ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

Twitter Spam Detection by Using Machine Learning Frameworks

Miss. Richa Ramesh Sharma, Prof. Yogesh S. Patil, Prof. Dinesh D. Patil

Master of Technology Student, Department of Computer Science & Engineering, Hindi Seva Mandal's, Shri Sant Gadge Baba College of Engineering & Technology, Bhusawal , Jalgaon, Maharashtra, India

Associate Professor, Department of Computer Science & Engineering, Hindi Seva Mandal's, Shri Sant Gadge Baba College of Engineering & Technology, Bhusawal , Jalgaon, Maharashtra, India

Associate Professor, Department of Computer Science & Engineering, Hindi Seva Mandal's, Shri Sant Gadge Baba College of Engineering & Technology, Bhusawal , Jalgaon, Maharashtra, India

ABSTRACT*:* Now a days users are increasing amount of time in social networks. However, due to the popularity of online social networks, cybercriminals are spamming on these platforms for potential victims. Spams invite users to external phishing sites or viruses downloads huge security issue online and undermined the user experience. However, current solutions do not reveal the Twitter spamming accurately and indeed. In this paper, we compared the performance of a wide range of conventional machine learning algorithms, with the aim of identify those that offer satisfactory detection and stability performance based on a large amount of true field data. With the objective in order to realize the real-time spam detection capability, we evaluated scalability algorithms. Performance the study evaluates the accuracy of the detection, the true positive/false positive and the F measure; stability analyses the stability of algorithms using randomly selected training samples of different sizes. Scalability aims to better understand the impact of in reducing training time learning algorithms.

KEYWORDS: Machine Learning, Parallel Computing, Spam Detection, Scalability, Twitter

I. 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.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

II. RELATED WORK

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 "Spam filtering 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 "Uncovering social 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.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

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 a number of attributes able to distinguish spam from non-spam tips

III. PROPOSED SYSTEM

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.

Advantages Of proposed system:

1. The system implements a method that will use the ML mechanism to detect if the post is spam or not.
2. Implementation of system can also be hosted online for use and data will be archived and retrieved from the server.
3. The user with the maximum amount of spam can be blocked by the system.

System Architecture:

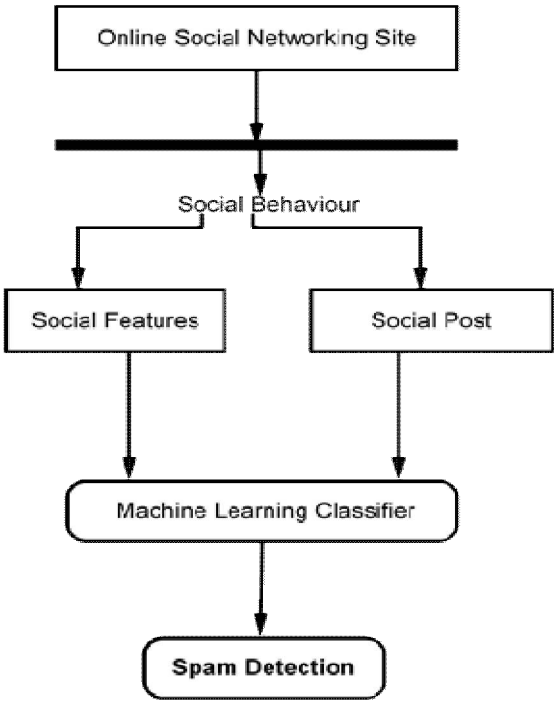


Fig 01. System Architecture

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

Methodology:

1) Feature Extraction: Extraction of 10-12 features and categories as Tag based features and URL based features. User-based features were extracted from the JSON object "user," such as account_age, which can be calculated by using the collection date minus the account created data. Other user-based features, like no_of followers, no_of followings, no_userfavourites, no_lists, and no_tweets, can be directly parsed from the JSON structure. Tweet-based features includes no_retweets, no_hashtags, no_usermentions, no_urls, no_chars, and no_digits. While no_chars and no_digits needs a little computing, i.e., counting them from the tweet text, others can also be straightforwardly extracted.
2) Feature Statistics: System evaluate the spam detection performance on dataset by using machine learning algorithms.
3) ML- Based SPAM Tweets Detection: This consist of,
i) Naïve Bays: This is mainly used for filtering the spam tweets and also used in text classification. Naive Bays classifiers work by correlating the use of tokens (typically words, or sometimes other things), with spam and non-spam tweets and then using Bayes' theorem to calculate a probability that a tweets is or is not spam.

Algorithm steps:
A. **Naive Bayes**

Step 1: Convert the data set into a frequency.

Step 2: Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.
For example:

Problem: Players will play if weather is sunny. Is this statement is correct?
We can solve it using above discussed method of posterior probability.

P(Yes | Sunny) = P(Sunny | Yes) * P(Yes) / P (Sunny)

Here we have P (Sunny |Yes) = 3/9 = 0.33, P(Sunny) = 5/14 = 0.36, P(Yes)= 9/14 = 0.64

Now, P (Yes | Sunny) = 0.33 * 0.64 / 0.36 = 0.60, which has higher probability.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems

4) Performance Evaluation: The evaluation can be done based on following factors:

i) Performance matrices such as TPR FPR Precision Recall etc.

ii) Impact of spam to Non-spam ratio

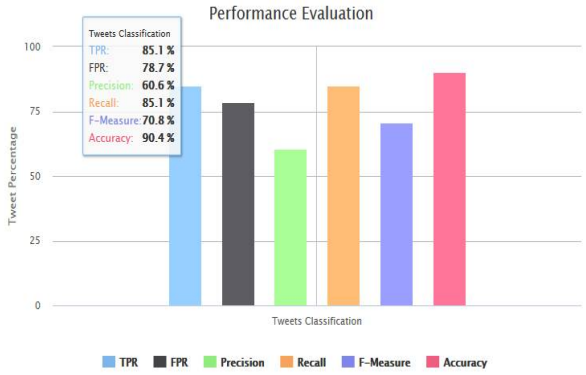
iii) Impact of Different Sampling method

iv) Investigation of time related data

IV. RESULT AND DISCUSSION

A. Offline Dataset Results

Naïve Bayes Performance:



Parameters	Percentage
TPR	85.1
FPR	78.7
Precision	60.6
Recall	85.1
F-Measure	78.8
Accuracy	94.4

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

B. Online Twitter Dataset Results:

Naïve Bayes Performance:

Parameters	Percentage
TPR	84.6
FPR	77.8
Precision	61.1
Recall	84.6
F-Measure	71.0
Accuracy	90.0

V. 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 ML-based 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.