International Journal of Current Engineering and Technology
©2021 INPRESSCO®, All Rights Reserved

E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

Spam Detection using SVM on Twitter Data

Mr.Yash Shailesh Thigale and Prof.Deipali V. Gore

Department of Computer EngineeringPES Modern College of Engineering Shivajinagar,Pune.

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Social network sites involve billions of users around the world wide. User interactions with these social sites, like twitter have a tremendous and occasionally undesirable impact implications for daily life. The major social networking sites have become a target platform for spammers to disperse a large amount of irrelevant and harmful information. Twitter, it has become one of the most extravagant platforms of all time and, most popular microblogging services which is generally used to share unreasonable amount of spam. Fake users send unwanted tweets to users to promote services or websites that do not only affect legitimate users, but also interrupt resource consumption. Furthermore, the possibility of expanding invalid information to users through false identities has increased, resulting in malicious content. Recently, the detection of spammers and the identification of fake users and fake tweets on Twitter has become an important area of research in online social networks (OSN). In this Paper, proposed the techniques used to detect spammers on Twitter. In addition, a taxonomy of Twitter spam detection approaches is presented which classifies techniques based on their ability to detect false content, URL-based, spam on trending issues. Twelve to Nineteen different features, including six recently defined functions and two redefined functions, identified to learn two machine supervised learning classifiers, in a real time da ta set that distinguish users and spammers.

Keywords: *Machine Learning, Parallel Computing, Spam Detection, Scalability, Twitter*

Introduction

Online social networking sites like Twitter, Facebook, Instagram and some online social networking companies have become extremely popular in recent years. People spend a lot of time in OSN making friends with people they are familiar with or interested in. Twitter, founded in 2006, has become one of the most popular microblogging service sites. Around 200 million users create around 400 million new tweets a day for spam growth. Twitter spam, known as unsolicited tweets containing malicious links that the non-stop victims to external sites containing the spread of malware, spreading malicious links, etc., hit not only more legitimate users, but also the whole platform Consider the example because during the election of the Australian Prime Minister in 2013, a notice confirming that his Twitter account had been hacked. Many of his followers have received direct spam messages containing malicious links.The ability to order useful information is essential for the academic and industrial world to discover hidden ideas and predict trends on Twitter. However, spam generates a lot of noise on Twitter. To detect spam automatically, researchers applied machine learning algorithms to make spam detection a classification problem. Ordering a tweet broadcast instead of a Twitter user as spam or non-spam is more realistic in the real world.

Literature Survey

Literature survey is the most important step in any kind of research. Before start developing we need to study the previous papers of our domain which we are working and on the basis of study we can predict or generate the drawback and start working with the reference of previous papers. In this section, we briefly review the related work on Spam Detection and their different techniques.

Nathan Aston, Jacob Liddle and Wei Hu*[1] describe the "Twitter Sentiment in Data Streams with Perceptron" in this system the implementation feature reduction we were able to make our Perceptron and Voted Perceptron algorithms more viable in a stream environment. In this paper, develop methods by which twitter sentiment can be determined both quickly and accurately on such a large scale.

Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro [2] describe the "Aiding the detection of fake accounts in large scale social online services".in this paper, SybilRank, an effective and efficient fake account inference scheme, which allows OSNs to rank accounts according to their perceived likelihood of being fake. It works on the extracted knowledge from the network so it detects, verify and remove the fake accounts.

G. Stringhini, C. Kruegel, and G. Vigna [3] describe the "Detecting spammers on social networks" in this paper, Help to detect spam Profiles even when they do not contact a honey-profile.The irregular behavior of user profile is detected and based on that the profile is developed to identify the spammer.

J. Song, S. Lee, and J. Kim [4] describe the "Spamfiltering in Twitter using sender receiver relationship" in this paper a spam filtering method for social networks using relation information between users and System use distance and connectivity as the features which are hard to manipulate by spammers and effective to classify spammers.

K. Lee, J. Caverlee, and S. Webb [5] describe the "Uncoveringsocial spammers: social honeypots and machine learning" in this System analyzes how spammers who target social networking sites operate to collect the data about spamming activity, system created a large set of "honey-profiles" on three large social networking sites.

K. Thomas, C. Grier, D. Song, and V. Paxson [6] describe the "Suspended accounts in retrospect: An analysis of Twitter spam" in this paper the behaviors of spammers on Twitter by analyzing the tweets sent by suspended users in retrospect. An emerging spam-as-a-service market that includes reputable and not-so-reputable affiliate programs, ad-based shorteners, and Twitter account sellers.

K.Thomas, C.Grier, J.Ma, V.Paxson, and D.Song [7] describe the "Design and evaluation of a real-time URL spam filtering" in this paper, service Monarch is a real-time system for filtering scam, phishing, and malware URLs as they are submitted to web services.Monarch's architecture generalizes to many web services being targeted by URL spam, accurate classification hinges on having an intimate understanding of the Spam campaigns abusing a service.

X. Jin, C. X. Lin, J. Luo, and J. Han [8] describe the "Social spam guard: A data mining based spam detection system for social media networks" in this paper ,Automatically harvesting spam activities in social network by monitoring social sensors with popular user bases.Introducing both image and text content features and social network features to indicate spam activities. Integrating with our GAD clustering algorithm to handle large scale data. Introducing a scalable active learning approach to identify existing spams with limited human efforts, and Perform online active learning to detect spams in real-time.

S. Ghosh et al [9] describe the "Understanding and combating link farming in the Twitter social network" in this paper Search engines rank websites/webpages based on graph metrics such as PageRank High in-degree helps to get high PageRank. Link farming in Twitter Spammers follow other users and attempt to get them to follow back.

H. Costa, F. Benevenuto, and L. H. C. Merschmann [10] describe the "Detecting tip spam in location-based social networks" in this paper identifying tip spam on a popular Brazilian LBSN system, namely Apontador.Based on a labelled collection of tips provided by Apontador as well as crawled information about users and locations, we identified anumber of attributes able to distinguish spam from non-spam tips.

Proposed Methodology

In proposed system, the process of Twitter spam detection by using machine learning algorithms. Before classification, a classifier that contains the knowledge structure should be trained with the prelabeled tweets. After the classification model gains the knowledge structure of the training data, it can be used to predict a new incoming tweet. The whole process consists of two steps: learning and classifying. Features of tweets will be extracted and formatted as a vector. The class labels i.e. spam and non-spam could be get via some other approaches. Features and class label will be combined as one instance for training. One training tweet can then be represented by a pair containing one feature vector, which represents a tweet, and the expected result, and the training set is the vector. The training set is the input of machine learning algorithm, the classification model will be built after training process. In the classifying process, timely captured tweets will be labelled by the trained classification model.

A. Algorithms

1. Support Vector Machine:

Support Vector Machine (SVM) is used to classify the fruit quality. SVM Support vector machines are mainly two class classifiers, linear or non-linear class boundaries.

The idea behind SVM is to form a hyper plane in between the data sets to express which class it belongs to.

The task is to train the machine with known data and then SVM find the optimal hyper plane which gives maximum distance to the nearest training data points of any class Steps:

Step 1: Read the test image features and trained features. Step 2: Check the all test features of image and also get all train features.

Step 3: Consider the kernel.

Step 4: Train the SVM using both features and show the output.

Step 5: Classify an observation using a Trained SVM Classifier.

B. Mathematical Model

Let S be a system having Input(I), Functions(F) and Output(O). S=$\{I,F,O\}$ where, I is a Input

I=$\{Input\}$ O is the Output.

O=$\{Output\}$

F is the set of functions used for execution of commands Features of the tweet will be executed and formulated as vector.

- Input:Dataset Provided • Output:Tweet's Features.

- Functions:
  – $F = (f_1, f_2, ..., f_n)$ where,

F is the features of tweets. f_1 is the first feature vector . f_2 is the second feature vector.

- Here features and class labels are combined into Training set
- Processing
  – $TS = (f_1, label1), (f_2, label2), ..., (f_n, labeln)$ where,

f_1 is the feature related to class label 1. f_2 is the feature related to class label 2.
$label_n$ is the various labels through which tweets will get classified.

- Now we get the Training set which will be the input of Machine Learning Algorithms
- $TS = (f_1), (f_2), ..., (f_n)$ where,

TS is the Training set for classification model. The classification model will be built after the training process.
Here, timely captured tweets will be labeled by training classification model.

C. Architecture

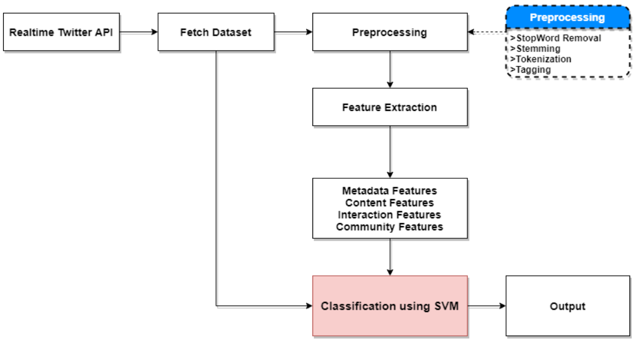

Fig. 1. System Architecture

D. Software Requirements and Specification

Result And Discussion

Experimental evaluation is done to compare the proposed system with the existing system for evaluating the performance. The simulation platform used is built using Java framework (version jdk 8) on Windows platform. The system does not require any specific hardware to run; any standard machine is capable of running the application.

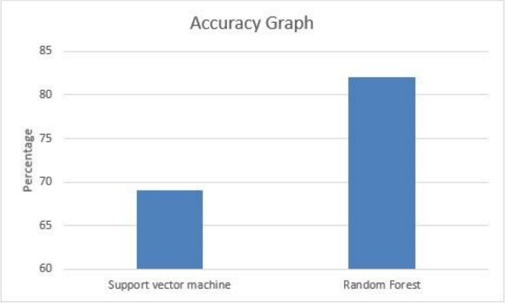

Fig. 2. Graph

Table 1:Comparative Result

Sr. No.	Existing System	ProposedSystem
1	69%	82%

Conclusion

In this Project, System found that classifiers ability to detect Twitter spam reduced when in a near real-world scenario since the imbalanced data brings bias. System also identified that Feature discretization was an important pre-process to MLbased spam detection. Second, increasing training data only cannot bring more benefits to detect Twitter spam after a certain number of training samples. System should try to bring more discriminative features or better model to further improve spam detection rate.

References

[1]. Nathan Aston, Jacob Liddle and Wei Hu*, "Twitter Sentiment in Data Streams with Perceptron," in Journal of Computer and Communications, 2014, Vol-2 No-11.

[2]. Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in Proc. Symp. Netw. Syst. Des. Implement. (NSDI), 2012, pp. 197–210.

[3]. G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in Proc. 26th Annu. Comput. Sec. Appl. Conf., 2010, pp. 1–9.

[4]. J. Song, S. Lee, and J. Kim, "Spam filtering in Twitter using sender receiver relationship," in Proc. 14th Int. Conf. Recent Adv. Intrusion Detection, 2011, pp. 301–317.

[5]. K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honeypots + machine learning," in Proc. 33rd Int. ACM SIGIR Conf. Res.Develop. Inf. Retrieval, 2010, pp. 435–442.

[6]. K. Thomas, C. Grier, D. Song, and V. Paxson, "Suspended accounts in retrospect: An analysis of Twitter spam," in Proc. ACM SIGCOMM Conf. Internet Meas., 2011, pp. 243–258.

[7]. K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time URL spam filtering service," in Proc. IEEE Symp. Sec. Privacy, 2011, pp. 447–462.

[8]. X. Jin, C. X. Lin, J. Luo, and J. Han, "Socialspamguard: A data mining based spam detection system for social media networks," PVLDB, vol. 4, no. 12, pp. 1458–1461, 2011.

[9]. S. Ghosh et al., "Understanding and combating link farming in the Twitter social network," in Proc. 21st Int. Conf. World Wide Web, 2012, pp. 61–70.

[10]. H. Costa, F. Benevenuto, and L. H. C. Merschmann, "Detecting tip spam in location-based social networks," in Proc. 28th Annu. ACM Symp. Appl.Comput., 2013, pp. 724–729.