© 2019 IJSRSET | Volume 6 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology
DOI : https://doi.org/10.32628/IJSRSET1962170

Statistical Analysis for Twitter Spam Detection

Ganesh Udge*, Mahesh Mohite, Shubhankar Bendre, Yogeshwar Birnagal, Mrs. Disha Wankhede
Department of Computer Engineering, VIIT, Pune, Maharashtra, India

ABSTRACT

The spreading and learning of new discoveries and information is made available using current online social networks. In Recent days, the solutions may be irrelevant to the actual content; also termed as attacks in the layman's term such attacks are been performed on Twitter as well and called as Twitter spammers. The quality of data is being compromised by addition of malicious and harmful information using URL, bio, emoticons, audio, images/videos & hash-tags through different accounts by exchanging tweets, personal messages (Direct Message's) & re-tweets. Misleading sites may be linked with the malicious links which may affect adverse effects on the user and also interfere in their decision making processes. To improve user-experience from the spammers attacks, the training twitter dataset are applied and then by extracting and using the 12 lightweight features like user's age, number of followers, count of tweets and re-tweets, etc. are used to distinguish the spam from non-spam. For enhancing the performance, the discretization of the function is important for transmission of spam detection between tweets. Our system creates classification model for Spam detection which includes binary classification and automatic learning algorithms viz. Naïve Bayes classifier or Support Vector Machine classifier which understands the behaviour of the model. The system will categorize the tweets from datasets into Spam and Non-spam classes and provide the user's feed with only the relevant information. The system will report the impact of data-related factors such as relationship between spam and non-spam tweets, size of training dataset, data sampling and detection performance. The proposed system's function is detection and analysis of the simple and variable twitter spam over time. The spam detection is a major challenge for the system and shortens the gap between performance appraisals and focuses primarily on data, features and patterns to identify real user and informing it about the spam tweets along with the performance statistics. The work is to detect spammed tweets in real time, since the new tweets may show patterns and this will help for training and updating dataset and in knowledge base.

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

I. INTRODUCTION

OSN's (Online social networking sites) like twitter; Facebook, Instagram, etc. are being utilized in daily routine. People connect with others using networking sites. Likely, twitter with around 335 million active users creates approximately 450 million new tweets every day contributing to spam growth. twitter spam can also be termed unsolicited tweets, containing malicious information and links that forwards users to external sites having irrelevant information. this affects certain amount of real users and also the whole platform. Consider an example, during the election of the American president in 2016, there is a very high possibility that the some hackers interfered with the elections through spreading and misguiding real users on the social media sites and changes the course of the elections. it is said that multiple fake accounts were created with them forwarding misleading information on both the candidates to sway the result in someone's favour. This information referred above is about spam tweets that should been removed or

blocked from the user's feed as the facts in them were misleading. Prediction of trends and patterns should be observed and avoided from spams to be performed. Spam generates noise over the platform and to detect spam automatically (periodically), researchers have applied classifying algorithms to make spam detection as classification problem. Categorizing the tweets distinguishing into Positive (Non-spam) and Negative (Spam) is betterment of the real world and will increase performance of the platform linearly.

II. RELATED WORK

Literature survey is most important and essential for research. Prior performed systems (papers) have to be studied of our domain to predict or generate the drawbacks and start working with the reference on these papers.
In this section we briefly review the implementation of Spam detection and its different methods,

A paper intending - Sybil Rank, an effective fake account detection scheme, that allows Social networks to rank accounts according to their weightage of being fake. It works on the extracted knowledge from the platform network so it detects, verifies and removes the fake accounts resulting in originality [1].

This paper enables amd unrolls detection of spam accounts, even when they don't contact honey-profile. The suspicious or irregular behaviour of the profile is detected and then identified as spammer on the account's previous knowledge [2].

Here a spam filtering method is developed for the Twitter using relational information between users and system also uses distance and connectivity as features which are hard for the spammers to manipulate and help in their classification [3].

This survey paper acknowledges the analysis of the system and how the spammers that target these OSNs to collect and manipulate the data of the spamming activity, system created a huge dataset of "honey-profiles" of the three major OSNs [4].

This system features the implementation in our own perception and algorithms which camoufloges into the stream environment. In this paper, the developed methods are for twitter sentiment that determines are quick and accurate on a very huge scale [5].

This paper determines the behaviours of spam spreading sources are by analysing the tweets sent by suspended users in retrospect[6].

The service names Monarch which is a real-time filtering system for permeating malware URLs, phishing and scam as they are included in web services. Monarch's architecture would be widely used by many web services that are being targeted by URL spam, content spam. Accurate classification of said filtering service hinges on having an intimate and thorough working of spam campaigns [7].

This paper determines automatically emerging spam activities by monitoring social accounts with popular user bases. Using information from image content, text content and network features to desirable spam activities. Integrating with clustering algorithms for Big Data with a scalable learning approach to identify existing spams with limited or none human efforts and perform online active, self-learning in real-time [8].

Here Search Engines rank sites/pages which uses graph metrics similar to PageRank High in-degree can guide in achieving a high PageRank. SEO (Search Engine Optimization) can also be helpful in some cases. Link Farming, allows spammers to follow and try to get them to follow back which results in more wide reached network of users to spread misleading information in the form of spam tweets [9].

In this paper identifying tweet spam on Apontador (Brazilian LBSN system). Based on labelled collection of tweets provided by Apontador. Gathering various information about users, user locations, we can identify number of attributes that distinguish between Positive (Non-spam) and Negative (Spam) [10].

III. EXISTING SYSTEM

Improvements over the years have been performed in this filed because of extensive usage and applications. This section determines the advances towards the requirements and achieves the purpose mentioned above (related work). These works are mainly differentiated by the algorithm used for classification model used in spam detection.

IV. PROPOSED SYSTEM

In the proposed system, we use the procedure for detection of spam tweets using Classifying algorithms. Prior to classification, a classifier that contains knowledge structure should be trained with pre-labelled tweets from a certain selected dataset for the purpose. Using the result of trained data, new tweets can be classified also during real time feed. Algorithm consists of two parts: 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 gained by some other techniques. Features and class labels will be combined as one instance for training. One training tweet can then be represented by a pair containing one feature vector that represents a tweet, and the expected result. The training is a vector. It is the input to the machine learning algorithm; classification model will be built after the training process. In the classifying process, timely captured tweets will be labelled by the trained classification model. This model will then recognise the nature of the incoming tweet and categorize them as spam or non-spam. This can also be implanted in real-time. The whole system can be made into a web application and run on the browser as an extension. With the system already trained and knowing what topics is the user interested in.

The proposed system will also be able to block or restrict contact to the particular user by checking the amount of fake/spam tweets broadcasted by that user.

Advantages:

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:

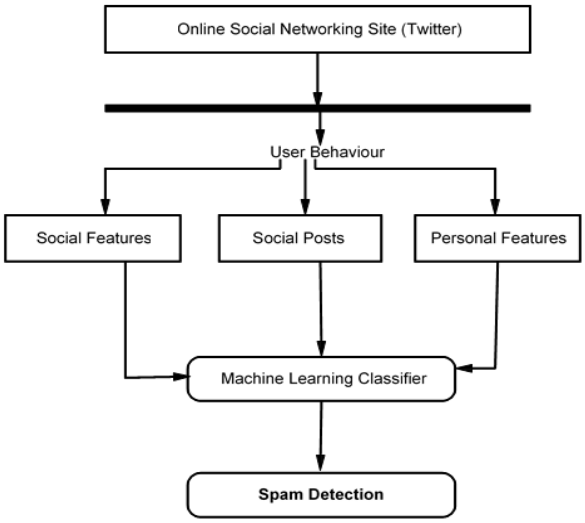


Fig 1. System Architecture

Methodology:

1) Feature Extraction: Selection of 10-12 features and classified as Tag based features and URL based features. User-based features selected from “user” - a JSON object, such as account age, which can be calculated by using the collection date minus the account created data. Other user-based features, like number of followers, number of followings, number of user favourites, number of tweets, and number of

lists which is directly parsed from the JSON structure. Number of retweets, number of hash tags, number of user mentions, number of URLs, number of chars, and number of digits are the part of Tweet based features . While number of chars and number of digits needs a little computing.

2) Feature analysis: System evaluate the spam detection performance on dataset by using classification (ML) algorithms.

3) Classification - Based Tweets Detection: This consist of,

i) Naïve Bays: Naive Bayes makes an assumption that the effect is independent of other attibutes' values on various attributes values inspite of a given class.

ii)SVM: Support Vector Machine is used to classify linear and non-linear data

A. Naive Bayes

Naïve Bayes is a classification algorithm based on Bayes Theorem and the Maximum A Posteriori hypothesis.

The algorithm performs better for categorical input variables.

It is useful for large data sets, is also called as Simple Bayes.

B. Support Vector Machine

Step 1: SVM uses a nonlinear mapping method to process and enhance original training data into higher dimension, it searches linear optimal separating hyper plane.

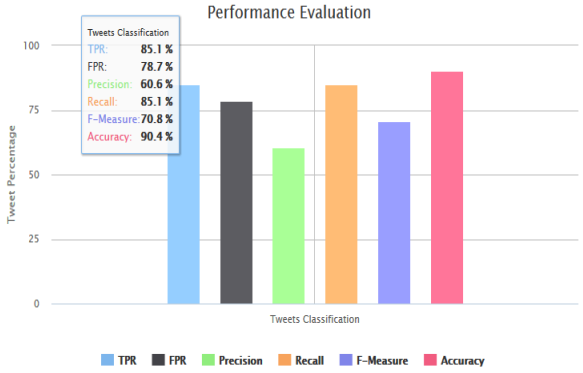
Step 2: SVM finds hyper plane using training tuples and margins.

Step 3: Output of linear function is greater than 1 , it is identified as one class and if the output is -1 it is identifies as another class.

V. RESULT AND DISCUSSION

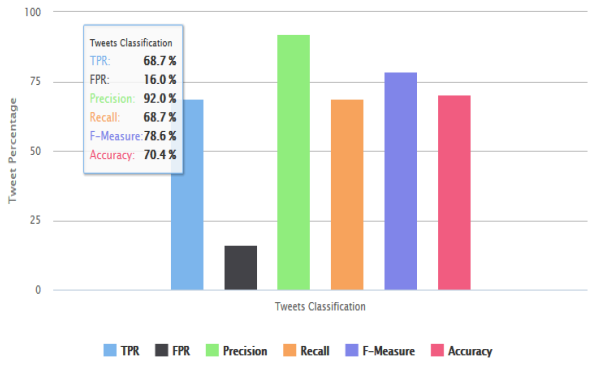
A. Offline Dataset Results

Naïve Bayes Performance:



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

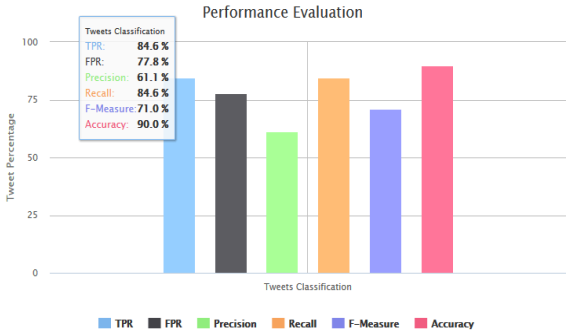
SVM Performance:



Parameters	Percentage
TPR	68.7
FPR	58.7
Precision	92.6
Recall	68.1
F-Measure	78.8
Accuracy	70.4

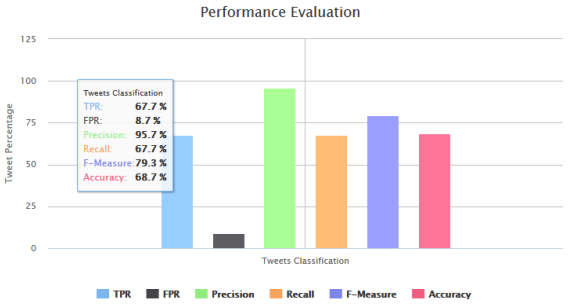
B. Online Twitter Dataset Results:

Naïve Bayes Performance:



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

SVM Performance:



Parameters	Percentage
TPR	67.7
FPR	56.3
Precision	95.7
Recall	67.7
F-Measure	79.3
Accuracy	68.7

VI. CONCLUSION

This project states the classification and evaluation of categorized tweets – spam and non-spam which then detected and reduced when compared to real-world occurrence of spammed and non-spammed tweets due to imbalanced data biasing. Also important distinct features were identified as pre-process for spam detection. Second, increasing training data only cannot bring more benefits to detect Twitter spam on specific amount of training samples. System should try to bring more discriminative features or better model to further improve spam detection rate.

VII. REFERENCES

[1]. 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.

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

[3]. 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.

[4]. 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.

[5]. 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.

[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.

Cite this article as :

Ganesh Udge, Mahesh Mohite, Shubhankar Bendre, Yogeshwar Birnagal, Mrs. Disha Wankhede, "Statistical Analysis for Twitter Spam Detection", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 2, pp. 624-629, March-April 2019. Available at doi : https://doi.org/10.32628/IJSRSET1962170
Journal URL : http://ijsrset.com/IJSRSET1962170