JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Twitter Bot Detection Using Reduced Feature Set

T. Sowmya Sri, B.Tech student, Department of IT, CMR Technical Campus, Hyderabad

K.L.S. Sri Vidya, B.Tech student, Department of IT, CMR Technical Campus, Hyderabad

B. Greeshma, B.Tech student, Department of IT, CMR Technical Campus, Hyderabad

V. Sri Suma, Assistant Professor, Department of IT, CMR Technical Campus, Hyderabad

Abstract: Malicious social bots generate fake tweets and automate their social relationships either by pretending like a follower or by creating multiple fake accounts with malicious activities. Moreover, malicious social bots post shortened malicious URLs in the tweet in order to redirect the requests of online social networking participants to some malicious servers. Hence, distinguishing malicious social bots from legitimate users is one of the most important tasks in the Twitter network. To detect malicious social bots, extracting URL-based features (such as URL redirection, frequency of shared URLs, and spam content in URL) consumes less amount of time in comparison with social graph-based features (which rely on the social interactions of users). Furthermore, malicious social bots cannot easily manipulate URL redirection chains. In this article, a learning automata-based malicious social bot detection (LA-MSBD) algorithm is proposed by integrating a trust computation model with URL-based features for identifying trustworthy participants (users) in the Twitter network. The proposed trust computation model contains two parameters, namely, direct trust and indirect trust. Moreover, the direct trust is derived from Bayes' theorem, and analyze the malicious behavior of a participant by considering URL-based features, such as URL redirection, the relative position of URL, frequency of shared URLs, and spam content. Experimentation has been performed on Twitter data sets, and the results illustrate that the proposed algorithm achieves improvement in precision, recall, Fmeasure, and accuracy.

Keywords: Bot Detection, Machine Learning, URL detection

1. Introduction

PROJECT SCOPE:

Malicious Social bot is a software program that pretends to be a real user in online social networks (OSNs). Moreover, malicious social bots perform several malicious attacks, such as spread social spam content, generate fake identities, manipulate online ratings, and perform phishing attacks. In Twitter, when a participant (user) wants to share a tweet containing URL(s) with the neighboring participants (i.e., followers or followers), the participant adapts URL shortened service (i.e., bit.ly) in order to reduce the length of URL (because a tweet is restricted up to 140 characters). Moreover, a malicious social bot may post shortened phishing URLs in the tweet. when a participant clicks on a shortened phishing URL, the participant's request will be redirected to intermediate URLs associated with malicious servers that, in turn, redirect the user to malicious web pages. Then, the legitimate participant is exposed to an attacker. This leads to Twitter network suffering from several vulnerabilities (such as phishing attack).

PROJECT PURPOSE:

The malicious behavior of participants is analyzed by considering features extracted from the posted URLs (in the tweets), such as URL redirection, frequency of shared URLs, and spam content in URL, to distinguish between legitimate and malicious tweets. To protect against the malicious social bot attack. Several approaches have been proposed to detect spam in the Twitter network. These approaches are based on tweet-content features, social relationship features, and user profile features. However, the malicious social bots can manipulate profile features, such as hashtag ratio, follower ratio, URL ratio, and the number of retweets. The malicious social bots can also manipulate tweet-content features, such as sentimental words, emoticons, and most frequent words used in the tweets, by manipulating the content of each tweet. The social relationship-based features are highly robust because the malicious social bots cannot easily manipulate the social interactions of users in CMRTC the Twitter network. However, extracting social relationship-based features consumes a huge amount of time due to the massive volume of social network graph.

PROJECT FEATURES:

Identifying the malicious social bots from the legitimate participants is a challenging task in the Twitter network. The existing malicious URL detection approaches are based on DNS information and lexical properties of URLs. The malicious social bots use URL redirections in order to avoid detection. However, for detectors, identification of all malicious social bots is an issue because malicious social bots do not post malicious URLs directly in the tweets. Thus, it is important to identify malicious URLs (i.e., harmful URLs) posted by malicious social bots in Twitter. Most of the existing approaches are based on supervised learning algorithms, where the model is trained with the labelled data in order to detect malicious bots in OSNs. However, these approaches rely on statistical features instead of analysing the social behaviour of users. Moreover, these approaches are not highly robust in detecting the temporal data patterns with noisy data (i.e., where the data is biased with untrustworthy or fake information) because the behaviour of malicious bots' changes over time in order to avoid detection. This motivated us to consider one of the reinforcements learning techniques (such as the learning automata (LA) model) to handle temporal data patterns. In this work, we design an LA model to detect malicious social bots with improved precision and recall.

PROBLEM DEFINITION:

Malicious social bot is a software program that pretends to be a real user in online social networks (OSNs). Moreover, malicious social bots perform several malicious attacks, such as spread social spam content, generate fake identities, manipulate online ratings, and perform phishing attacks. In Twitter, when a participant (user) wants to share a tweet containing URL(s) with the neighboring participants (i.e., followers or followers), the participant adapts URL shortened service in order to reduce the length of URL (because a tweet is restricted up to 140 characters). Moreover, a malicious social bot may post shortened phishing URLs in the tweet

2. OVERVIEW OF THE SYSTEM

3.1 Existing System

Existing methods analyzed social botnet attack on Twitter. The authors have presented that social bot use URL shortening services and URL redirection in order to redirect users to malicious web pages. Existing methods presented methods to detect, retrieve, and analyze botnet over thousands of users to observe the social behavior of bots. In, a social bot hunter model has been presented based on the user behavioral features, such as follower ratio, the number of URLs, and reputation score.

.

3.1.1 Disadvantages of Existing System

The existing methods analyzed that the low trust value of a user indicates that the information spread by the user is considered as untrustworthy.

- These studies consider user profile features, which can easily be modified by malicious bots. To avoid feature manipulation.
- Moreover, profile features and social interaction features may not help in detecting malicious URL.

3.2 Proposed System

The proposed trust computation model contains two parameters, namely, direct trust and indirect trust. Moreover, the direct trust is derived from Bayes' theorem, and the indirect trust is derived from the Dempster– Shafer theory (DST) to determine the trustworthiness of each participant accurately. Experimentation has been performed on two Twitter data sets, and the results illustrate that the proposed algorithm achieves improvement in precision, recall, F-measure, and accuracy compared with existing approaches for MSBD.

ADVANTAGES OF THE PROPOSED SYSTEM

- Algorithm used in this system provides better performance compared with the existing algorithms, such as learning from unlabeled tweets.
- Analyze the malicious behavior of a participant by considering URL-based features, such as URL redirection, the relative position of URL, frequency of shared URLs, and spam content in URL.
- We evaluate the trustworthiness of tweets (posted by each participant) by using the Machine Learning algorithm

3.3 Proposed System Design

The project involved analysing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigation from one screen to the other well-ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible. Here we are proposing three modules and those are as following:

Dataset

The data given is in the form of a comma-separated values files with tweets and their corresponding sentiments. The training dataset is a csv file of type tweet_id, sentiment tweet , URL-based features, such as URL redirection, the relative position of URL, frequency of shared URLs, and spam content in URL where the tweet_id is a unique integer identifying the tweet, sentiment is either 1 (positive) or 0 (negative), and tweet is the tweet enclosed in "". We will use this single dataset to cross validate our model. The dataset is a mixture of words, bag of words bot, symbols, URLs, and references to people. Words contribute to predicting the sentiment, but URLs and references to people don't. Therefore, URLs and references can be ignored. The words are also a mixture of misspelled words, extra punctuation, and words with many repeated letters. The tweets, therefore, have to be preprocessed to standardize the dataset. Range Index: 99989 entries, 0 to 99988.

Step A: Preparing the Training Set

Step B: Preparing the Test Set

● Step A.1: Getting the authentication credentials

● Step A.2: Authenticating our Python script

● Step A.3: Creating the function to build the Test set

Step C: Pre-processing BOT Tweets in The Data Sets

Step D: Naive Bayes Classifier

● Step D.1: Building the vocabulary

● Step D.2: Matching tweets against our vocabulary

● Step D.3: Building our feature vector

● Step D.4: Training the classifier

3. ARCHITECTURE

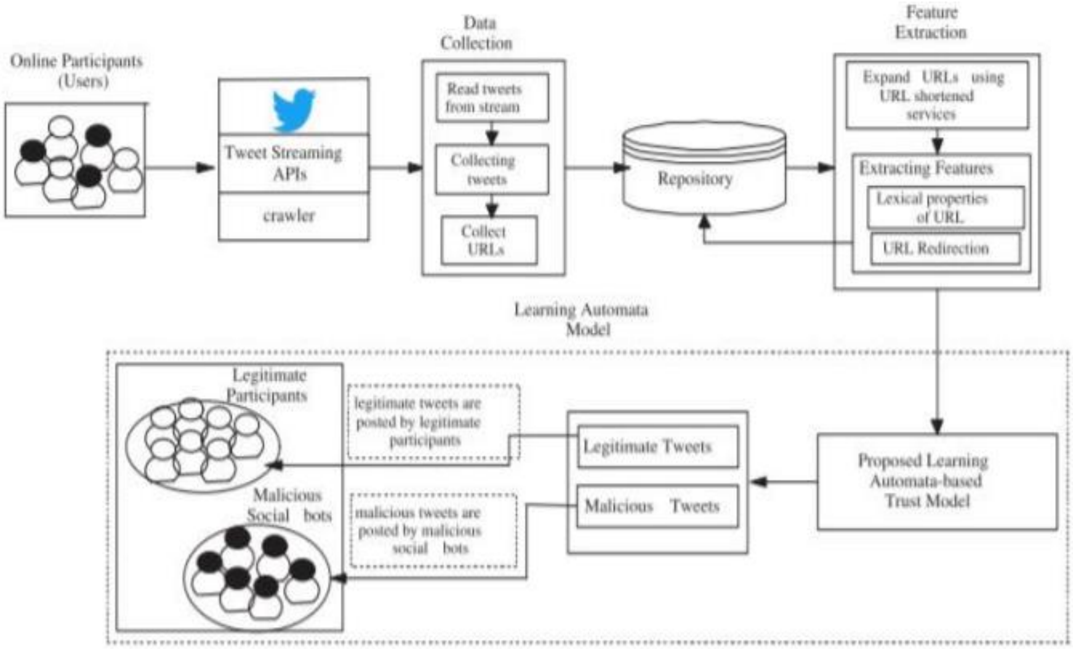
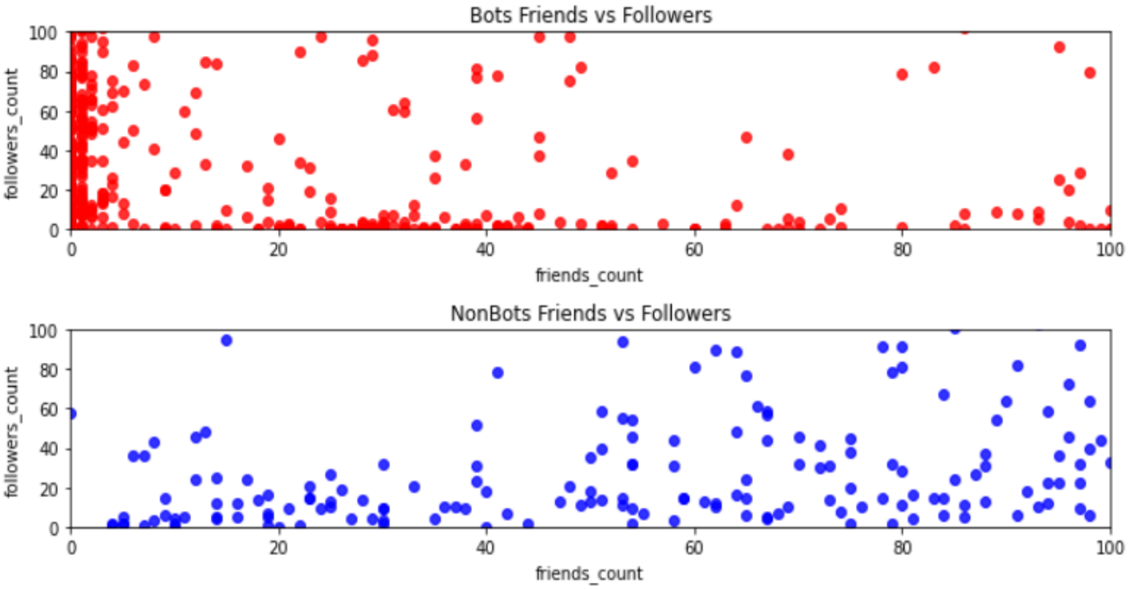


Fig 1: Architecture diagram

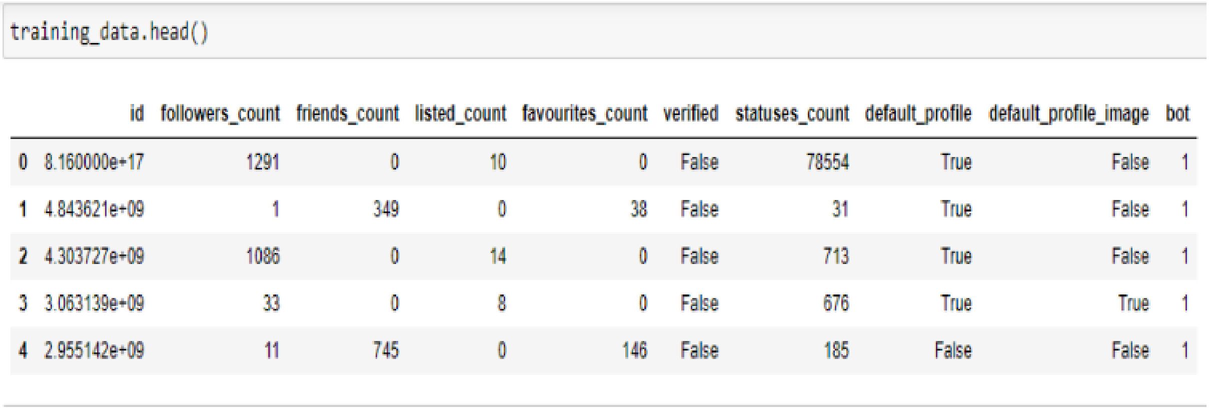
4. RESULTS SCREEN SHOTS

Plotting Bot Friends vs Followers and Non-Bot Friends vs Followers



Plotting Bot Friends vs Followers and Non-Bot Friends vs Followers

Head data of training dataset

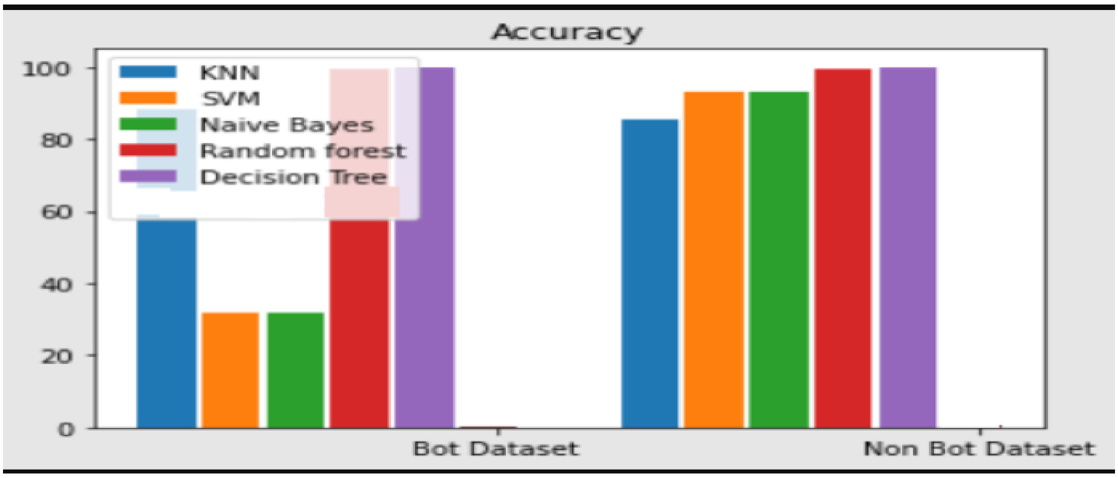


```
training_data.head()
```

	id	followers_count	friends_count	listed_count	favourites_count	verified	statuses_count	default_profile	default_profile_image	bot
0	8.160000e+17	1291	0	10	0	False	78554	True	False	1
1	4.843621e+09	1	349	0	38	False	31	True	False	1
2	4.303727e+09	1086	0	14	0	False	713	True	False	1
3	3.063139e+09	33	0	8	0	False	676	True	True	1
4	2.955142e+09	11	745	0	146	False	185	False	False	1

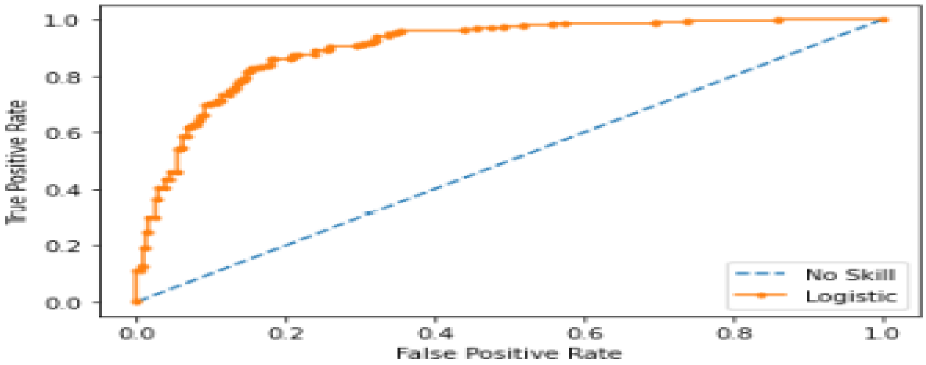
First 5 records of dataset screen shot

Plotting all the algorithms based on their accuracy



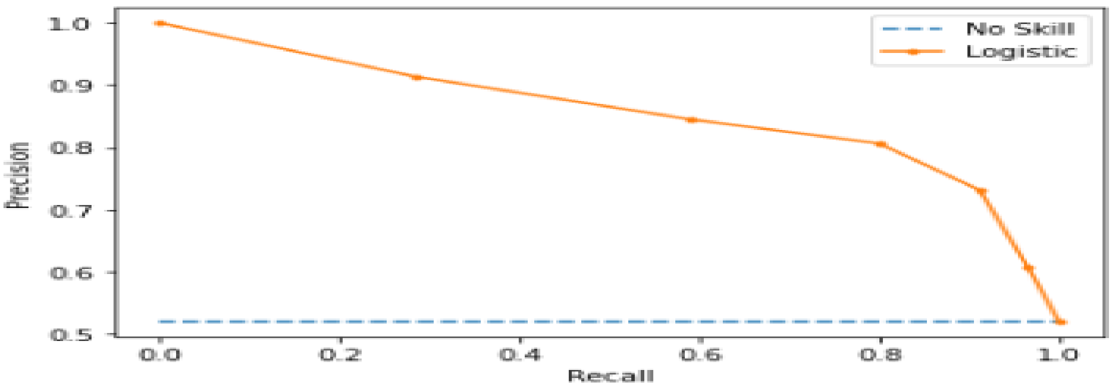
Plotting accuracy of algorithms based on bot vs non bot dataset

No Skill vs Logistic regression ROC AUC



True positive rate vs False Positive Rate

No Skill vs Logistic regression ROC AUC



Precision vs Recall

5. CONCLUSION

This article presents an Multi model Navie Bayes algorithm by integrating a trust computational model with a set of URL-based features for MSBD. In addition, we evaluate the trustworthiness of tweets (posted by each participant) by using the Bayesian learning and DST. Moreover, the proposed Multi model Navie Bayes algorithm executes a finite set of learning actions to update action probability value (i.e., probability of a participant posting malicious URLs in the tweets). The proposed Multi model Navie Bayes algorithm achieves the advantages of incremental learning. Two Twitter data sets are used to evaluate the performance of our proposed Multi model Navie Bayes algorithm. The experimental results show that the proposed LAMSBD algorithm achieves up to 7% improvement of accuracy compared with other existing algorithms. For The Fake Project and Social Honeypot data sets, the proposed Multi model Navie Bayes algorithm has achieved precisions of 95.37% and 91.77% for MSBD, respectively. Furthermore, as a future research challenge, we would like to investigate the dependence among the features and its impact on MSBD.

FUTURESCOPE:

In future work, we envisage quantifying the performance gain during the training of the classifier using the reduced feature set. There is also the intention to explore different one-class algorithms, as they have the property of classifying new types of bots. In the near future, we intend to perform real-time classification using an agent-based approach. Finally, we may carry out an analysis of the behavior of bots during the next elections along the lines of the research conducted

6. REFERENCES

S. Madisetty and M. S. Desarkar, “A neural network-based ensemble approach for spam detection in Twitter,” IEEE Trans. Comput. Social Syst., vol. 5, no. 4, pp. 973–984, Dec. 2018.

• H. B. Kazemian and S. Ahmed, “Comparisons of machine learning techniques for detecting malicious webpages,” Expert Syst. Appl., vol. 42, no. 3, pp. 1166–1177, Feb. 2015.

• H. Gupta, M. S. Jamal, S. Madisetty, and M. S. Desarkar, “A framework for real-time spam detection in Twitter,” in Proc. 10th Int. Conf. Commun. Syst. Netw. (COMSNETS), Jan. 2018, pp. 380–383

• T. Wu, S. Liu, J. Zhang, and Y. Xiang, “Twitter spam detection based on deep learning,” in Proc. Australas. Comput. Sci. Week Multiconf. (ACSW), 2017, p. 3.

• Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, “Key challenges in defending against malicious socialbots,” Presented at the 5th USENIX Workshop Large-Scale Exploits Emergent Threats, 2012, pp. 1–4.

• G. Yan, “Peri-watchdog: Hunting for hidden botnets in the periphery of online social networks,” Comput. Netw., vol. 57, no. 2, pp. 540–555, Feb. 2013.