International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

DETECTION OF MALICIOUS SOCIAL BOTS USING ML TECHNIQUE IN TWITTER NETWORK

V N P SAI SIRI DANTU [1] , JHANSI DEVI TELU[2] ,PADMA SREE KUNCHAM [3] ,GURUDATTA PILLA [4]

[1234] Final Year B.Tech, CSE, Sanketika Vidya Parishad Engineering College, Visakhapatnam,A.P, India
Guided by: Mrs. Gudiwaka Vijayalakshmi, Associate Professor, SVPEC, Visakhapatnam, A.P, India

---

ABSTRACT:

Malicious (spam) social bots generate and spread fake tweets and automate their social relationships by pretending like a follower and by creating multiple fake accounts with malicious activities. Furthermore, malicious social bots post shortened malicious URLs in the tweet in order to redirect the requests of online social networking participants to some malicious and suspicious servers. Hence, distinguishing malicious social bots from legitimate users is one of the most tasks in the Twitter network. To detect malicious or suspicious social bots, extracting URL-based features that include frequency of shared URLs, DNS fluxiness feature, network features, link popularity features and spam content presents in URLrequires less amount of time comparatively with social graph-based features (which rely on the social interactions of users). Moreover, malicious social bots cannot quickly manipulate URL redirection chains. In this, a learning automata-based malicious social bot detection (LA-MSBD) algorithm is a Machine Learning approach proposed by integrating a Naïve Bayes algorithm model with URL-based features(URL Classification and Feature Extraction) for identifying trustworthy participants (users) in the Twitter network. Experimentation has been performed on 2 Twitter data sets, and the results obtained illustrate that the proposed algorithm achieves improvement in precision and detection accuracy.

KEY WORDS: ***Learning Automata, URLfeatures, Malicious Social Bots, URL Classification, Feature Extraction, Online social network.***

INTRODUCTION

Twitter being a micro-blogging platform used by an increasing population of users of different age groups over the last decade. Generally, people post tweets and interact with other users as well. More specifically, they (users) can follow (following/friends) their favorite politicians, celebrities, athletes, entrepreneur, artists, friends and get followed by them (followers). Furthermore, Twitter generates a list of the topics being discussed day-to-day updates, that so called trending topics. Hence, users can get informed about the hot topics of discussion on a daily basis. And generally online social networks (OSNs) are increasingly used by automated accounts, well known as bots, due to their immense popularity across a wider range of user categories.It is estimated that over 15% of accounts on Twitter are automated bot accounts.A customer support chatbot is a prime example of a Twitter bot.It can help and improve the overall customer support experience by improving the response time. Following few are the most useful and amazing bots on Twitter.

@HundredZeros: Twitter bot that frequently recommends amazing and thought provoking e-books that are free on Amazon. This helps followers and avid readers find great titles and content to read.

@MagicRealismBot: Magic realism is quite an amusing Twitter bot that argues the existence and significance of magic in the real world. The tweets posted by this bot are some of the funniest tweets that one can find.

@DearAssistant: Virtual Assistants (Google Assistant, Siri, Alexa) became the most used medium to extract relevant information in any aspects. @DearAssistant is a Twitter bot developed to provide answers to questions like the definition of words, the distance between places, and many other things. There do exists different facet of those automatic bots which ends in a very nice loss. Spam bots faux like legitimate users by making pretend accounts and ID's and posting same tweets repeatedly and spreading pretend news and conjointly tweets which will aid to malicious servers and successively ends up in forceful consequences. Their main purpose is that the dissemination of pretend news, the promotion of specific ideas and merchandise, the manipulation of the securities market. By posting tweets very often, they influence measures together with the trending topics. As a consequence, legitimate users cannot distinguish between real trending topics and pretend ones. In Twitter, once a participant (user) desires to share a tweet containing URL(s) with the neighboring participants (followers or followees), the participant adapts uniform resource locator shortened service so as to cut back the length of uniform resource locator (because a tweet is restricted up to one hundred forty

characters). Moreover, a malicious social larva might post shortened phishing URLs within the tweet. Twitter bots are often an excellent facilitate in several distinctive ways that, there square measure cases wherever they were used unethically and illicitly. Hence, it's vital to identifying malicious social bots from legitimate users is one in all the foremost vital tasks within the Twitter network. The researchers at Indiana and North-eastern University had developed a brand new tool referred to as BOTOMETER, that tells concerning the chance of a Twitter network user being a larva. It's extremely troublesome to see AN threshold share to observe the bots however or so the score is nearer to 100%, the likelihood of the account being a larva will increase. The systems square measure being trained to acknowledge the larva behavior and analyze supported the patterns in a very dataset of over thirty,000 accounts that were initial verified by the human researchers as either bots or non-bots. Botometer a tool that "reads" over a thousand different characteristics, or "features," for each account and then assigns the account a score between 0 and 1. The higher the score, the more the chances the account is automated. By the process and experiments being done one can estimate and understand the level of difficulty is too high and even time consuming to detect an twitter account is being or not and then the main task comes into role while the account if detected as bot is being purposive or subversive. Detecting an account is automated or not involves complicated steps and again detecting that automated account is malicious or legitimate is more complicated. Several techniques, including supervised, unsupervised and reinforcement learning, have been proposed to detect bots and its malicious activities in Twitter. These techniques mainly use a limited number of features extracted for identifying the automated accounts at account-level. However, there exists bots that have grown mechanisms to mimic human behavior and avoid detections. Therefore, new techniques should be proposed for securing the legitimate users from the proliferation of malicious accounts in the Twitter Network.

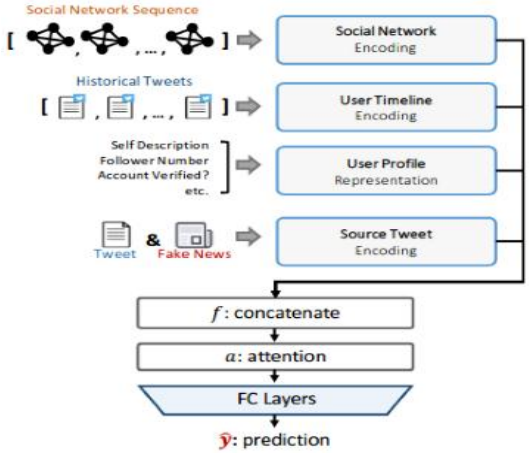
METHODOLOGY:

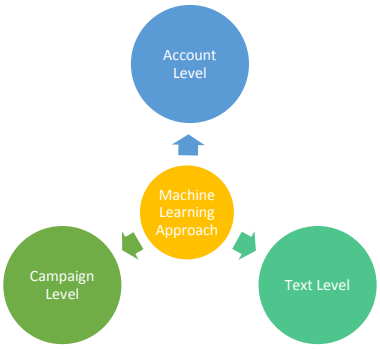
Several approaches depend on the usage of machine learning and the extraction of certain types of features from user accounts, posts, or social graphs. These approaches have proven successful but, they require control from the user, high resources, and more processing time. During recent times, deep-learning methods have surpassed previous approaches in the speed and efficiency, and they have no necessary user intervention. Extracting features from text rather than regular data has became the dominant trend in recent deep-learning-based on studies. This happens due to the ability of deep-learning algorithms to extract and track hidden patterns within the available texts that an existing traditional approach may not be able to detect or predict accurately.Presented an up-to-date survey and analysis regarding related research to spam detection on Twitter accounts and data being shared in it.Collected a large tweet dataset from thetwitter API using a premium Twitter API that provides unique features not available to the public. Such a dataset could be served as a benchmark dataset for other researchers including URL classifications and feature extraction from the data collected in the field.Developing a novel framework that combines text-based and metadata-based features for detecting malicious accounts on Twitter using machine-learning algorithms.Investigated the effectiveness of combining accessible metadata with textual data integrating URL based features when identifying spam accounts on Twitter.Benchmarking the proposed framework with the most prominent models using URL based features of machine learning and deep learning applied to spam detection in twitter network.

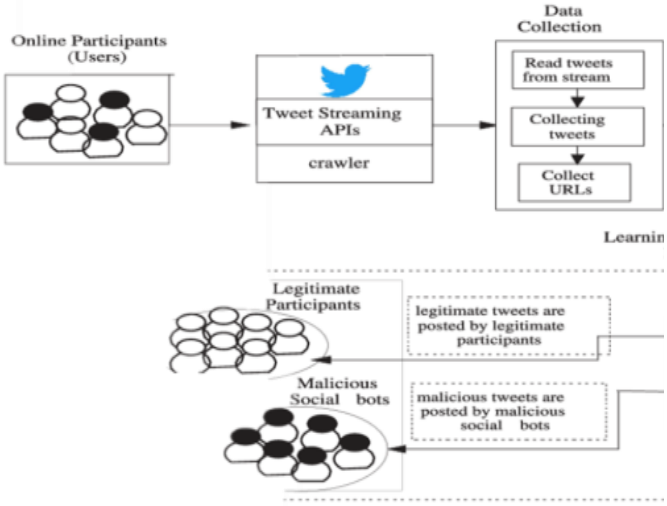
EXISTING SYSTEM:

There are few existing systems that have only considered social graph based features such as user timeline, account details, and tweets posted and the response time when tweeted and re-tweeted . Existing system aimed to design a framework by considering the features set to be evaluated the trust value of each online social network account and to detect the malicious social chat bots in the Twitter network effectively and efficiently. Furthermore, it defined two trust components included in the algorithm, namely, direct and indirect trusts to determine the trust value of each participant or user in the twitter network. Learning Automata is one of the specified reinforcement machine learning algorithm. LearningAutomata is a modifiable and flexible decision-making sector and learns an optimal action by repeatedly interacting with the surrounding environment. At each iteration, Learning Automata chooses a specific action from the specified finite set of actions and provides a response (or reinforcement signal) in the terms of reward and penalty. Based on the available response from the environment, Learning Automata updates and proceeds its action probability value to obtain the maximum reward from the surrounding environment. Learning Automata can be represented through a six-tuple < L A, A, pr,β,rs, F >, where as L A = {la1,la2,...,lan} that represents the finite set of states of an automata, A = {a1, a2,..., an} that represents the set of actions to be performed (also known as outputs of LA), and pr = {pr1, pr2,..., prn} that represents the set of action probability values of each learning action being selected. Let β = {β1, β2,...βn} represent a finite set of the reinforcement signals, where βi∈ {0, 1}, 1 represents reward, 0 represents

penalty, and F : pr×β → pr represents theimplementation and improvement of sequence of action probability values with respect to the current action probability value and the response from the environment simultaneously. However, the work is different from other existing works in the sense that we focus on detecting malicious social bots based on the LA model with the trust computational model through social graph based features. The LA has also been successfully applied in various areas, such as Internet of Things (IoT), cloud computing, online social networks, wireless networks, and image processing .







PROPOSED SYSTEM:

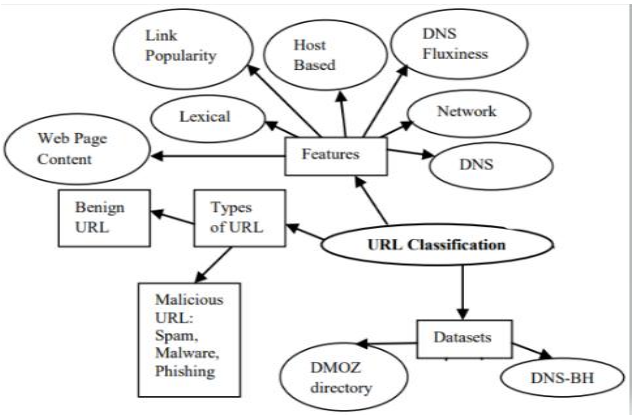
A Learning Automata model has been proposed to identify the spatio-temporal patterns in given noisy sequences. Learning Automata is mostly robust to handle the noisy data in any adversarial environment. Moayedikia proposed a Learning Automata-based method to label the ground truth without human intervention and to avoid genuine interpretation by manually observing users' behavioral patterns. A simple probabilistic classifier based on applying the Bayes theorem from Bayesian statistics with strong naïve independence assumptions are known as Naïve-Bayes classifier. Furthermore, the fundamental probability model is being described as "independent feature model". In simple words, a Naïve-Bayes classifier algorithm assumes and predicts that the presence or absence of a particular feature of a class is not related to the presence or absence of any other feature. Even if these features are actually dependent on each other or upon the existence of the other features However, the work is different from other existing works in the sense that we focus on detecting malicious social bots based on the LA model with the Naïve Bayes algorithm with URL based features.

Naïve-Bayes

PROBLEM FORMULATION:

According to given Twitter network G = (P, E), where P represents a participant set P = {p1, p2,..., pn} and E (i.e., E ⊆ P × P) represents a social relationship set (or directed edges) between the participants (users). If there exists a social relationship between two participants, then they are considered as neighbour's (that can be called either followers or followees). According to a given Twitter network with n participants and series of m tweets twpi = {twi1, twi2,...,twim} posted by each participant pi, a feature finite set F = { f1, f2,..., fn } can be constructed from each individual tweet posted by each participant. In this work, we predict that features are independent to each other. Based on the URL-based features such as URL redirection, frequency of shared initial URLs, and spam content in

Proposed system framework works effectively including mainly 3 stages. Stage-1 consist oftraining of the data being collected during the data collection and Stage 2 involves supervised learning with the available training data and Stage 3 involves malicious URL detection and attack type Identification. These three stages can operate consecutively (one after another) as in batched learning, or in an interleaving manner and furthermore, additional data is also collected to incrementally train the classification algorithm model while the model is being used in the detection and there by identification. Here, learning process is entirely based on the feature extraction module in which the training set is combined with the available number of URL's in the learning process.



FEATURE EXTRACTION MODULE:

It is one of the mandatory and integral process involved in which feature extraction is mainly based on following six features and they are Lexical Features, Link Popularity Features ,DNS ,DNS Fluxiness, Network Features, Web page content Features. Datasets are trained and those data sets are of two types ,one is of static and other is of live that is run time, DMOZ directory is of static type. For an unbiased classification in the proposed framework, we split the entire dataset (static –DMOZ directory) and 50% of the instances are used for the training while the rest of the data set that is 50% is used for testing purposes. Performance can be reported based on the accuracy that is obtained at each level of the classifier. The necessity and the essentiality of an online learner lies in the fact that it builds a classification model in such an incremental manner when it fed with a number of URL training data in an 'on-the-fly learning(online learning)' pattern or fashion.

As mentioned the 6 features of the feature extraction module works as follows;

LEXICAL FEATURES:

Malicious and spam URLs, specifically those for the kind of phishing attacks, usually have distinguishable and vivid patterns in their URL exists. Among these lexical options that are present, eventually the typical domain/path token length (delimited by '.', '/', '?', '=', '-', '') and specified name and other specifications presence were driven from a study by McGrath and Gupta that phishing URLs show completely various distinct and different lexical patterns. It also includes domain token count, path token count, domain token length, path token length and also involves longest domain token length and longest path token length.

LINK POPULARITY FEATURES:

One among the available foremost essential and necessary options utilized in this technique are "link popularity (feature)", that is calculable by the examination of the amount that is being evaluated of incoming links from existing alternative websites. Malicious (spam) sites tend to acquire or possess a minimum value amount of link popularity, however several surviving benign sites tend to acquire or possess a highest (maximum value of) amount of link quality. Each link popularity of an registered address and link popularity of the specified URL's domain are utilized in this technique.

NETWORKFEATUREEXTRACTION:

Domain look up time, Download speed, Actual downloaded bytes are being calculated in implementing the network feature module as attackers might take the help of redirections and address shortening and redirection such as i-frame. DNS FEATURES: Domain Naming System can easily and instantly map to a new IP (internet protocol) address if the host's IP address gets modified and also they are easier andquick to remember

than an IP address. It allows the organizations to utilize a domain name hierarchy that is independent of any other IP address assignment.

DNS FLUXINESS FEATURES:

A freshly rising fast-flux service network (FFSN) establishes a proxy network to host extralegal online services with a really high efficient and convenience. To detect URLs which are served by FFSNs, it uses

$$\Phi = \frac{N_{ip}}{N_{single}} \text{Fluxiness} = \frac{\text{Total no. of unique ip's}}{\text{no. of ip's}}$$

THIS process involves receiving of URLs to be tested and extracting features associated with the URL in the classification module. Then employing Naïve-Bayes model for detecting and classifying malicious URL's. After that further classification takes place whether it is malicious URL or not, if output obtained as malicious it can be further classified into following types such as spam, phishing and malware. The proposed algorithm is based on Naive-Bayes algorithm and it takes input of training data and URLs to be tested and evaluated output is obtained as testing domain names with their attack type. Mathematical Model, however the Naive-Bayes algorithm Rule is the basis for several machine-learning algorithms and data mining methods. The mentioned algorithm model is used to create and analyse models with the predictive assertions and the capabilities. It specifically provides new ways further of exploring and understanding the available data. It is used when data is highly available and efficiency to be obtained in the output is comparatively high when compared to other methods.

ALGORITHM USING NAÏVE-BAYES:

Following algorithm is the proposed algorithm to detect the genuine users in the twitter network and to detect the malicious social bots by feature extraction of URL's.

STEP 1: Input; Training set and URL's are given that are to be tested.

STEP 2: For the extracted and available feature , calculate its sub-features through training set for the purpose of training.

STEP 3: Step 2: using a Gaussian distribution a training set is created mean and variance of each sub feature is calculated.

STEP 4: Probability of each and every individual class is also calculated.

STEP 5: Posterior for each class (Benign, SpamandMalware) is calculated and evaluated. posterior=(prior*likelihood)/evidence.

Step 6: Perform the analysis of the posterior values of each individual class.

Step 7: Among the available 4 classes, class with greater value ofposterior is assigned to testing domain.

CONCLUSION:

This article presents an LA-MSBD (Learning Automata - Malicious Social Bot Detection) algorithm by integrating a Naïve-Bayes algorithm model with a set of URL-based feature extraction for malicious social bot detection in twitter network. In this research work ,it is an extension for the existing system that is based on trust computational model of direct and indirect trust and the proposed system is an enhancement of the existing system. In the current system proposed using Naive-Bayes algorithm and through extracting URL based features and implementing the classifier algorithm model and URL features extracted from the tweets are tested and evaluated and finally tweets from malicious bots and legitimate users are distinguished.

RESULT:

REFERENCES:

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] A. K. Jain and B. B. Gupta, "A machine learning based approach for phishing detection using hyperlinks information," J. Ambient Intell. Hum. Comput., vol. 10, no. 5, pp. 2015–2028, May 2019.

T. M. Chen and V. Venkataramanan, "Dempster-shafer theory for intrusion detection in ad hoc networks," IEEE Internet Comput., vol. 9, no. 6, pp. 35–41, Nov. 2005.

[21] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in Proc. 26th Int. Conf. World Wide Web Companion- (WWW Companion), 2017, pp. 963–972.

[22] K. Lee, B. D. Eoff, and J. Caverlee, "Seven months with the devils: A long-term study of content polluters on Twitter," in Proc. ICWSM, 2011, pp. 1–8.

[23] C. Besel, J. Echeverria, and S. Zhou, "Full cycle analysis of a largescale botnet attack on Twitter," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2018, pp. 170–177.

[24] J. Echeverria and S. Zhou, "Discovery, retrieval, and analysis of the'star wars' botnet in twitter," in Proc. 2017 IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining 2017, 2017, pp. 1–8

[25] A. Dorri, M. Abadi, and M. Dadfarnia, "SocialBotHunter: Botnet detection in Twitter-like social networking services using semisupervised collective classification," in Proc. IEEE 16th Int. Conf. Dependable, Autonomic Secure Comput., 16th Int. Conf. Pervasive Intell. Comput., 4th Intl Conf Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech), Aug. 2018, pp. 496–503.

[26] M. Agarwal and B. Zhou, "Using trust model for detecting malicious activities in Twitter," in Proc. Int. Conf. Social Comput., Behav.-Cultural Modeling, Predict. Springer, 2014, pp. 207–214.

C. Yang, R. Harkreader, and G. Gu, "Empirical evaluation and new design for fighting evolving Twitter spammers," IEEE Trans. Inf. Forensics Security, vol. 8, no. 8, pp. 1280–1293, Aug. 2013.

[29] M. Al-Janabi, E. D. Quincey, and P. Andras, "Using supervised machine learning algorithms to detect suspicious URLs in online social networks," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, Jul. 2017, pp. 1104–1111

[30] S. Lee and J. Kim, "WarningBird: A near real-time detection system for suspicious URLs in Twitter stream," IEEE Trans. Dependable Secure Comput., vol. 10, no. 3, pp. 183–195, May 2013.

BIOGRAPHIES:

G.VIJAYALAKSHMI
Currently Associate professor from Department of computer science and Engineering at Sanketika Vidya Parishad Engineering college.

D V N P SAI SIRI
Pursuing B-tech from Department of computer science and Engineering at Sanketika Vidya Parishad Engineering College

T.JHANSI DEVI
Pursuing B-tech from Department of computer science and Engineering at Sanketika Vidya Parishad Engineering College

K.PADMA SREE
Pursuing B-tech from Department of computer science and Engineering at Sanketika Vidya Parishad Engineering College

P.GURUDATTA
Pursuing B-tech from Department of computer science and Engineering at Sanketika Vidya Parishad Engineering College