International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 3, pp: 501-504 www.ijaem.net ISSN: 2395-5252

Spam Detection in Twitter Stream

Liya Prakash, Mrs. Annes Philip
M.tech Computer Science and Engineering MES College of Engineering, Kuttipuram Kerala, India
Assistant Professor Dept. of CSE
MES College of Engineering, Kuttipuram
Corresponding author: Liya Prakash

Date of Submission: 26-07-2020 Date of Acceptance: 05-08-2020

ABSTRACT—online social networks (OSN) and microblogging websites are attracting in ternet users more than any other kind of website. Services such those offered by Twitter, Facebook and Instagram are more and more popular among people from different backgrounds, cultures and interests.Proposing a semi- supervised spam detection framework for spam detection at tweet-level.And also hate speech is an important problem.Hate speech refers to the use of aggressive, violent or offensive language, targeting a specific group of people sharing a common property, whether this property is their gender (i.e., sexism), their ethnic group or race (i.e., racism) or their believes and religion, etc.

Index Terms—SVM, Preprocessing, Testing, Training

I. INTRODUCTION

Twitter is a social networking site where people interact with each other through tweets. Spammer tweets pose either as advertisements,scams and help to perpetrate phishing attacks or the spread of malware through the embedded URLs. Twitter is an attractive platform for spammers so the spamming activities are increasing Only the registered users can post the tweets but unregistered users can read it.An unwanted content appearing in twitter can be said as spam. Tweets contain URL and links which after clicking directs users to some website which contain viruses, malware,scams etc .It is necessary to save users and system from such spammers. Here proposing a spam detection framework in machine learning. Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. As it is evident from the name, it gives the computer that which makes it more similar to humans: The ability to learn. Twitter spam is one of the most important problems that professionals have to deal with in social networks on the internet. For this problem, the researchers presented some solutions, mostly based on a number of different methods considering learning. The means and techniques used at the current time has achieved a good ratio of the accuracy based on the so-called methods of blacklisting in order to determine the undesirable activities in relation to send and receive an e-mail on social networks that based on the conclusions obtained from previous experiments and studies.A semi-supervised frame-
work for spam tweet detection is proposed.

The framework mainly consists of two main modules: 1) four lightweight detectors in the spam tweet detection module for detecting spam tweets in real time and 2) updating module to periodically update the detection models based on the confidently labeled tweets from the previous time window.

While most of the online social networks and microblogging websites forbid the use of hate speech, the size of these networks and websites makes it almost impossible to control all of their content. Therefore, arises the necessity to detect such speech automatically and filter any content that presents hateful language or language inciting to hatred. An approach to detect hate expressions on Twitter is proposed. This approach is based on unigrams and patterns that are automatically collected from the training set. These patterns and unigrams are later used, among others, as features to train a machine learning algorithm.

Machine learning is the filed of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. And it is evident from the name,it gives the computer that which makes it more similar to humans, that is the ability to learn.

II. LITERATURE SURVEY

A. State of the art

1) Detecting spammer on Twitter: Approaching[1] the problem of detecting spammers on Twitter. First col lected a large

dataset then the dataset was manually classified into spammers and non-spammers. The irregular behavior of user profile is detected and based on that the profile is developed to identify the spammer.

Consider the problem of detecting spammers on Twitter.A large dataset of Twitter that includes more than 54 million users, 1.9 billion links, and almost 1.8 billion tweets is collected. Using tweets related to three famous trending topics from 2009, a large labeled collection of user is constructed, manually classified into spammers and non-spammers.Then a number of characteristics related to tweet content and user social behavior is identified, which could potentially be used to detect spammers. Using these characteristics as attributes of machine learning process for classifying users as either spammers or nonspammers. The strategy is to succeeds at detecting much of the spammers while only a small percentage of non-spammers are misclassified

2) Spam filtering in Twitter using sender receiver relation- ship : S.Lee peoposing[2] a novel spam filtering system that detects spam messages in Twitter. Instead of using account features, using relation features, such as the distance and connectivity between a message sender and a message receiver, to decide whether the current message is spam or not. sender receiver relationship is used to detect the spam message. J48 classifier is used in this approach. Here, the relation feature approach is very difficult to calculate.

3) Statistical Features-Based Real-Time Detection of Drifted Twitter Spam : The Statistical Features-Based Real- Time Detection of Drifted Twitter Spam d[3]is the scheme first carry out a deep analysis on the statistical features of one million spam tweets and one million non-spam tweets, and then propose a novel Lfun scheme. The proposed scheme can discover spam tweets from unlabeled tweets and incorporate them into classifier's training process. This proposed Lfun can effiectively detect Twitter spam by reducing the impact of Spam Drift" issue.

simply compare some representative statistics, such as the mean values of fea tures to show the "Spam Drift" problem. To further illustrate the changing of the statistical features in a dataset, a natural approach is to model the distribution of the data . There are two kinds of approaches: parametric and non- parametric. Parametric approaches are very powerful when the specific distribution of the dataset, like Normal Distribution, is already known. However, the distribution of the Twitter spam data is unknown, thus it is not possible to apply parametric ap proaches. Consequently, non-parametric methods, such as statistical tests, which make no assumptions of the dataset distributions are used by researchers.

4) Real Time Detection of Drifted Twitter Spam Based on Sta tistical Features: The real time detection of drifted twitter spam based on statistical features is a system[4] with elaborated Lfun scheme which helps to deal with twitter spam by reducing impact of Spam Drift issue. For this considered the existing dataset to study twitter spam. The proposed system focus on URL Thread Detection which helps to detect whether the particular URL is malicious or not. For this considered the most recent tweet containing URL. This URL is send to virus total website which analyze suspicious rules and URLs to detect types of malware including viruses, worms, and Trojans. Finally produce the result showing whether the URL is suspicious or not.

System which elaborated Lfun scheme which helps to deal with twitter spam by reducing impact of Spam Drift issue. For this considered the existing dataset to study twitter spam. In this proposed system have to focus on URL Thread Detection which helps to detect whether the particular URL is malicious or not. Considering the most resent tweet containing URL. This URL is send to virus total website which analyze suspicious files and URLs to detect types of malware includ- ing viruses, worms, and Trojans. Finally produce the result showing whether the URL is suspicious or not.

III. PROPOSED METHOD

The proposed system contains two main modules.Assuming that have all the information(e.g., a blacklist of spamming domains and trained classification models), the tweets are labeled as spam and nonspam (also known as "ham") tweets using the four detectors in real time. The required information is updated periodically based on the confidently labeled tweets from the previous time window, in a semi-supervised manner. Next, detailing the main modules.

A. System Architecture

The system architecture mainly consist of two phases. They are a training phase and an testing phase

1) Trianing Phase : The training in supervised machine learning is also known. It is the task of inductive learning or classification. It is the task of inferring a function (classifier) from a supervised (labeled) training phishing websites

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 3, pp: 501-504 www.ijaem.net ISSN: 2395-5252

dataset. A supervised learning algorithm analyzes the training phishing websites dataset and produces a classifier, which can predict the correct class for unseen dataset and effectively detect the newly created phishing websites. Once the significant features are selected properly using the wrapper approach, the machine learning techniques can be trained in order to correctly classify the website, as either a phishing or legitimate website.

2) Testing Phase: In the training phase, a learning algo- rithm uses the training data to gen erate a classification model (classifier). In testing phase, the learned classifier is evalu- ated using the testing dataset to get the correct classification accuracy. If the correct classification accuracy for the testing dataset is acceptable, the trained 11 classifier can be used in real-world applications.

The proposed approach automatically detects spam patterns and most common unigrams and use these along with senti- mental and semantic features . The proposed frame work is mainly deals with

- Import modules Will be using pandas, numpy and Multi- nomial naive Bayes classifier for building a spam detec- tor. Pandas will be used for performing operations on data frames. Furthermore using numpy, we will perform necessary mathematical operations.
- Reading the dataset and preparing it for basic processing First, read the csv using pandas read function. Then modify the column names for easy references. In this dataset, the target variable is categor ical (ham, spam) and need to convert into a binary variable. Remember, machine learning models always take numbers as input and not the text hence we need to convert all categorical variables into numerical ones. Replace ham with 0 and spam with 1 .
- Cleaning text is one of the interesting and very important steps before per forming any kind of analysis over it. Text from social media and another platform may contain many irregularities in it. People tend to express their feeling while writing and you may end up with words like gooood or goood or gooooooooooood in your dataset. Essentially all are same but we need to regularize this data first.Have made a function below which works fairly well in removing all the inconsistencies from the data.
- following steps are done tro clean data.....

i Removing web links from the text data as they are not pretty much useful
ii Correcting words like poooor and baaaaad to poor and bad
iii Removing punctuations from the text
iv Removing apostrophes from the text to correct words like I'm to I am
v Correcting spelling mistakes

B. Feature extraction

An n-gram is a contiguous sequence of n items from a given sequence of text. Given a sentence, have to construct a list of n-grams from s finding pairs of words that occur next to each other. For example, given the sentence "I am Malu" you can construct bigrams (n-grams of length 2) by finding consecutive pairs of words which will be ("I", "am"), ("am", "Malu").

A consecutive pair of three words is known as tri-grams. This will help us to understand how exactly a sequence of tokens together determines whether an incoming message is a spam or not. In natural language processing (NLP), n-grams hold a lot of importance as they determine how sequences of words affect the meaning of a sentence.

C. Classification Algorithm

SVM (Super Vector Machine)is the algorithm used in the proposed method .It will provide highest accuracy more than other classifiers.

IV. EXPERIMENTS AND RESULTS

System provides a fundamental evaluation of ML algorithms on the detection of streaming spam tweets. The system will identify spam in twitter. Through this system, the aim is to implement the detection of spam using machine learning approach. The task will be done by extracting the features of spam and updating the models.Propose a semi-supervised spam detection framework. Utilizes four lightweight detectors to detect spam tweets on real-time basis and update the models periodically in batch mode. The experiment results will demonstrate the effectiveness of semi-supervised approach in spam detection framework.

Fig. 1. Reading the dataset and preparing it for basic processing

	precision	recall	f1-score	support
0	0.95	0.99	0.97	566
1	0.70	0.21	0.32	34
accuracy			0.95	600
macro avg	0.83	0.60	0.65	600
weighted avg	0.94	0.95	0.94	600

Accuracy: 0.95

Fig. 2. Result after Testing Dataset

V. CONCLUSION

Through this system, the aim is to implement the detection of spam using machine learning approach. The task will be done by extracting the features of spam and updating the models.

REFERENCES

[1] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida,"Detecting spammer on twitter", in Proc. 7th Annu. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf., Jul. 2012, p. 12

[2] 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. 301317.

[3] Chao Chen, Yu Wang, Jun Zhang, Yang Xiang, WanleiZhou, "Statistical Features-Based Real-Time Detection of Drifted TwitterSpam, Twit-terSpam,"IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 12, NO. 4, APRIL 2017.

[4] Sayali Kamble,S.M.Sangve, ""Real Time Detection of Drifted Twit- terSpam Based on StatisticalFeatures",Features"2018 nternational Conference on Information, Communication, Engineering and Technol- ogy (ICICET) .

[5] Surendra Sedhai, Aixin Sun, "Semi Supervised Spam Detection Twitter Stream", "IEEE Transactions On Computational Social Systems 2017.

[6] Hajime Watabe, Mondher Boduazizi,And Tomoaki Ohtsuki, ""Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection", "IEEE Transactions On Computational Social Systems 2018.

International Journal of Advances in Engineering and Management
ISSN: 2395-5252

IJAEM

Volume: 02 Issue: 01 DOI: 10.35629/5252

www.ijaem.net
Email id: ijaem.paper@gmail.com