check each and every message and give whether a message is spam or not by doing the above three steps. In this project we have taken only one data set in that there are totally 11 million tweets. We have considered only 2 million tweets in this 1,40,000 are spam messages and the remaining messages are not spam. In this some messages are not found as they are deleted from by the twitter. The algorithms will tell a message as spam by marking as spam mark on the message and not spam by using not spam mark. For better understanding of the users we are giving red coloured mark to the spam messages and green coloured mark to the not spam messages.

As this algorithms are supervised learning algorithms we have to give a data set and train the algorithm to know which is spam and which is not spam. If we keep on training the algorithm we will get better results.

5. Validation:-

This is the step where it validate our algorithm. It will validate the algorithm by how accurately the algorithm is working and number of correct results the algorithm is giving. The clear process is seen with the block diagram. It will give a brief idea on how the whole process is done. The above figure shows how a tweet is classified and how it will predict the message as spam or not spam is shown in the above figure.

5. Methodology

Training Data

Training data is most important part of the whole system as training of the system is wholly depends on it and classification of testing data is done on the basis of this result only.

Algorithm

1. Online tweet data is collected.
2. Processing the raw data and converted into useful format(files).
3. Divide the word into training and testing dataset.

4. Sentiment analysis divides the sentence into spam or not spam based on the words.
5. If the word are not detected train the model for the word and gotostep1
6. Repeat the process.

6. Results

The algorithm is working better than the older algorithms and the results are shown by the help of neat graph. We are also comparing different algorithms to show which one is performing better and accurate.

The algorithms will be easy to understand to the users as it will give a clean information on what the algorithm is doing and how it is processing the data. This is done because the convenience of the user to understand about the queries that are related with the algorithms.

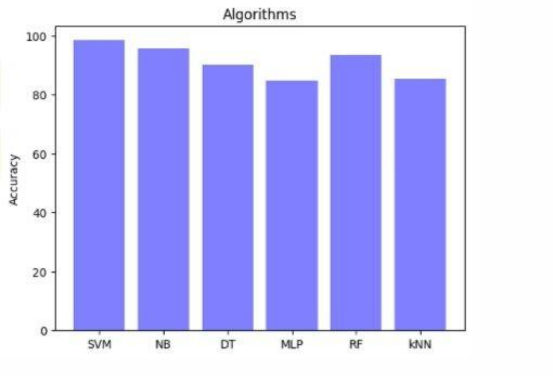

Figure 2: Efficiency of an algorithm

The above table shows how a efficiently a algorithm is working and we compared six algorithms in which SVM is giving faster and accurate results.SVM algorithm is showing more accurate results than any of the algorithm, it is showing up to 99 percent accuracy where as the NB is the second highest by 98 percent accuracy and the RF stands in third place with 96percent accuracy.

7. Application

We are doing this on detecting spam tweets in twitter it can be used in twitter it can also be used in different social media platforms like Facebook, Instagram and also in different platforms which the users are using the most .Currently Facebook and Instagram are blocking the third parties to be accessed so it difficult to keep spam messages in them if they give acesss it can be also used in Faccebook and Instgram.

8. Conclusion

The model which we developed is giving better results than the existing one and also it is showing exact results on which message is spam and which message is not spam so we have done the training on a large dataset and got better results.

We have taken a data set which is consisting of 10 million tweets and it showed 60000 spam message and the accuracy is nearly 98%.

References

[1] X. Zhang, Z. Li, S. Zhu, and W. Liang, “Detecting spam and promoting campaigns in Twitter,” ACM Trans. Web, vol. 10, no. 1, 2016, Art. no.4.

[2] S. Sedhai and A. Sun, “Semi-supervised spam detection in Twitter stream,” IEEE Trans. Comput. Social Syst., vol. 5, no. 1, pp. 169–175, Mar. 2017.

[3] B. Wang, A. Zubiaga, M. Liakata, and R. Procter, “Making the most of tweet-inherent features for social spam detection on Twitter,” in Proc. Making Sense Microposts, 2015, pp.10–16.

[4] C. Chen et al., “A performance evaluation of machine learning-based streaming spam tweets detection,” IEEE Trans. Comput. Social Syst., vol. 2, no. 3, pp. 65–76, Sep.2015.

[5] A. Töscher, m. Jahrer, and r. M. Bell,“the Bigchaos solution to the netflix grand prize,” Netflix prize document, pp. 1–52, sep. 2009